Data-Free Deep Image Recovery from Sparsely Corrupted Incomplete Measurements

Xiang Li
forkobe@umich.edu

Can Yaras
cjyaras@umich.edu

Abstract

A key task in recovering an image from incomplete measurements is the correct choice of prior over the signal space. Recently, deep image prior (DIP) has shown that randomly initialized convolutional networks have a strong inductive bias for natural image generation, and have been demonstrated to yield SOTA performance for various inverse problems in image recovery without learning from any extra data. However, due to the over-parameterization of the network, the reconstruction process eventually overfits any noise in the measurements, requiring early-stopping. To overcome this issue for *sparsely* corrupted measurements, we further over-parameterize it with a two-layer diagonal network in the measurement space, whose implicit bias towards sparsity matches the nature of the noise. Our proposed CS-DODIP algorithm demonstrates near-optimal performance on a variety of image recovery tasks with sparsely corrupted measurements such as Gaussian sensing, inpainting, and super-resolution. Code for all experiments can be found at `https://github.com/cjyaras/cs-dodip`.

1 Introduction

In the past few decades, significant progress has been made in compressed sensing and inverse problems in signal processing, leading to advances across numerous scientific and engineering fields in applications such as medical imaging [1], astronomy [2], and photography [3], among many others. The central task in inverse problems is to estimate a ground-truth signal $x^* \in \mathbb{R}^n$ from its measurements

$$y = \mathcal{A}(x^*) + \epsilon \tag{1}$$

where $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a (typically non-invertible) operator and $\epsilon \in \mathbb{R}^m$ is noise. In many applications of interest, it is desirable to reconstruct x^* from as few measurements as possible - for example, in *magnetic resonance imaging* (MRI), each entry of y is a single Fourier component obtained by applying a spatially varying magnetic field to the sample, for which collecting every possible spatial frequency is time consuming. Therefore, it is often the case that $m \ll n$ and so the system is highly underdetermined, meaning that without any prior information about the nature of the signal x^*, the problem is ill-posed.

Fortunately, many signals in practice are well-structured and can be recovered from measurements y under certain regularity conditions for the sensing operator $\mathcal{A}$ and the signal x. For example, the Nyquist–Shannon sampling theorem [4] ensures that for a bandlimited signal x^* with no frequency content above f, if the sampling rate of $\mathcal{A}$ is at least $2f$ then the signal x^* can be perfectly reconstructed (when $\epsilon = 0$) via sinc interpolation. In compressed sensing, the signal x is assumed to be sparse in a proper basis Ψ – then under certain incoherence conditions for $(\mathcal{A}, \Psi)$, the signal x can be recovered exactly (when $\epsilon = 0$) with high probability via ℓ_1 minimization [5].

The question that remains then is, *what is the correct prior for image recovery?* While many previous approaches have considered handcrafted or data learned priors, recently it has been shown that *untrained, generative convolutional neural networks* (CNNs) with fixed random inputs exhibit an inductive bias toward natural images, known as *deep image prior* (DIP) [6]. More specifically, for $f_\theta(z)$ where z a fixed random input and θ are network parameters, the problem

$$\min_\theta \|f_\theta(z) - x\|^2 \tag{2}$$

converges to zero loss much quicker for a noiseless natural image x as opposed to a noisy natural image or even purely noise x. In the original work, it is demonstrated that by parameterizing an image as $\hat{x}(\theta) = f_\theta(z)$, the DIP can be leveraged to solve various image recovery problems such as denoising, inpainting, super-resolution, and network feature inversion from measurements of the form in (1) via

$$\min_\theta \|\mathcal{A}(\hat{x}(\theta)) - y\|^2 \tag{3}$$

meanwhile outperforming handcrafted features and approaching the performance of trained networks. However, in the presence of measurement noise, since the network f is over-parameterized it will eventually overfit to the noise, meaning that early-stopping is required to achieve an accurate reconstruction. To address this, the work [7] introduces additional over-parameterized terms to the DIP for the image denoising problem, dubbed *double over-parameterization* (DOP). Specifically, for an observation $x = x^* + s$ where s is sparse noise, the authors solve the optimization problem

$$\min_{\theta,g,h} \|\hat{x}(\theta) + g \odot g - h \odot h - x\|^2 \tag{4}$$

using gradient descent with discrepant step sizes for θ and g, h. The idea is that the implicit bias of the DIP term $\hat{x}(\theta)$ fits the natural image x^*, whereas the implicit sparsity bias of the additional term $g \odot g - h \odot h$ fits the sparse noise s. However, the work [7] only considers the image denoising problem, which is the simplest case of the inverse problem in (1). In particular, the operator $\mathcal{A}$ is the identity (measurement space $\equiv$ signal space) which is *invertible* and does not entirely capture the typical challenges associated with most inverse problems.

In this work, we consider the general setting of recovering image $x^* \in \mathbb{R}^{C\times H\times W}$ from measurements

$$y = \mathcal{A}(x^*) + s \in \mathbb{R}^\ell \tag{5}$$

where $\mathcal{A}$ is non-invertible (and even $\ell \ll C \cdot W \cdot H$) and s is sparse noise of arbitrary energy. Sparsely corrupted measurements are often prevalent in applications where only a small subset of the measurement sensors are faulty, e.g., a camera with defective pixels. We utilize the DIP as a data-free image prior, and introduce a two-layer diagonal network for denoising measurements, rather than the original signal as in [7]. We demonstrate that this approach which we call CS-DODIP (*compressive sensing with doubly over-parameterized deep image prior*) outperforms similar data-free or untrained approaches across a variety of inverse problems such as Gaussian sensing, inpainting, and super-resolution.

Related Work Prior to the dominant success of deep learning, most approaches to the image recovery problem were inspired by the compressed sensing and sparse recovery literature, where the image prior is induced via choice of basis for which the underlying signal is sparse, such as discrete cosine transform (DCT) [8] and discrete wavelet transform (DWT) [9]. The image prior can also be specified in the original signal space, such as with total-variation regularization [10], which assumes that integral of absolute image gradient is small. In contrast to handcrafted bases or regularizers, the image prior can also be *learned* from data. For instance, overcomplete dictionary learning [11] computes a sparse local patch-based dictionary or basis from natural (unlabeled) image data that frequently outperforms pre-fixed bases such as wavelets on various image recovery tasks [12]. A more general approach involves an implicit probability model for the image prior at a local patch level, such as Gaussian mixture models [13] or Dirichlet processes [14].

For the past decade, deep learning-based approaches comprise nearly all state-of-the-art methods for image recovery problems such as inpainting and super-resolution. Among supervised methods, the most common approach is to train (using collected image-measurement pairs) an end-to-end network to predict the original image from a given measurement. Choices of network architecture include denoising auto-encoders [15], U-Nets [16], convolutional framelets [17], among many others, but overall largely depend on the specific task. On the other hand, generative networks such as variational auto-encoders (VAEs) [18] and generative adversarial networks (GANs) [19] can be trained on unlabeled image data to capture a task agnostic statistical prior. The generator can then be applied to the inverse problem by optimizing over the latent space [20]. The number of deep learning-based approaches for inverse problems are innumerable – see [21] for a comprehensive survey on the topic.

The deep image prior (DIP) [6] is uniquely distinct from other deep learning-based methods, in the sense that there is no training (supervised or unsupervised) prior to estimation. Rather than exploiting the approximation power of deep neural networks, DIP utilizes the implicit *structure* of randomly initialized CNNs to give a powerful handcrafted prior for natural images. DIP has previously been applied to compressive sensing problems with added total-variation regularization and learned regularization terms, dubbed CS-DIP [22]. However, CS-DIP tends to lose detail in the image to due to the TV regularization, and furthermore the learned regularization term requires some training data to tune properly. To our knowledge, the proposed CS-DODIP method is the only deep-learning based approach for compressive sensing with sparse noise that yields robust SOTA performance without early-stopping or training data.

2 Problem Formulation & Methods

In this section, we state the proposed CS-DODIP method in detail, and also describe comparable baselines for benchmarking our approach in Section 3.

2.1 CS-DODIP

Due to the over-parameterization of f_θ, the original DIP formulation for inverse problems will eventually overfit any noise in the measurements, yielding a subpar recovery. Inspired by [7], we (somewhat counter-intuitively) introduce additional over-parameterization in the form of a two-layer diagonal network $n(g, h) = g \odot g - h \odot h$ to "predict" the sparse corruption in the measurements and solve the problem

$$\min_{\theta, g, h} \|\mathcal{A}(f_\theta(z)) + n(g, h) - y\|^2 \tag{6}$$

for fixed random z. The key idea behind CS-DODIP is that for sparsely corrupted measurements of the form $y = \mathcal{A}(x^*) + s$, the implicit *sparsity* bias (see Corollary 2 in [23]) of GD for the network $n(g, h)$ matches s (for small initialization of g and h) while the implicit *natural image* bias of the network $f_\theta(z)$ matches x^*. Therefore, if $n(g, h)$ can match s closely, then (6) reduces to the original noiseless DIP problem.

The main novelty of our approach compared to the double over-parameterization (DOP) approach for DIP introduced in [7] is that we model noise in the compressed measurement space rather than the original signal space, so we can tackle non-trivial inverse problems rather than the special case $\mathcal{A} = \mathcal{I}$. In general, CS-DODIP may not work – indeed, one can easily construct pathological examples of operators $\mathcal{A}$ for which image recovery is ill-posed. However, in Section 3 we show that for several interesting inverse problems, the method is highly performant.

In our algorithm, f_θ is a U-Net with skip connections optimized with Adam with learning rate η_1, while $n(g, h)$ is optimized with SGD using learning rate η_2.

Algorithm 1: CS-DODIP

Data: Sensing operator $\mathcal{A}$, measurements y, learning rates η_1, η_2, initialization scale α, number of iterations T
Result: Recovered image x^*
Initialize U-Net f_θ;
Initialize $g, h \sim \mathcal{N}(0, \alpha^2 I)$;
Sample and fix $z \sim \mathcal{N}(0, I)$;
for $= 0, \ldots, T-1$ **do**
  $n(g, h) \leftarrow g \odot g - h \odot h$;
  $\mathcal{L}(\theta, g, h) \leftarrow \|\mathcal{A}(f_\theta(z)) + n(g, h) - y\|^2$;
  Update θ with Adam optimizer step on $\mathcal{L}$ using learning rate η_1;
  Update g, h with SGD optimizer step on $\mathcal{L}$ using learning rate η_2;
$x^* \leftarrow f_\theta(z)$;

2.2 Baselines

For fair comparison, we only consider baselines that do not require any training data. Furthermore, we use an ℓ_1 penalty for the data fidelity term, since these methods do not account for sparse noise in the measurements.

2.2.1 DCT-Lasso

Using a traditional compressive sensing approach, we solve the Lasso-type problem

$$\min_{w} \|\mathcal{A}(\Psi(w)) - y\|_1 + \lambda\|w\|_1 \tag{7}$$

where Ψ is the two-dimensional inverse discrete cosine transform (2D-IDCT) and $\lambda > 0$. We consider a cosine basis over wavelets, as it was reported in [22] that cosines outperforms wavelets across a variety of CS problems. The recovered image is given by $\hat{x} = \Psi(w^*)$.

2.2.2 Robust-DIP

We simply solve the original DIP problem with an ℓ_1 loss, i.e., we solve

$$\min_{\theta} \|\mathcal{A}(f_\theta(z)) - y\|_1 \tag{8}$$

where f_θ is a U-Net with skip connections optimized with Adam and z is a random fixed input. The recovered image is given by $\hat{x} = f_{\theta^*}(z)$.

2.2.3 TV-DIP

We add total-variation (TV) regularization to the ℓ_1 DIP problem, i.e., we solve

$$\min_{\theta} \|\mathcal{A}(f_\theta(z)) - y\|_1 + \lambda\|f_\theta(z)\|_{\mathrm{TV}} \tag{9}$$

where $\lambda > 0$ and

$$\|x\|_{TV} = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|. \tag{10}$$

This approach is closely related to that of [22] without the learned regularizer, which requires some training data. The recovered image is given by $\hat{x} = f_{\theta^*}(z)$.

3 Experimental Results

In this section, we demonstrate the improved performance of CS-DODIP over similar baselines outlined in the previous section for sparsely corrupted Gaussian sensing, inpainting, and super-resolution problems. To benchmark each method, we employ peak signal-to-noise ratio (PSNR), a commonly used metric for measuring reconstruction quality, defined as

$$\mathrm{PSNR}(x, \hat{x}) = -20\log_{10}\left(\frac{\|x - \hat{x}\|_2}{N}\right) \tag{11}$$

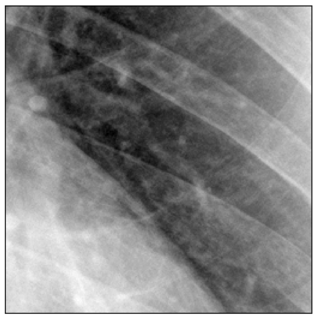
where N is the number of pixels in each image.

3.1 Gaussian Sensing

In the context of Gaussian sensing, (5) takes the form

$$y_i = \langle A_i, x^* \rangle + s_i,\ i = 1, \ldots, \ell \tag{12}$$

where $\{A_i\}_{i=1}^{\ell}$ have *i.i.d.* $\mathcal{N}(0, 1/\sqrt{\ell})$ distributed entries. In our example, we assume that the noise s is p-sparse with $p = 0.1$, and the non-zero entries of s are *i.i.d.* $\mathcal{N}(0, \sigma^2)$ distributed with $\sigma = 1$. We take $\ell = 8000$ measurements of an X-ray image shown in Figure 1a. The results of all methods are shown in Figure 3a, Figure 2a, and Table 1. We see that DCT-Lasso cannot produce a high-resolution image, while Robust-DIP begins to overfit to the noisy measurements. In this case, TV-DIP and CS-DODIP perform very similarly in PSNR, although the CS-DODIP results in more detailed reconstruction image.

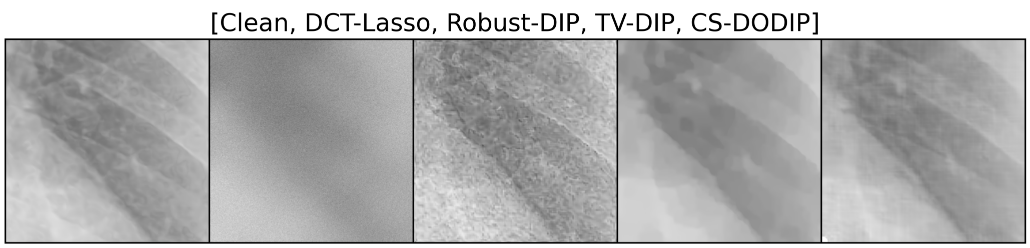


(a) **X-ray**

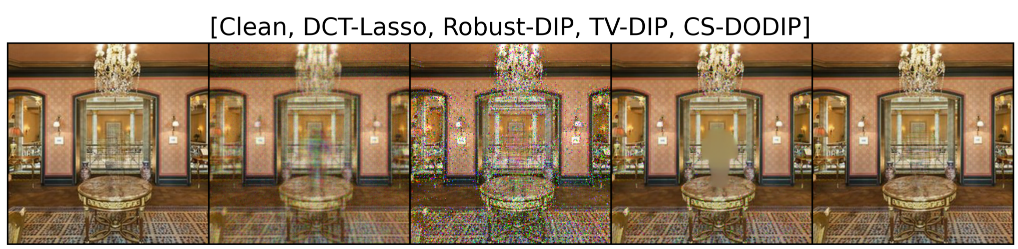
(b) **Vase**

(c) **Zebra**

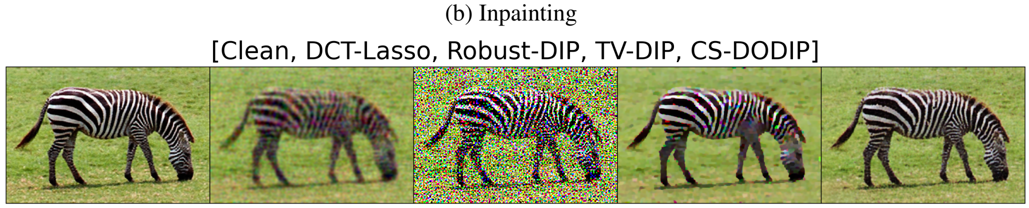
Figure 1: **Examples for Image Recovery Problems.** (a) Grayscale image used for Gaussian sensing problem. (b) Image used for inpainting problem. *Left*: Original image. *Middle*: Image after removing vase region. *Right*: Image after masking as well as salt-and-pepper noise. (c) Image used for super-resolution problem. *Left*: Original image. *Middle*: Image after Lanczos downsampling. *Right*: Image after downsampling as well as salt-and-pepper noise.



(a) Gaussian sensing

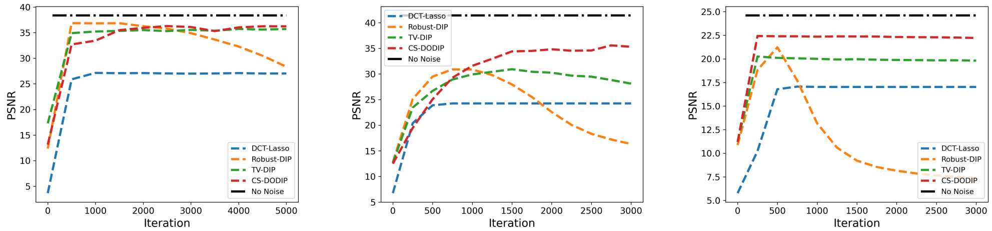


(b) Inpainting



(c) Super-resolution

Figure 2: **Reconstructed Images**. Resulting images after running each method to final iteration.



(a) Gaussian Sensing (b) Inpainting (c) Super-resolution

Figure 3: **PSNR vs. Iteration**. PSNR measured between original image (or non-masked portion of image for inpainting) and reconstructed image.

3.2 Inpainting

Inpainting refers to the problem of filling in an image with a missing dense region. In this case, (5) takes the form

$$y_i = \langle A_i, x^* \rangle + s_i,\ i = 1, \ldots, \ell \tag{13}$$

where A_i contains all zeros except for a single 1 in the position of an observed pixel and s is p-sparse salt-and-pepper noise. For our example, we have $p = 0.2$ and $\mathcal{A}$ is given by the mask illustrated in Figure 1b. The results of all methods are shown in Figure 1b, Figure 2b, and Table 1. We see that CS-DODIP produces the cleanest image whereas all other methods produce some artifacts in the inpainted region. There is also a large PSNR gap to the closest approach which is TV-DIP. In this example, CS-DODIP performs drastically better than the other approaches.

3.3 Super-resolution

Super-resolution refers to the problem of enhancing the resolution of an image from its downsampled or degraded version. In our case, (5) takes the form

$$y = \mathcal{A}(x) + s \tag{14}$$

where $\mathcal{A}$ is a downsampling operator and s is p-sparse salt-and-pepper noise. For our example, we have $p = 0.2$ and $\mathcal{A}$ is a non-linear Lanczos interpolation downsampler, the effect of which is shown in Figure 1c. The results of all methods are shown in Figure 1c, Figure 2c, and Table 1. We see that CS-DODIP produces the sharpest image whereas all other methods are either blurry or contain artifacts. Similarly, CS-DODIP has 2-3 greater PSNR than TV-DIP, the next best method.

Table 1: Final PSNR measured between original image (or non-masked portion of image for inpainting) and reconstructed image.

	DCT-Lasso	Robust-DIP	TV-DIP	**CS-DODIP**	Clean
Gaussian Sensing	27.0	28.3	35.7	**36.2**	38.4
Inpainting	24.3	16.3	28.1	**35.3**	41.6
Super-resolution	17.0	7.4	19.8	**22.2**	24.6

4 Conclusion

In this report, we proposed CS-DODIP, an extension of the method of double over-parameterization [7] for deep image prior to more general settings of image recovery from sparsely corrupted measurements beyond simple image denoising. We demonstrated the effectiveness of our approach on several important and well-known inverse problems such as Gaussian sensing, inpainting, and super-resolution, showing that CS-DODIP outperforms similar data-free approaches.

As future work, we would like to characterize conditions on the forward process $\mathcal{A}$, signal x^*, and noise s for which recovery of x^* is possible under CS-DODIP, similar to incoherence conditions in the sparse compressive sensing framework. Preliminary evidence suggests that a frequency based analysis may be fruitful.

References

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

[2] Jérôme Bobin, Jean-Luc Starck, and Roland Ottensamer. Compressed sensing in astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):718–726, 2008.

[3] Ashok Veeraraghavan, Dikpal Reddy, and Ramesh Raskar. Coded strobing photography: Compressive sensing of high speed periodic videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):671–686, 2010.

[4] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[5] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[6] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.

[7] Chong You, Zhihui Zhu, Qing Qu, and Yi Ma. Robust recovery via implicit bias of discrepant learning rates for double over-parameterization. *Advances in Neural Information Processing Systems*, 33:17733–17744, 2020.

[8] Bao-Lei Liu, Zhao-Hua Yang, Xia Liu, and Ling-An Wu. Coloured computational imaging with single-pixel detectors based on a 2d discrete cosine transform. *Journal of Modern Optics*, 64(3):259–264, 2017.

[9] Mário AT Figueiredo and Robert D Nowak. A bound optimization approach to wavelet-based image deconvolution. In *IEEE International Conference on Image Processing 2005*, volume 2, pages II–782. IEEE, 2005.

[10] 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.

[11] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[12] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

[13] Guoshen Yu, Guillermo Sapiro, and Stéphane Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing*, 21(5):2481–2499, 2011.

[14] Geng Ji, Michael C Hughes, and Erik B Sudderth. From patches to natural images via hierarchical dirichlet processes.

[15] Ali Mousavi, Ankit B Patel, and Richard G Baraniuk. A deep learning approach to structured signal recovery. In *2015 53rd annual allerton conference on communication, control, and computing (Allerton)*, pages 1336–1343. IEEE, 2015.

[16] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[17] Jong Chul Ye, Yoseob Han, and Eunju Cha. Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM Journal on Imaging Sciences*, 11(2):991–1048, 2018.

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

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[20] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International Conference on Machine Learning*, pages 537–546. PMLR, 2017.

[21] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.

[22] Dave Van Veen, Ajil Jalal, Mahdi Soltanolkotabi, Eric Price, Sriram Vishwanath, and Alexandros G Dimakis. Compressed sensing with deep image prior and learned regularization. *arXiv preprint arXiv:1806.06438*, 2018.

[23] Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Implicit regularization in matrix factorization. *Advances in Neural Information Processing Systems*, 30, 2017.