ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10890318

Some Intriguing Observations on the Learnt Matrices in Deep Unfolded Networks

Kartheek Kumar Reddy Nareddy †§ , Inbasekaran Perumal ‡, Chandra Sekhar Seelamantula †
† Department of Electrical Engineering, Indian Institute of Science, Bengaluru-560012
§ Institute of Data Science, German Aerospace Center, 07745 Jena, Germany
‡ Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Mangaluru-575025
Email: {nareddyreddy@iisc.ac.in, kartheek.nareddy@dlr.de}, inba2002.p@gmail.com, css@iisc.ac.in

***Abstract*—Deep-unfolded networks (DUNs) have set new performance benchmarks in fields such as compressed sensing, image restoration, and wireless communications. DUNs are built from conventional iterative algorithms, where an iteration is transformed into a layer/block of a network with learnable parameters. Despite their huge success, the reasons behind their superior performance over their iterative counterparts are not fully understood. This paper focuses on enhancing the explainability of DUNs by investigating potential reasons behind their superior performance over traditional iterative methods. We concentrate on the Learnt Iterative Shrinkage-Thresholding Algorithm (LISTA), a foundational contribution that achieves sparse recovery with significantly fewer layers than its iterative counterpart, ISTA. Our findings reveal that the learnt matrices in LISTA always have Gaussian distributed entries regardless of whether the sensing matrix is random Gaussian, Bernoulli, exponential, or uniform. The findings also show that the singular values of the learnt matrices exceed unity, despite which, the reconstruction scheme is stable. We conjecture that the activation function may have a role to play in ensuring stability. We also present an unbiasing technique that substantially improves the sparse recovery performance by reestimating the amplitudes based on the converged support.**

***Index Terms*—deep-unfolded networks, sparse signal recovery, iterative shrinkage-thresholding algorithm (ISTA), learnt ISTA (LISTA).**

I. Introduction

Sparse signal recovery addresses the challenge of reconstructing a sparse vector $\boldsymbol{x} \in \mathbb{R}^n$ from its lower-dimensional projection $\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{w}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ ($m \ll n$) is the sensing matrix and $\boldsymbol{w}$ denotes additive measurement noise [1]–[3]. This under-determined problem, foundational to compressed sensing, relies on sparsity for signals reconstruction [4], [5]. Advancements in convex optimization have greatly enabled sparse recovery methods [6]–[11]. Applications are numerous, spanning medical imaging [12], [13], signal/image processing [14], optical-coherence tomography [15], [16], snapshot imaging [17], compressed image recovery [18], magnetic resonance imaging [19], seismic imaging [20]–[23], unlimited sampling [24], radar imaging [25]–[28], and single-pixel cameras [29].

An unconstrained optimization objective for solving the sparse recovery problem, where the signal is sparse in canonical basis, is given as

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{x}\|_1, \tag{1}$$

where λ balances data fidelity with sparsity. Efficient solutions to this problem include proximal gradient methods like the iterative shrinkage thresholding algorithm (ISTA) [7], its accelerated version FISTA [30], and approximate message passing (AMP) [31]. Other notable algorithms for various optimization tasks include the Dantzig selector [32], orthogonal matching pursuit [33], CoSAMP [34], and LASSO [6]. The $\|\boldsymbol{x}\|_1$ penalty in these methods encourages sparsity but is known to introduce bias in the amplitude estimates [1], [30]. Non-convex penalties like minimax-concave penalty (MCP) [35] and smoothly clipped absolute deviation (SCAD) penalty [36] have shown higher accuracy in support and amplitude estimates [35]–[43].

Deep Unfolded Networks (DUNs) integrate iterative algorithms with deep learning, turning iterations into learnable layers to improve efficiency and accuracy [43]–[52]. ISTA is frequently employed to solve the sparse recovery problem (1) using a proximal gradient method, starting with the initialization $\boldsymbol{x}^{(0)} = \mathbf{A}^\top \boldsymbol{y}$, a step-size parameter $\eta = 1/\|\mathbf{A}\|_2^2$, and a threshold parameter λ. The update rule is as follows:

$$\boldsymbol{x}^{(k+1)} = \mathcal{T}_{\eta\lambda}\left(\boldsymbol{x}^{(k)} - \eta\mathbf{A}^\top\left(\mathbf{A}\boldsymbol{x}^{(k)} - \boldsymbol{y}\right)\right), \tag{2}$$

where $\mathcal{T}_{\eta\lambda}$ is the shrinkage function given by

$$\mathcal{T}_{\eta\lambda}(\boldsymbol{x}) = \text{sgn}(\boldsymbol{x}) \cdot \max(|\boldsymbol{x}| - \eta\lambda\mathbf{1}, \mathbf{0}). \tag{3}$$

which is applied element-wise. Null vector is denoted by $\mathbf{0}$ and the vector of all ones by $\mathbf{1}$.

LISTA is an unfolded version of ISTA with l_{max} layers, incorporating weights $\mathbf{S}$ and $\mathbf{W}$. The activation function $\mathcal{T}_{\eta^{(l)}\lambda^{(l)}}$ serves as the proximal operator for the sparsity-promoting ℓ_1 penalty, with learnable parameters $\{\lambda^{(l)}, \eta^{(l)}\}$. The update step is given below

$$\begin{aligned} \boldsymbol{z}^{(k+1)} &= \mathbf{S}\boldsymbol{x}^{(k)} + \mathbf{W}\boldsymbol{y}, \\ \boldsymbol{x}^{(k+1)} &= \mathcal{T}_{\eta^{(k)}\lambda^{(k)}}(\boldsymbol{z}^{(k)}). \end{aligned} \tag{4}$$

LISTA is effectively a recurrent neural network with learnable network weights $\mathbf{S}$, $\mathbf{W}$, and the nonlinearity $\mathcal{T}_{\eta^{(k)}\lambda^{(k)}}$. The network weights are typically shared between the layers. The network has a finite number of layers (L), and the output from the last layer ($\boldsymbol{x}^{(L)}$) is considered as the reconstructed signal. The network is trained on squared error (quadratic loss) $\|\boldsymbol{x}^{(L)} - \boldsymbol{x}^*\|_2^2$ [44]. Observations from [44] have shown that the number of iterations required for accurately reconstructing the sparse signal $\boldsymbol{x}^*$ has reduced from several hundred iterations in the case of ISTA to about a dozen iterations/layers in the case of LISTA. Moreover, the reconstruction accuracy has improved with LISTA, although it requires more extensive training [44].

A. Contributions

This paper focuses on the Learnt Iterative Shrinkage-Thresholding Algorithm (LISTA), exploring its improved efficiency and effectiveness over traditional ISTA. We analyze the singular value distribution of the learnt matrices in LISTA vs. ISTA. We also show that the distributions of the learnt weights is Gaussian regardless of the distribution of the entries of the sensing matrix. We also perform unbiasing by recomputing the amplitudes using least-squares regression on the estimated support.

This work is funded by the Qualcomm Innovation Fellowship and the Science and Engineering Research Board (SERB) Core Research Grant.

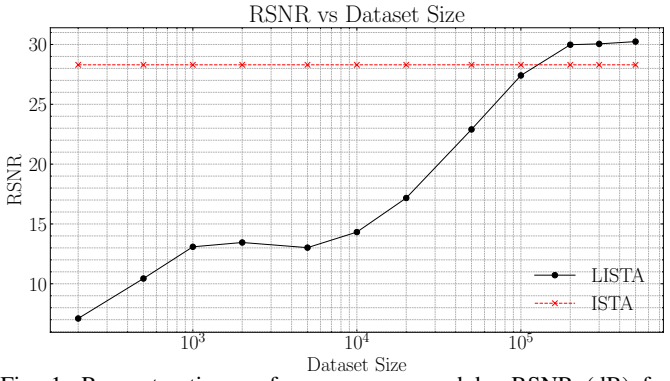


Fig. 1: Reconstruction performance measured by RSNR (dB) for LISTA and ISTA models across various dataset sizes. RSNR improves with larger datasets for LISTA.

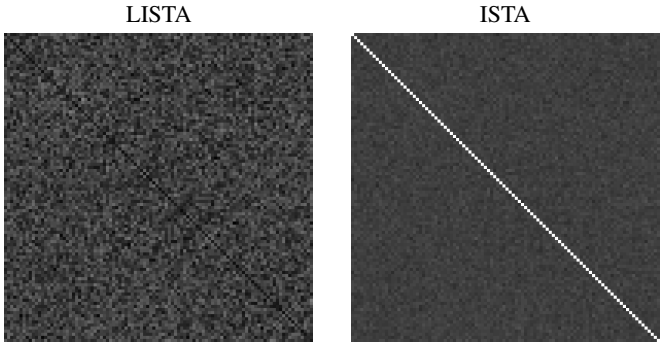


Fig. 2: Comparative visualization of learnt $\mathbf{S}$ weights from LISTA and conventional ISTA.

II. Experimental Setup

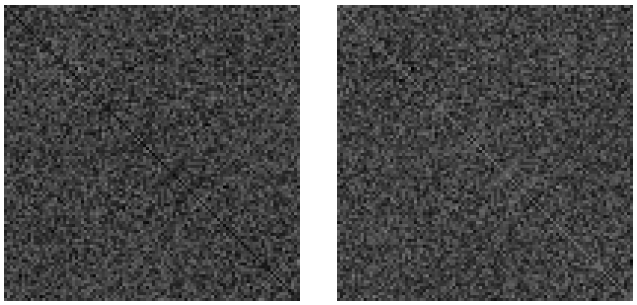
The support of the ground-truth sparse signal $\boldsymbol{x}^* \in \mathbb{R}^n$ is selected from a uniform distribution, and the amplitudes from a normal distribution. The sparsity factor denoted by ρ is defined as $\rho = \|\boldsymbol{x}^*\|_0/n$. The sensing matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ has i.i.d. normally distributed elements, with $n = 100$ and $m = 70$. The measurement vector $\boldsymbol{y} \in \mathbb{R}^m$ is obtained as $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}^* + \sigma\,\boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$ and σ is the desired standard deviation. Sparse recovery performance is measured by the reconstruction signal-to-noise ratio (RSNR), defined as $\text{RSNR} = 10\log_{10}\left(\frac{\|\boldsymbol{x}^*\|_2^2}{\|\hat{\boldsymbol{x}}-\boldsymbol{x}^*\|_2^2}\right)$ dB, where $\hat{\boldsymbol{x}}$ is the estimated sparse vector. RSNR is averaged over 1,000 test samples. LISTA has 15 layers, with 17,030 learnable parameters, trained for 100 epochs using the Adam optimizer (learning rate 10^{-4}, halved at epochs 50 and 75) with mean-square error (MSE) loss and 100,000 training samples. σ is chosen to set the measurement SNR at 40 dB, and sparsity factor ρ is set to 0.1.

We varied the training dataset size and recorded the network performance. Since ISTA is an iterative method, it does not require training data, and its performance remains constant across dataset sizes. As shown in Figure 1, LISTA needs a large dataset for training, around 100,000 training samples to outperform ISTA.

III. A Closer Look at the Learnt Parameters

This section analyzes the learnt network weights of LISTA and compares them with those from ISTA, along with ablation studies highlighting key properties.

The updates for ISTA and LISTA can be expressed as $\boldsymbol{x}^{(k+1)} = \mathcal{T}_{\eta^{(k)}\lambda^{(k)}}\left(\mathbf{S}\boldsymbol{x}^{(k)} + \mathbf{W}\boldsymbol{y}\right)$, where $\mathbf{S} = \mathbf{I} - \eta\mathbf{A}^\top\mathbf{A}$ and $\mathbf{W} = \eta\mathbf{A}^\top$



(a) (b)

Fig. 3: Comparison of learnt $\mathbf{S}$ matrices from LISTA: (a) original and (b) sign-inverted. The inversion reveals a diagonal dominance trend similar to ISTA, suggesting improved regularization and stability.

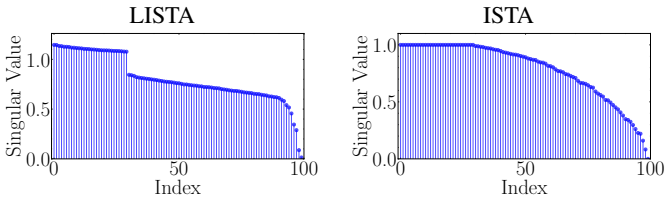


Fig. 4: Singular values of learnt $\mathbf{S}$ matrices from LISTA and ISTA.

in the case of ISTA, whereas they are learnt in the case of LISTA. Figure 2 compares the $\mathbf{S}$ matrices. ISTA shows diagonal dominance, as expected from its $\mathbf{S} = \mathbf{I} - \eta\mathbf{A}^\top\mathbf{A}$ structure and LISTA stands out by reversing diagonal dominance with negative diagonal values.

Figure 3 shows the original and sign-inverted $\mathbf{S}$ matrices in LISTA. Sign inversion, taking the negative of the values for the purpose of display, shows that the learning process suppresses the diagonal dominance in the learnt $\mathbf{S}$ matrix.

Figure 4 shows the singular value distributions of the $\mathbf{S}$ matrices. In the case of ISTA, the singular values are below 1, indicating stability in iterations. By contrast, some of the dominant singular values of the learnt $\mathbf{S}$ matrix in LISTA are greater than 1, which indicates potential instability with iterations. Despite this, LISTA achieves stable convergence in practice, primarily due to the non-expansive property of the shrinkage-thresholding function (cf. Eq. 3).

Figure 5 illustrates the learnt threshold values in LISTA for networks with 15 and 100 layers. As the network depth increases, we observe a decreasing trend in threshold values, indicating a saturation effect in deeper architectures. This behavior suggests that LISTA adapts its thresholding strategy based on network depth, potentially optimizing the trade-off between signal amplification and stability. The relatively high threshold values learnt by LISTA, particularly in the first few layers, help maintain stability.

We examined the role of the shrinkage-threshold activation function in LISTA's stability by examining the squared ℓ^2 norm of the reconstructed signal $\hat{\boldsymbol{x}}$ across 15, 100, 500, and 1000 layers while fixing the $\mathbf{S}$ and $\mathbf{W}$ matrices, which are obtained from training a 15-layer LISTA. With the activation function, $\|\hat{\boldsymbol{x}}\|_2^2$ remains nearly constant, indicating stability (cf. Figure 6). On the other hand, removing the activation function leads to an exponential increase in the norm, clearly indicating instability as the number of layers increases (cf. Figure 7).

We examine how these matrices evolve during training. The initialization of $\mathbf{S}$ matrix is inspired from ISTA and does not have Gaussian distributed entries. However, upon training, the distribution

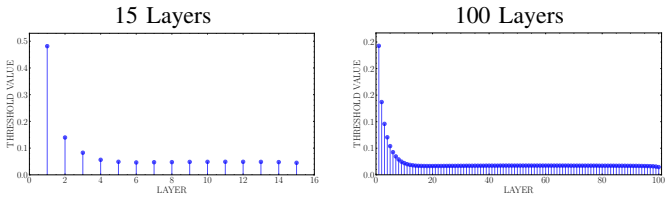


Fig. 5: ℓ_1 threshold values in LISTA across layers. The threshold decreases as the number of layers increase, demonstrating a saturation effect in deeper networks.

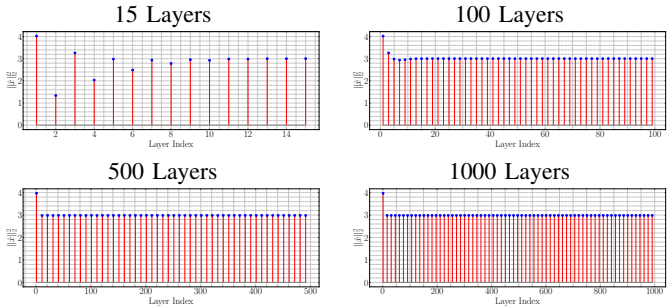


Fig. 6: Stability of LISTA (activation included) for different choices of layers. It shows the squared ℓ_2 norm $\|\hat{x}\|_2^2$ for 15, 100, 500, and 1000 layers, indicating stability, and reflecting robustness with shrinkage-threshold activation.

settles to a Gaussian (cf. Figure 9). Figure 10 shows how the distribution of the entries of $\mathbf{W}$ evolves during training. While the distribution remains Gaussian, the variance increases with epochs.

We next analyze the effect of the choice of the sensing matrix on the distributions of the entries of the learnt $\mathbf{S}$ and $\mathbf{W}$ matrices. Figure 8 reveals that the entries are nearly Gaussian distributed with the $\mathbf{S}$ matrix having a narrower distribution than $\mathbf{W}$ across all choices of the sensing matrix. The empirical convergence of the distribution to a Gaussian regardless of the choice of the distribution of the sensing matrix is rather interesting. While Gaussian matrices (rectangular) serve as good sensing matrices in view of their *restricted isometry property*, it is intriguing that the learnt $\mathbf{S}$ matrix, which is square, is also Gaussian distributed. Random Gaussian matrices are known to have full rank with probability one [53]. It is also quite likely that the Central Limit Theorem [54] may be at play.

Based on the preceding observations, we consider the potential advantages of a Gaussian initialization for $\mathbf{S}$ and $\mathbf{W}$ matrices for LISTA. Figure 11 compares the reconstruction accuracy with ISTA initialization and random Gaussian initialization. While Gaussian initialization initially yields a higher reconstruction loss, it leads to faster convergence than ISTA-based initialization. This trend is consistent across both training and validation losses, indicating that Gaussian initialization improves training efficiency. This trend was found to be consistent for various initializations as confirmed by the error bars in Figure 11.

IV. Unbiasing the Amplitude Estimates

The shrinkage-thresholding function introduces bias in the reconstructions [35]. The bias occurs because ℓ_1 regularization tends to shrink even the large-valued coefficients, leading to underestimation of amplitudes. To address this issue, we propose an *unbiasing approach* that refines the amplitude estimates obtained from a DUN or its iterative counterpart.

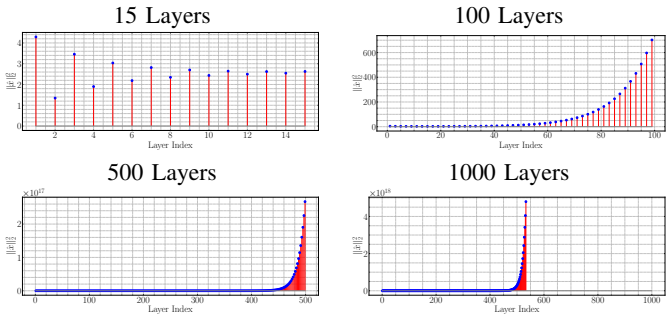


Fig. 7: Impact of removing the shrinkage-threshold activation in LISTA. The plots track the squared ℓ_2 norm $\|\hat{x}\|_2^2$ without activation across 15, 100, 500, and 1000 layers, demonstrating significant increases in the norm, especially for higher number of layers, which suggests instability.

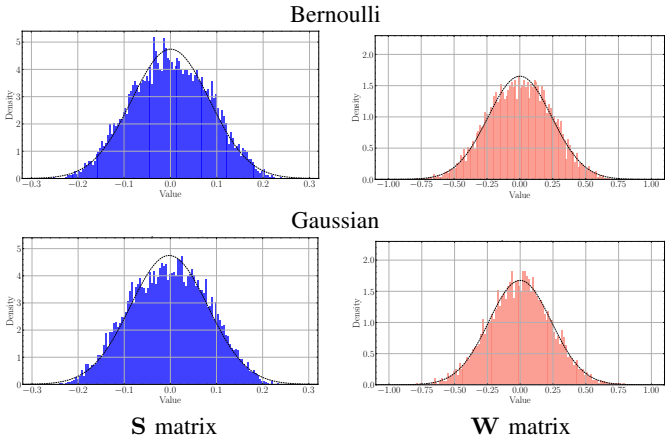


Fig. 8: Histograms of the entries of the learnt matrices in LISTA and corresponding Gaussian fits, demonstrating close adherence to a Gaussian distribution.

The crucial problem in sparse signal recovery is support estimation. Given the support, the problem of amplitude recovery is relatively simpler because it can be performed using least-squares regression. We make use of this observation to refine and unbias the amplitudes estimated by LISTA or ISTA.

Initially, we estimate the support set $\mathcal{S}$, which includes the indices where the sparse signal is non-zero, using the final reconstruction $\hat{\boldsymbol{x}}$ obtained by methods such as ISTA and LISTA. Next, we select the columns of the sensing matrix $\mathbf{A}$ corresponding to $\mathcal{S}$ to form a submatrix $\mathbf{A}_{\mathcal{S}}$. Unbiased amplitude estimates are then calculated through least-squares regression by computing the Moore-Penrose pseudo-inverse of $\mathbf{A}_{\mathcal{S}}$ and multiplying it by the measurement vector $\boldsymbol{y}$. The refined, unbiased estimate of the sparse signal, $\hat{\boldsymbol{x}}_{\mathrm{u}}$, is defined as:

$$\hat{\boldsymbol{x}}_{\mathrm{u},\mathcal{S}} = \mathbf{A}_{\mathcal{S}}^{\dagger}\boldsymbol{y},$$

where $\mathbf{A}_{\mathcal{S}}^{\dagger}$ denotes the pseudo-inverse of $\mathbf{A}_{\mathcal{S}}$. The entries of $\hat{\boldsymbol{x}}_{\mathrm{u}}$ not in $\mathcal{S}$ are set to zero. By doing so, we mitigate the bias introduced by shrinkage-thresholding and improve the accuracy of the reconstructed signal. This method is particularly effective in cases where the initial support estimate is accurate, allowing the least-squares step to refine the amplitude estimates, with a modest overhead of computing the pseudo-inverse.

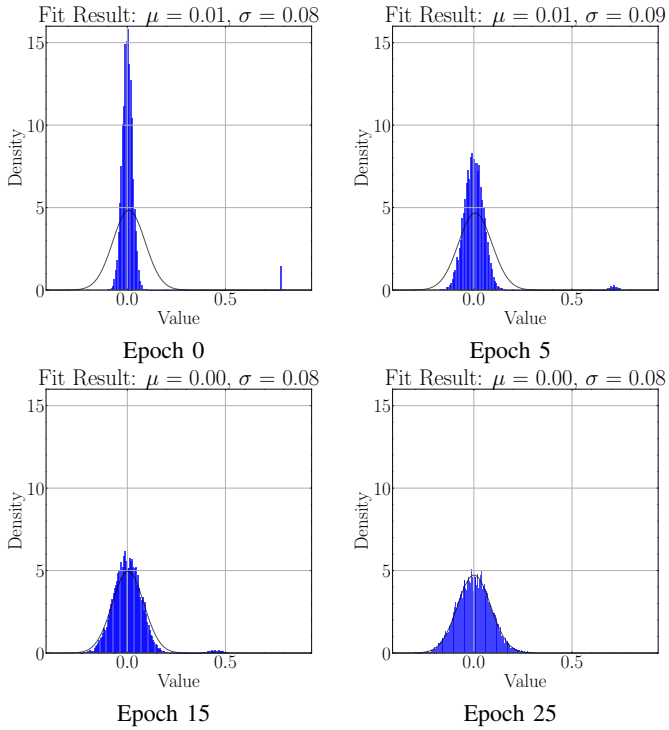


Fig. 9: Evolution of the entries of the $\mathbf{S}$ matrix over training epochs. The weights are nearly Gaussian distributed. The sample mean and sample variance are also indicated.

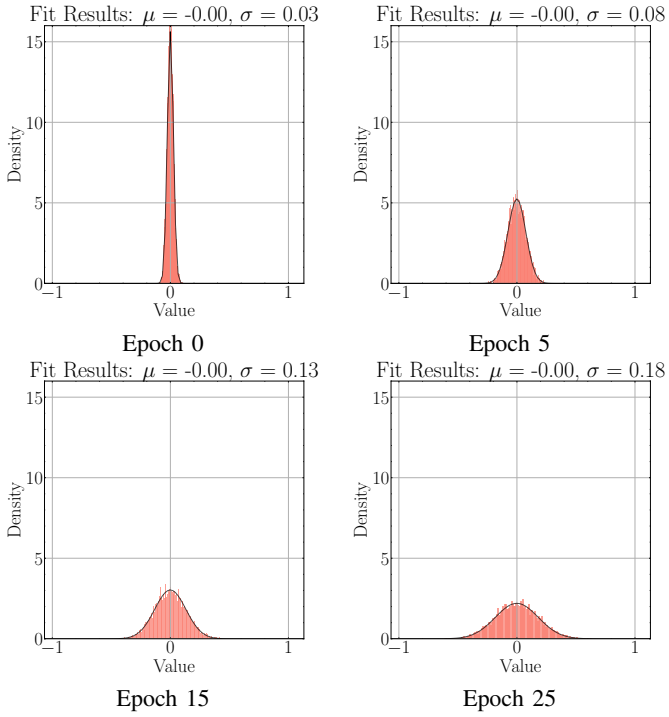


Fig. 10: Evolution of the distribution of the entries of $\mathbf{W}$ over training epochs. The Gaussian distribution is an accurate fit to the distribution of the learnt weights.

TABLE I: Effect of unbiasing on sparse recovery performance for different measurement SNRs. Prefix U- denotes unbiased variant. The performance is measured in terms of reconstruction SNR (RSNR). Values represent mean RSNR $\pm$ standard deviation.

Technique	250 dB	50 dB	40 dB	30 dB
ISTA	44.5 $\pm$ 4.3	43.5 $\pm$ 4.3	37.8 $\pm$ 2.7	28.2 $\pm$ 2.5
LISTA	30.4 $\pm$ 2.3	30.5 $\pm$ 2.3	30.1 $\pm$ 2.3	26.4 $\pm$ 3.0
U-ISTA	238.3 $\pm$ 55.5	54.3 $\pm$ 3.8	41.1 $\pm$ 6.3	32.2 $\pm$ 3.7
U-LISTA	115.7 $\pm$ 52.3	51.4 $\pm$ 7.7	44.0 $\pm$ 4.2	32.6 $\pm$ 4.1

We evaluate the effect of amplitude unbiasing on ISTA and LISTA. The unbiased variant is prefixed with 'U-'. With the sparsity parameter set to $\rho = 0.1$, we tested noise levels of 250 dB (practically noiseless scenario), 50 dB, 40 dB and 30 dB. The observed findings are tabulated in Table I. Our findings highlight consistent improvement in reconstruction quality with the unbiasing technique across all methods and noise levels. Notably, at lower noise levels (250 dB and 50 dB), they significantly outperform their biased counterparts. This observation underscores the potential for simpler models, equipped with unbiasing techniques, to achieve superior reconstruction fidelity over more complex models. The unbiasing technique relies on accurate support estimation. Better the accuracy of support estimation, better the performance of the unbiasing technique.

V. Conclusions

We explored the transition from the Iterative Shrinkage-Thresholding Algorithm (ISTA) to its data-driven counterpart, LISTA, detailing the distinctions between the two methods by analyzing the distributions of weights and singular values. Notably, the singular values of LISTA's learnt matrices exceed one. However, the potential instability appears to be mitigated by the higher threshold values of the nonlinearity, suggesting a balancing effect that could explain LISTA's enhanced performance over ISTA. The transition of $\mathbf{S}$ and $\mathbf{W}$ matrices over the course of training, and convergence to a Gaussian distribution inspired a new initialization scheme for LISTA based on random weights from Gaussian distribution, which resulted in faster convergence over the conventional initialization based on ISTA. We also proposed an unbiasing technique to further improve the reconstruction accuracy obtained from ISTA and LISTA. The unbiased variants outperformed their biased counterparts significantly. The unbiasing technique is applicable to other sparse recovery techniques as well. One could deploy the two key contributions of this paper — Gaussian weight distribution and unbiasing technique — to other inverse problems in computational imaging such as deconvolution, compressed image recovery, super-resolution, etc., where deep-unfolded networks have a crucial role to play.

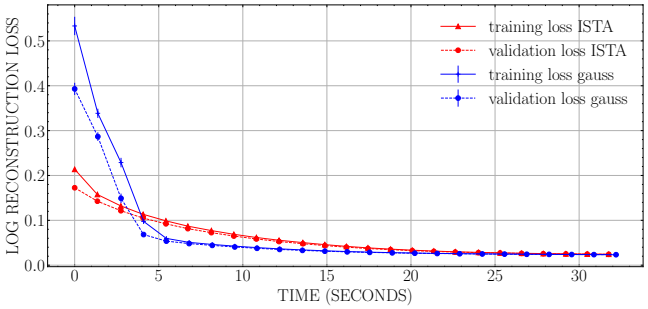


Fig. 11: Comparison of ISTA and Gaussian weight initialization for LISTA. Solid lines represent training loss, dotted lines show validation loss, averaged over 25 random seeds.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[3] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, 2007.

[4] G. Kutyniok, "Theory and applications of compressed sensing," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 79–101, 2013.

[5] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[6] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Royal Statistical Society (Series B)*, 1996.

[7] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, pp. 183–202, 01 2009.

[9] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM J. Imaging Sci.*, vol. 1, no. 1, pp. 143–168, 2008.

[10] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 253–276, 2010.

[11] J.-F. Aujol and C. Dossal, "Stability of over-relaxations for the forward-backward algorithm, application to fista," *SIAM J. Optim.*, vol. 25, no. 4, pp. 2408–2433, 2015.

[12] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," *Adv. Neural. Inf. Process. Syst.*, vol. 29, 2016.

[13] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Modl: Model-based deep learning architecture for inverse problems," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 394–405, 2018.

[14] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Process.* Springer, 1st ed., 2010.

[15] S. Mukherjee and C. S. Seelamantula, "Fienup algorithm with sparsity constraints: Application to frequency-domain optical-coherence tomography," *IEEE Trans. on Sig. Proces.*, vol. 62, no. 18, 2014.

[16] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, and D. Psaltis, "Optical tomographic image reconstruction based on beam propagation and sparse regularization," *IEEE Trans. Comput. Imaging*, vol. 2, no. 1, pp. 59–70, 2016.

[17] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 65–88, 2021.

[18] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3618–3632, 2014.

[19] W. Lu and N. Vaswani, "Modified compressive sensing for real-time dynamic MR imaging," in *Proc. IEEE Intl. Conf. on Image Proces. (ICIP)*, pp. 3045–3048, 2009.

[20] S. Mache, P. K. Pokala, K. Rajendran, and C. S. Seelamantula, "DuRIN: A deep-unfolded sparse seismic reflectivity inversion network," *CoRR*, vol. abs/2104.04704v2, 2021.

[21] S. Mache, P. K. Pokala, K. Rajendran, and C. S. Seelamantula, "*Nu*SPAN: A proximal average network for nonuniform sparse model — application to seismic reflectivity inversion," *CoRR*, vol. abs/2105.00003v2, 2021.

[22] S. Yuan and S. Wang, "Spectral sparse Bayesian learning reflectivity inversion," *Geophys. Prospect.*, vol. 61, no. 4, pp. 735–746, 2013.

[23] S. Mache, P. K. Pokala, K. Rajendran, and C. S. Seelamantula, "Introducing nonuniform sparse proximal averaging network for seismic reflectivity inversion," *IEEE Trans. on Comput. Imaging*, pp. 1–14, 2023.

[24] S. Rudresh, A. Adiga, B. A. Shenoy, and C. S. Seelamantula, "Wavelet-based reconstruction for unlimited sampling," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 4584–4588, 2018.

[25] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. on Sig. Proces.*, vol. 57, no. 6, 2009.

[26] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proc. IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.

[27] A. De Maio, Y. C. Eldar, and A. M. Haimovich, *Compressed sensing in radar signal Process.* Cambridge University Press, 2019.

[28] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for mimo radar," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 419–430, 2014.

[29] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[30] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 693–696, 2009.

[31] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.

[32] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *The Ann. Stat.*, vol. 35, no. 6, pp. 2313 – 2351, 2007.

[33] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[34] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[35] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Ann. Stat.*, vol. 38, no. 2, pp. 894–942, 2010.

[36] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, pp. 1348–1360, 02 2001.

[37] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. on Sig. Proces.*, vol. 65, no. 17, pp. 4481–4494, 2017.

[38] J. Woodworth and R. Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Inverse Problems*, vol. 32, no. 7, 2016.

[39] K. Bredies and D. A. Lorenz, "Iterated hard shrinkage for minimization problems with sparsity constraints," *SIAM Journ. on Scient. Comput.*, vol. 30, no. 2, pp. 657–683, 2008.

[40] S. Foucart and M. J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, 2009.

[41] E. Chouzenoux, A. Jezierska, J. C. Pesquet, and H. Talbot, "A Majorize-Minimize subspace approach for $\ell_2 - \ell_0$ image regularization," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 563–591, 2013.

[42] P. Yin, Y. Lou, Q. He, and J. Xin, "Minimization of ℓ_{1-2} for compressed sensing," *SIAM J. Sci. Comput.*, vol. 37, no. 1, pp. A536–A563, 2015.

[43] P. K. Pokala, A. G. Mahurkar, and C. S. Seelamantula, "FirmNet: A sparsity amplified deep network for solving linear inverse problems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 2982–2986, 2019.

[44] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn. (ICML)*, p. 399–406, 2010.

[45] J. Liu and X. Chen, "ALISTA: Analytic weights are as good as learned weights in LISTA," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019.

[46] X. Chen, J. Liu, Z. Wang, and W. Yin, "Theoretical linear convergence of unfolded ista and its practical weights and thresholds," *Advances in Neural Information Process. Systems (NuerIPS)*, vol. 31, 2018.

[47] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1828–1837, 2018.

[48] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Trans. Signal Process.*, vol. 65, pp. 4293–4308, Aug 2017.

[49] D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," *IEEE Trans. Signal Process.*, vol. 67, pp. 3113–3125, 2019.

[50] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, no. 1, pp. 11–29, 2020.

[51] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, S. van, J. G. Ruud, and Y. C. Eldar, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Trans. Med. Imaging*, vol. 39, pp. 1051–1063, 2020.

[52] G. Kutyniok and Y. C. Eldar, *Compressed Sensing: Theory and Applications.* Cambridge, U.K.: Cambridge Univ. Press, 2012.

[53] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis, Springer New York, 2013.

[54] P. Billingsley, *Probability and measure*. A Wiley-Interscience publication, Wiley, 3. ed ed., 1995.