Theoretical linear convergence of deep unfolding network for block-sparse signal recovery

Rong Fu, Yimin Liu, Xiuhong Li

*Shanghai AI Laboratory, Shanghai, China; bDepartment of Electronic Engineering, Tsinghua University, Beijing, China

ABSTRACT

In this paper, we consider the recovery of the high-dimensional block-sparse signal from a compressed set of measurements, where the nonzero coefficients of the recovered signal occur in several blocks. Adopting the idea of deep unfolding, we explore the block-sparse structure and put forward a block-sparse reconstruction network named Ada-BlockLISTA, which performs gradient descent on every single block followed by a block-wise shrinkage. Furthermore, we prove the linear convergence rate of our proposed network, which also theoretically guarantees exact recovery for a potentially higher sparsity level based on the underlying block structure. Numerical results indicate that Ada-BlockLISTA yields better signal recovery performance than existing algorithms, which ignore the additional block structure in the signal model.

Keywords: Deep unfolding network, block-sparse, linear convergence rate

1. INTRODUCTION

As it is crucial to minimize the sample size required in estimation problems, compressive sensing (CS) plays an indispensable role in the field of signal processing. Many existing problems can be formulated in terms of compressed sensing (CS), including sparse channel estimation, beam pattern synthesis, direction-of-arrival (DOA) estimation, and range-Doppler estimation. The recovery is completed by finding a sparse representation of the time-domain signal over a dictionary matrix, which is always a row-sampled discrete Fourier transform (DFT) matrix.

Since this kind of problem is ill-posed, various methods have been proposed. One of the well-known $\ell_1$-norm regularization techniques is the iterative shrinkage thresholding algorithm (ISTA), which is often used in a wide range of applications. Since ISTA is solving a convex optimization problem, it is guaranteed to converge to a solution under the correct circumstances. As an iterative solver, it often costs too much time for ISTA to converge, thus difficult to apply in various real-time applications. Some pre-defined optimization parameters such as the step size and regularization parameter are also required to be set carefully based on prior knowledge, which may become quite challenging in some cases.

To address this problem, Gregor and LeCun have proposed a recurrent neural network (RNN) to solve sparse coding problems, named learned ISTA (LISTA). Based on the iterative structure of ISTA, the authors free the traditional parameters in ISTA to data-driven variables and unfold ISTA algorithms into a RNN structure. LISTA networks show improved performance in terms of convergence speed in both the theoretical analysis and empirical results. Many works have modified LISTA to increase its adaptability by embedding the dictionary in the network architecture. Reference couples the learned matrices in LISTA by embedding the dictionary matrix into the network structure. Based on the work of Reference, robust ALISTA is proposed in Reference, which explicitly calculates the learned matrices by solving a coherence minimization problem. Another adaptive variation of LISTA, named Adaptive-LISTA (AdaLISTA), generalizes the application of robust-ALISTA to various model scenarios. Beyond enjoying accelerated convergence speed, AdaLISTA can serve different dictionaries using the same weight matrix without retraining the whole network.

However, all these LISTA-type networks are sensitive to the signal sparsity level. While increasing the number of nonzero elements in the sparse signal of interest, LISTA and its variants will suffer decreased signal recovery accuracy. In practical applications, the recovered signal may not be sufficiently sparse, such as block-sparse problems which recover an unknown, block-sparse signal, where the nonzero elements are distributed in blocks. Based on AdaLISTA, a block-sparse reconstruction network, named Ada-BlockLISTA, has been proposed in Reference, which makes good use of the natural
structure of a block-sparse signal and can recover its nonzero blocks correctly, while original AdaLISTA fails, especially with many nonzero blocks.

In this paper, we formulate the signal model with block structure and introduce relevant iterative algorithms for solving the block-sparse recovery problem in Section 2. By exploring the block structure of the signal model, we introduce AdaBlockLISTA network in Section 3, and further analyze its convergency rate as well as support recovery guarantee. In numerical simulations, we demonstrate its convergence acceleration and a significant advantage in recovery performance.

2. SYSTEM MODEL AND TRADITIONAL ALGORITHMS

In this section, we introduce the signal model with block-sparse and review traditional algorithms such as ISTA and LISTA, which leverage the assumption of sparsity in the ground truth signal.

2.1 System model

We aim to recover an unknown, sparse signal from noisy observations taken from a known, under-determined dictionary \( \Phi \in \mathbb{C}^{N \times M} \), with \( N < M \). In this ill-conditioned problem, noisy observation \( y \in \mathbb{C}^N \) can be formulated as

\[
y = \Phi x^* + w
\]

where \( x^* \) is the ground truth, and \( w \) is additive random noise present in the system. Next, we define block sparsity according to Reference\(^{21}\). Suppose that a block-sparse signal \( x^* \) is divided into \( Q \) blocks of length \( P \). If there are at most \( K \) nonzero blocks (\( K < Q \)) in \( x^* \), it is called \( K \)-block-sparse. Sharing the same structure with \( x^* \), the dictionary matrix \( \Phi \) is also divided into \( Q \) blocks. A visualization of the block-sparse model is shown in Figure 1.

![Figure 1. A visualization of the block-sparse signal model where each color in \( \Phi \) and \( \hat{x} \) corresponds to a different block. White corresponds to zero entries in \( \hat{x} \).](image)

2.2 Brief review of traditional algorithms

This subsection briefly reviews traditional algorithms such as ISTA and LISTA, designed to solve sparse problems. In the framework of CS, the \( \ell_1 \)-regularized regression is a well-known technique to harness prior knowledge of sparsity. Standard ISTA solves the \( \ell_1 \)-regularized optimization problem by iteratively performing Lipschitz gradient descent with respect to the measurement error\(^{9,22}\).

While the signal of interest \( x \) possesses a block-sparse structure, many researchers have designed specialized algorithms in principle to utilize the block-sparse structure, such as Block ISTA\(^{20}\), block-OMP\(^{23}\), block-based CoSaMP\(^{24}\) and block-sparse Bayesian learning\(^{25}\). To compute an estimate which maintains a block-sparse structure, the original \( \ell_1 \)-norm optimization problem turns into the following mixed-norm optimization problem as follows,

\[
\hat{x} = \arg \min_x \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda \|x\|_{2,1}
\]
where \( \hat{x} \) is the estimated signal, and \( \lambda \) is a regularization parameter controlling the sparsity penalty characterized by minimizing the \( \ell_{2,1} \) norm.

ISTA demonstrates considerable accuracy in recovering sparse signals but takes too much time to converge\(^{26}\). As an unfolded version of ISTA iterations, LISTA is a RNN containing only \( T \) layers, where each layer corresponds to an ISTA iteration\(^{13}\). To build a more generalized network for various dictionary choices, AdaLISTA, a variant of LISTA, has been proposed to increase the adaptability of the deep unfolding network\(^{19}\). One layer of AdaLISTA is defined as,

\[
X^{(k+1)} = S_{\gamma^{(k)}} \left( X^{(k)} - \lambda^{(k)} W \left( \Phi X^{(k)} - y \right) \right) \quad k = 0, 1, \ldots, T - 1
\]

where \( X^{(k)} \) is the output signal of the \( k \)-th layer, \( W \in \mathbb{C}^{N \times N} \) is the learned weight shared across different layers, while \( \theta^{(k)} \), \( \gamma^{(k)} \) are the learned soft threshold and step size of the \( k \)-th layer in AdaLISTA.

### 3. BLOCK-SPARSE RECOVERY NETWORK AND CONVERGENCE ANALYSIS

Motivated by \( \ell_{2,1} \) minimization methods such as Block-ISTA, we explore the block structure of the signal model and propose our Ada-BlockLISTA network, which is derived from AdaLISTA.

As \( \Phi \) and \( x^* \) share the same block structure, i.e., \( x^* = [x_1^* \ x_2^* \ \ldots \ x_Q^*]^T \) and \( \Phi = [\Phi_1 \ \Phi_2 \ \ldots \ \Phi_Q] \), we learn an individual weight matrix \( \{W_q\}_{q \leq Q} \) for each block and encourage block sparsity for the nonlinear shrinkage function by applying soft thresholding to each block individually. At the \( k \)-th layer of Ada-BlockLISTA, the update rule of the \( q \)-th block in the block-sparse signal is formulated as,

\[
\begin{align*}
z_q^{(k+1)} &= x_q^{(k)} + \gamma_q^{(k)} \Phi_q^H W_q^H \left( y - \sum_{i=1}^Q \Phi_i x_i^{(k)} \right) \\
x_q^{(k+1)} &= z_q^{(k+1)} \left( 1 - \theta_q^{(k)} \|z_q^{(k+1)}\|_2 \right)
\end{align*}
\]

where \( z_q^{(k+1)} \) is the \( q \)-th block of \( z_q^{(k+1)} \) after block-wise gradient descent, and \( x_q^{(k+1)} \) is the \( q \)-th block of the output signal \( x^{(k+1)} \) after block-wise soft thresholding. In Ada-BlockLISTA, we need to learn the individual weight \( W_q \) for the \( q \)-th block, as well as the learned soft threshold \( \theta_q^{(k)} \) and the learned step size \( \gamma_q^{(k)} \).

As shown above, Ada-BlockLISTA consists of two main steps.

- First, we perform block-wise gradient descent and update the value of each block according to the current residual value.

- Then, we use block-wise soft thresholding to force the updated signal \( \left[ z_q^{(k+1)} \right]_{q \leq Q} \) in the previous step to be block-sparse. Figures 2 and 3 illustrate the network structure of both AdaLISTA and Ada-BlockLISTA.
Comparing Figures 2 and 3, we can see that Ada-BlockLISTA updates each block in $x$ independently, while AdaLISTA uses the same weight across different blocks. Furthermore, AdaLISTA and Ada-BlockLISTA use the soft threshold parameters $\{\theta^{(k)}\}_{k \in \mathbb{Z}}$ in different ways. In AdaLISTA, we perform element-wise soft thresholding and set each element in $x^{(k)}$ with an absolute value less than $\theta^{(k)}$ to 0. On the other hand, we use $\theta^{(k)}$ as a block-wise threshold in Ada-BlockLISTA, which is compared with the $\ell^2$ norm of every single block $z_q^{(k+1)}$ rather than the absolute value of elements.

Then we show some main theoretical results, i.e., the linear convergence rate and support recovery guarantee of Ada-BlockLISTA. We generalize the two basic definitions of coherence\(^2\), i.e., the mutual coherence within a block, also referred to as sub-coherence, and block-coherence of the dictionary, which describe its local and global properties, respectively.

Here we illustrate the linear convergence rate of our Ada-BlockLISTA in the following theorem (More details on the convergence analysis of Ada-BlockLISTA can be found in the arXiv article: https://arxiv.org/abs/2111.09801). The proof of Theorem 1 mimics the corresponding proof steps in References\(^{18,19}\) and can be viewed as an extension to the block-sparse case.

**Theorem 1 (Convergence rate of Ada-BlockLISTA).**

We first consider the three quantities,

$$
\tilde{\nu}_1 = \max_{k \geq 0} \max_{l \in \mathbb{Z}, j \in \mathcal{Q}} \left\{ \max_{l \in \mathbb{Z}, j \in \mathcal{Q}} \left\| \frac{\Phi_q^H W_q}{\mu^b} \Phi_{j,l} \right\| \right\}
$$

$$
\bar{\mu}_b = \max_{k \geq 0} \max_{l \in \mathbb{Z}, j \in \mathcal{Q}} \left\{ \max_{l \in \mathbb{Z}, j \in \mathcal{Q}} \left\| \frac{1}{P} \left\| \Phi_q^H \Phi_{j,l} \right\| \right\}
$$

$$
C_w = \max_{k \geq 0} \max_{l \in \mathbb{Z}, j \in \mathcal{Q}} \left\| \left\| \frac{\Phi_q^H W_q}{\mu^b} \right\|_{2,1} \right\ |
$$

Suppose $x^*$ to be sufficiently small such that $s = \|x^*\|_{2,0} \leq \frac{1}{2P} \left( \frac{\bar{\mu}_b^{-1} + P - (P - 1) \tilde{\nu}_1}{\bar{\mu}_b} \right)$, and the soft thresholds are set as $\theta^{(k)} = P \tilde{\nu}_1 \mathrm{sup} \|x^{(k)} - x^*\|_{2,1} + C_w \sigma$ where $\sigma$ is the standard deviation of noise, then Ada-BlockLISTA converges linearly, i.e., its recovered error decreases exponentially with the number of layers.

The above theorem can be interpreted in two aspects:

- Theorem 1 shows that Ada-BlockLISTA can converge at a linear rate under sufficient conditions, which is faster than the original ISTA and FISTA.
Compared to the recovery condition of AdaLISTA [19, Theorem 1], the recovery condition, which is the same as the block-sparse recovery condition of [25, Theorem 3], shows that exploring the block structure in $x^*$ leads to ensuring recovery of $x^*$ with higher sparsity level.

4. NUMERICAL RESULTS

We simulate various experiments with different signal dimensions and the number of nonzero blocks. In our simulation, we set $P = 16$ and $Q = 64$, respectively. We generate observed signals according to Reference 27. In Figures 4 and 5, we compared the convergence speed and recovered the 2D block-sparse signal plane of four methods (ISTA, Block-ISTA, AdaLISTA and our Ada-BlockLISTA) when the block sparsity $K$ is 2.

![Figure 4](image1.png)

Figure 4. The NMSE of four methods in each iteration/layer.

![Figure 5](image2.png)

Figure 5. The reconstructed block sparse signal, which has two nonzero blocks.

The results show that Block ISTA and Ada-BlockLISTA successfully reconstruct two blocks, while ISTA and AdaLISTA fail. It demonstrates that block sparse recovery methods can recover a larger number of blocks than their non-block counterparts, which is also proved in Theorem 1.

5. CONCLUSION

In this work, we considered the block-sparse signal recovery problem, where nonzero entries of recovered signals occur in clusters. A block-sparse reconstruction network, Ada-BlockLISTA, explores the block structure in the signal model. We prove that the unrolled block-sparse reconstruction network enjoys a linear convergence rate and provides a sufficient condition on block-sparsity to guarantee the successful recovery of block-sparse signals. We performed extensive simulations to verify the recovery performance of the proposed network. The numerical results show that Ada-BlockLISTA yields better reconstruction properties than the original AdaLISTA network, especially with a large number of nonzero blocks. We demonstrate that explicit use of block-sparsity in unfolded deep learning networks improves block-sparse recovery performance by conducting theoretical and experimental analyses.
REFERENCES

[1] Xu, W. and Hassibi, B., “Efficient compressive sensing with deterministic guarantees using expander graphs,” 2007 IEEE Information Theory Workshop, 414-419 (2007).
[2] Indyk, P., “Sparse recovery using sparse random matrices,” in: A. Lo’pez-Ortiz (Ed.), [LATIN: Theoretical Informatics], Springer, Berlin, Heidelberg, 157 (2010).
[3] Eldar, Y. C. and Kutyniok, G., [Compressed Sensing: Theory and Applications], Cambridge University Press, (2012).
[4] Berger, C. R., Wang, Z., Huang, J. and Zhou, S., “Application of compressive sensing to sparse channel estimation,” IEEE Communications Magazine 48(11), 164-174 (2010).
[5] Xiong, J. and Wang, W., “Sparse reconstruction-based beampattern synthesis for multi-carrier frequency diverse array antenna,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3395-3398 (2017).
[6] Balakrishnan, R. D. and Kwon, H. M., “A new inverse problem based approach for azimuthal DOA estimation,” GLOBECOM'03 Global Telecommunications Conference, 2187-2191 (2003).
[7] Xenaki, A., Gerstoft, P. and Mosegaard, K., “Compressive beamforming,” The Journal of the Acoustical Society of America 136(1), 260 (2014).
[8] Huang, T. and Liu, Y., “Compressed sensing for a frequency agile radar with performance guarantees,” IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), 1057-1061 (2015).
[9] Beck, A. and Teboulle, M., “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” Siam J. Imaging Sciences, 2(1), 183-202 (2009).
[10] Combettes, P. L. and Wajs, V. R., “Signal recovery by proximal forward-backward splitting,” Multiscale Model Simul. 4(4), 1168-1200 (2006).
[11] Antonello, N., Stella, L., Patrinos, P. and van Waterschoot, T., “Proximal gradient algorithms: Applications in signal processing,” arXiv preprint arXiv:1803.01621, (2018).
[12] Zhang, J. and Ghanem, B., “ISTA-Net: Iterative shrinkage-thresholding algorithm inspired deep network for image compressive sensing,” arXiv:1706.07929, (2017).
[13] Gregor, K. and LeCun, Y., “Learning fast approximations of sparse coding,” International Conference on International Conference on Machine Learning, 399-406 (2010).
[14] Borgerding, M., Schniter, P. and Rangan, S., “AMP-inspired deep networks for sparse linear inverse problems,” IEEE Transactions on Signal Processing 65(16), 4293-4308 (2017).
[15] Borgerding, M. and Schniter, P., “Onsager-corrected deep learning for sparse linear inverse problems,” 2016 IEEE Global Conference on Signal and Information Processing, 227-231 (2016).
[16] Fu, R., Huang, T., Liu, Y. and Eldar, Y. C., “Compressed LISTA exploiting Toeplitz structure,” 2019 IEEE Radar Conference (RadarConf), 1-6 (2019).
[17] Chen, X., Liu, J., Wang, Z. and Yin, W., “Theoretical linear convergence of unfolded ista and its practical weights and thresholds,” in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), [Advances in Neural Information Processing Systems], Curran Associates, Inc., 2018.
[18] Liu, J., Chen, X., Wang, Z. and Yin, W., “ALISTA: Analytic weights are as good as learned weights in LISTA,” International Conference on Learning Representations, (2019).
[19] Aberdam, A., Golts, A. and Elad, M., “Ada-lista: Learned solvers adaptive to varying models,” arXiv:2001.08456, (2020).
[20] Fu, R., Monardo, V., Huang, T. and Liu, Y., “Deep unfolding network for block-sparse signal recovery,” 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2880-2884 (2021).
[21] Eldar, Y. C., Kuppinger, P. and Bolcskei, H., “Block-sparse signals: Uncertainty relations and efficient recovery,” IEEE Transactions on Signal Processing 58(6), 3042-3054 (2010).
[22] Wu, K., Guo, Y., Li, Z. and Zhang, C., “Sparse coding with gated learned ISTA,” International Conference on Learning Representations, (2020).
[23] Eldar, Y. C., Kuppinger, P. and Bolcskei, H., “Block-sparse signals: Uncertainty relations and efficient recovery,” IEEE Transactions on Signal Processing 58(6), 3042-3054 (2010).
[24] Baraniuk, R. G., Çevher, V., Duarte, M. F. and Hegde, C., “Model-based compressive sensing,” IEEE Transactions on Information Theory 56(4), 1982-2001 (2010).
[25] Zhang, Z. and Rao, B. D., “Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning,” IEEE Journal of Selected Topics in Signal Processing 5(5), 912-926 (2011).
[26] Draganic, A., Orovic, I. and Stankovic, S., “On some common compressive sensing recovery algorithms and applications—Review paper,” Facta universitatis-series: Electronics and Energetics 30, 477-510 (2017).
[27] Chi, Y. and Chen, Y., “Compressive two-dimensional harmonic retrieval via atomic norm minimization,” IEEE Transactions on Signal Processing 63(4), 1030-1042 (2015).