Block-Sparse Recovery Network for Two-Dimensional Harmonic Retrieval

Rong Fu, Tianyao Huang, Lei Wang, Yinmin Liu

As a typical signal processing problem, multidimensional harmonic retrieval (MHR) has been adapted to a wide range of applications in signal processing. Block-sparse signals, whose nonzero entries appearing in clusters have received much attention recently. An unfolded network named Ada-BlockLISTA was proposed to recover a block-sparse signal at a small computational cost, which learns an individual weight matrix for each block. However, as the number of network parameters is increasingly associated with the number of blocks, the demand for parameter reduction becomes very significant, especially for large-scale MHR. Based on the dictionary characteristics in two-dimensional (2D) harmonic retrieval problems, we introduce a weight coupling structure to shrink Ada-BlockLISTA, which significantly reduces the number of weights without performance degradation. In simulations, our proposed block-sparse reconstruction network, named AdaBLISTA-CP, shows excellent recovery performance and convergence speed in 2D harmonic retrieval problems.

Introduction: Multidimensional harmonic retrieval, associated with many practical applications including direction-of-arrival (DOA) estimation \cite{1, 2} and range-Doppler estimation \cite{3}, has been extensively studied in the signal processing literature. As it is crucial to minimize the required sample size, a myriad of compressed sensing (CS) methods have been developed, such as iterative shrinkage thresholding algorithm (ISTA).

Many algorithms in principle designed for CS can be extended to solve block-sparse recovery problems according to the partition of blocks, such as ISTA \cite{4}. To solve the \(f_2\)-norm minimization problem in (4), we briefly review Block-ISTA as an extension of ISTA, which iteratively performs the following two steps (block-wise gradient descent and soft-thresholding) for every block \(q \in [1, Q]:\)

\[
\begin{align*}
    z_q^{(t+1)} &= z_q^{(t)} + \frac{1}{\gamma_q^t} \left( y - \Phi^H \left( z_q^{(t)} - \rho_q^t \tau_q^t \right) \right), \\
    x_q^{(t+1)} &= z_q^{(t+1)} - \theta_q^t \left( \| z_q^{(t+1)} \|_1 \right). 
\end{align*}
\]

where \(\gamma\) denotes \(\max(\cdot, 0)\), and threshold \(\theta > 0\) is block-wise soft-thresholding parameter which forces blocks in the updated signal \(z_q^{(t+1)}\) to 0 if its \(f_2\) norm is less than \(\theta\). Block-ISTA demonstrates considerable accuracy in recovering block sparse signals but takes hundreds or thousands of iterations for convergence.

Deep unfolding methods: Given the outburst in application of deep neural networks (DNNs) in CS, LISTA and its variants have been proposed to speed up the rate of convergence by freeing the traditional parameters in ISTA to data-driven variables and unfolding ISTA algorithms into a \(T\)-layer RNN \((T \approx 10)\).

Hence, AdaLISTA can successfully recover \(s\)-sparse signals with exponential convergence under the condition that weights satisfies a low enough mutual coherence \(\mu(W, \Phi) = \max_{i \neq j} |\langle W_i, \Phi_j \rangle| / \min_{i \neq k} \| W_i \|_2 \| \Phi_k \|_2 = 1\), where \(W_i\) is the \(i\)-th columns of \(\Phi\). Furthermore, \cite{5} establishes a minimum-coherence criterion between the desired weights and the dictionary, i.e., the optimal learned matrix \(W\) is ought to approach the infimum of the generalization mutual coherence. Thus, we have \(\Phi^H W^H \Phi \approx I\).

When it comes to block sparse case, a block-sparse reconstruction network \cite{6}, named Ada-BlockLISTA, has been proposed. Motivated by Block-ISTA, Ada-BlockLISTA makes use of block structure in AdaLISTA and thus learns an individual weight matrix \(W_q\) for each \(q\)-th block, whose update rule at the \(t\)-th layer is formulated as

\[
\begin{align*}
    z_q^{(t+1)} &= z_q^{(t)} + \gamma_q^t \left( W_q y - \Phi_q^H \left( z_q^{(t)} - \rho_q^t \tau_q^t \right) \right), \\
    x_q^{(t+1)} &= z_q^{(t+1)} - \theta_q^t \left( \| z_q^{(t+1)} \|_1 \right),
\end{align*}
\]

where \(\{ W_1, \ldots, W_Q \}, \{ \theta(\gamma_q^t) \}_{q=1}^{Q} \) are network parameters to learn. Based on the definition of sub-coherence and block-coherence \cite{7}, a similar condition of each weight matrix \(W_q\) in Ada-BlockLISTA ensuring recovery of block-sparse signals can be developed as

\[
\Phi_q^H W_q \Phi_q \approx I, \quad \Phi_q^H W_q \Phi_q \approx I, \quad q, \neq q'.
\]

Although AdaBlockLISTA shows both rapid convergence speed and great block-sparse recovery performance, it is difficult to train a large neural network with many weight matrices \(\{ W_1, \ldots, W_Q \}\), especially when the number of blocks \(Q\) is large. Therefore, we couple the learned matrices in AdaBlockLISTA and propose a structured network named AdaBLISTA-CP based on the dictionary characteristics in 2D harmonic retrieval problems.

**COUPLED NETWORK FOR 2D HARMONIC RETRIEVAL:** While AdaLISTA only learns a single weight \(W\), AdaBlockLISTA needs to learn a different weight \(W_q\) for the \(q\)-th block, whose number of network parameters is up to \(O(W^2)\), where \(N\) is the number of observation samples. These large matrices consume costly memory/computation resources and require a huge amount of labeled data for training. Motivated

**BRIEF REVIEW OF BLOCK SPARSE RECOVERY:** Signal model: The recovery of block sparse signals involves solving a system of linear equations of the form

\[
y = \Phi x + \omega,
\]

where \(\omega \in \mathbb{C}^M\) is additive random noise in the system and \(x \in \mathbb{C}^M, M = PQ\), is the ground truth, which can be divided into \(Q\) sub-vectors as below

\[
x = [x_1, \ldots, x_{P1}, x_2, \ldots, x_{P2}, \ldots, x_{Q}, \ldots, x_{PQ}]^T.
\]

The vector \(x\) is said to be \(K\)-block-sparse, if there are at most \(K\) nonzero blocks \((K \leq Q)\). Sharing the same nested structure with \(x\), the dictionary matrix \(\Phi\) is also divided into \(Q\) blocks, i.e.,

\[
\Phi = [\phi_1, \ldots, \phi_{P1}, \phi_1, \ldots, \phi_{P2}, \ldots, \phi_1, \ldots, \phi_{PQ}]^T.
\]

**Traditional Iterative Algorithms:** To harness block sparsity, we estimate \(x^*\) by solving the mixed-norm optimization problem,

\[
\min_{x} \frac{1}{2} || y - \Phi x ||_2^2 + \lambda ||x||_2^2,
\]

where \(\lambda\) is a regularization parameter controlling the block sparsity penalty characterized by \(f_2, \lambda\) norm defined as \( ||x||_2^2 = \sum_{q=1}^{Q} ||x_q||_2^2\).

Many algorithms in principle designed for CS can be extended to solve block-sparse recovery problems according to the partition of blocks, such as ISTA \cite{8}. To solve the \(f_2\)-norm minimization problem in (4), we briefly review Block-ISTA as an extension of ISTA, which iteratively performs the following two steps (block-wise gradient descent and soft-thresholding) for every block \(q \in [1, Q]:\)

\[
\begin{align*}
    z_q^{(t+1)} &= x_q^{(t)} + \frac{1}{\gamma_q^t} \left( y - \Phi^H \left( x_q^{(t)} - \rho_q^t \tau_q^t \right) \right), \\
    x_q^{(t+1)} &= z_q^{(t+1)} - \theta_q^t \left( \| z_q^{(t+1)} \|_1 \right),
\end{align*}
\]

where \(\gamma\) denotes \(\max(\cdot, 0)\), and threshold \(\theta > 0\) is block-wise soft-thresholding parameter which forces blocks in the updated signal \(z_q^{(t+1)}\) to 0 if its \(f_2\) norm is less than \(\theta\). Block-ISTA demonstrates considerable accuracy in recovering block sparse signals but takes hundreds or thousands of iterations for convergence.
by the minimum-coherence criterion for block-sparse recovery described in (8), it is possible to reduce the network parameters by exploring the model structure and introducing some specific relationships between different weights. In this section, we recall the signal model of 2D harmonic retrieval problems and reveal its specific characteristics of the dictionary matrices, which helps us couple learned variables in AdaBlockLISTA.

Following the signal model in [13], 2D harmonic retrieval problem can be formed into a block-sparse signal estimation problem. We uniformly sample 2D harmonic frequencies into $P$ and $Q$ points, encapsulated in the set of grid points $\{p/P\}_{p=0}^{P-1}$ and $\{q/Q\}_{q=0}^{Q-1}$, respectively. The full observation matrix $\Psi$ is defined as a 2D discrete Fourier transform (DFT) matrix, given by $\Psi = F_Q \otimes F_P$, where operator $\otimes$ represents Kronecker product, and $F_Q$ is a discrete Fourier matrix of size $Q \times Q$, whose $((i,k)-th$ entry is $[F_Q]_{i,k} = \omega_{Q}^{(i-1)(k-1)}$, where $\omega_{Q} = e^{2\pi i/\gamma_{Q}}$, $i, k = 0, 1, \ldots, Q - 1$. Another discrete Fourier matrix $F_P$ is computed in the same manner.

Considering compressive measurements, we have the dictionary $\Phi \in \mathbb{C}^{N \times M}$ in (1) consists of $N$ sub-sampled rows of the full dictionary $\Psi$. To store the indices of the selected rows, we define a subset $\Omega$ of cardinality $N$ randomly chosen from the set $M = \{1, 2, \ldots, M\}$. We use $\mathcal{R}$ as a row-sampled matrix to select $N$ rows corresponding to the elements in $\Omega$, i.e., $[\mathcal{R}]_{i,m} = 1$, where $m$ is the $n$th element of $\Omega$ while other entries in the $n$th row are zeros. Thus, the dictionary $\Phi$ can be computed as $\Phi = \mathcal{R} (F_Q \otimes F_P)$.

According to the definition of Kronecker product, we find that each sub-matrix $\Phi_q \in \mathbb{C}^{Q \times \gamma_{Q}}$ in (3) can be computed from the first sub-matrix $\Phi_1$ as $\Phi_q = \Lambda^{Q-1} \Phi_1$, where $\Lambda$ is a diagonal matrix defined as:

$$
\Lambda = \begin{bmatrix}
I & 0 & 0 & \cdots & 0 \\
0 & \omega_Q I_P & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \omega_Q I_P & 0 \\
0 & 0 & \cdots & 0 & I_P
\end{bmatrix}
$$

Thus, (3) becomes $\Phi = \{\Phi_1, \Phi_2, \ldots, \Lambda^{Q-1} \Phi_1\}$, where $\Lambda^H \Lambda = I$.

As we have $\Phi_q = \Lambda^{Q-1} \Phi_1$, in 2D harmonic signal model, we propose a weight coupling method motivated by the minimum-coherence criterion for block-sparse recovery in (8); if $\tilde{W}_1$ satisfies the condition of (8), then $\tilde{W}_q = \Lambda^{Q-1} \tilde{W}_1 (\Lambda^{Q-1})^H$ is also the best parameter for (8), which motivates us to couple the matrices $\{\tilde{W}_q\}_{q=1}^{Q}$ in AdaBlockLISTA thus leads to a considerable reduction in the number of trained parameters. Therefore, we propose our coupled network, named AdaBLISTA-CP, whose iteration is:

$$
\tilde{x}_{q}^{(t+1)} = x^t_q + \tau_q (\tilde{y}_q - \tilde{W}_q^{(t)} (\Lambda^{Q-1})^H (y - \sum_{i=1}^{Q} \Lambda^{-1} \Phi_i x^t_i)),
$$

where we only learn a single weight $\tilde{W}_1$ and generate other weights by multiplication with $\Lambda$. Thus, the number of learned parameters in AdaBLISTA-CP is reduced to $O(N^2)$, which contributes to lower storage burden and higher sample efficiency. We illustrate the network structure of AdaBLISTA-CP in Fig. 1.

**NUMERICAL RESULTS:** We compare the performance of three deep unfolding networks (AdaLISTA, Ada-BlockLISTA, and our AdaBLISTA-CP) in block sparse recovery. In our simulation, we generate noisy observed signals according to (1) where the block-sparse signal $x^* \in \mathbb{C}^{p \times q}$ has $Q = 64$ blocks each with block size $p = 4$, and the number of non-zero blocks in $x$ is $K \in \{1, 2, 3, 4, 5\}$. Note that the inputs to our reconstruction network are complex-value data, thus we transform every operator above to its complex value counterparts by following complex-value extension methods presented in [7].

As shown in Fig. 2, we evaluate the support recovery performance of block-sparse signals in terms of hit rate versus signal-to-noise ratio (SNR) and block sparsity. The SNR is computed as $SNR = 10 \log_{10} \frac{\sigma_2}{\sigma_1}$, where $\sigma_2$ is the noise variance. The hit rate is defined as the percentage of successes in finding the nonzero blocks in $x$ against noise. In Fig. 2, a larger area of

**CONCLUSION:** In this work, we considered the block-sparse signal model in 2D harmonic retrieval problem, where nonzero entries of recovered signal occur in clusters, and derived our AdaBLISTA-CP network by leveraging the particular dictionary structure in the signal model. AdaBLISTA-CP inherits the structure of dictionary and reduces the number of parameters to train by exploring the relationship between different weights across blocks. In simulation, our proposed network with coupled weight matrices shows excellent recovery performance in terms of hit rate, better than Ada-LISTA and comparable to Ada-BlockLISTA.

**Acknowledgment:** This project is funded by the National Natural Science Foundation of China under Grants No. 61801258 and 62171259.

Rong Fu, Tianyao Huang, Lei Wang, Yimin Liu (Department of Electrical Engineering, Tsinghua University, Beijing, China)

Correspondence: huangtianyao@tsinghua.edu.cn

**References**

1. R. D. Balakrishnan and H. M. Kwon, “A new inverse problem based approach for azimuthal doa estimation,” in Proc. IEEE Global Commun. Conf. (IEEE GLOBECOM), vol. 4, 2004, pp. 2187–2191.

2. A. Xenaiki, P. Gerofti, and K. Mougeard, “Compressive beamforming,” The Journal of the Acoustical Society of America, vol. 136, p. 260, 07 2014.

3. Y. Li, T. Huang, X. Xu, Y. Liu, L. Wang, and Y. C. Eldar, “Phase transitions in frequency agile radar using compressed sensing,” IEEE Transactions on Signal Processing, vol. 69, pp. 4801–4818, 2021.

4. K. Gregor and Y. Lecun, “Learning fast approximations of sparse coding,” in International Conference on International Conference on Machine Learning, 2010, pp. 399–406.

5. M. Borgedirg, P. Schniter, and S. Rangan, “AMP-inspired deep networks for sparse linear inverse problems,” IEEE Transactions on Signal Processing, vol. 65, no. 16, pp. 4293–4308, Aug 2017.

6. M. Borgedirg and P. Schniter, “Onsager-corrected deep learning for sparse linear inverse problems,” in 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Dec 2016, pp. 227–231.

7. R. Fu, Y. Liu, T. Huang, and Y. C. Eldar, “Structured lista for multidimensional harmonic retrieval,” IEEE Transactions on Signal Processing, vol. 69, pp. 3459–3472, 2021.

8. R. Fu, Y. Liu, T. Huang, and Y. C. Eldar, “ALISTA: Analytic weights are as good as learned weights in LISTA,” in International Conference on Learning Representations, 2019, [Online]. Available: https://openreview.net/forum?id=B1lnzn0ctQ

9. A. Aberdam, A. Golts, and M. Elad, “Ada-lista: Learned solvers adaptive to varying models,” arXiv:2001.08456, 2020.

10. R. Fu, V. Monardo, T. Huang, and Y. Liu, “Deep unfolding network for block-sparse signal recovery,” in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 2880–2884.

11. A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” Siam J Imaging Sciences, vol. 2, no. 1, pp. 183–202, 2009.

12. Y. C. Eldar, P. Kuppinger, and H. Bolcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” IEEE Transactions on Signal Processing, vol. 58, no. 6, pp. 3042–3054, 2010.

13. Y. Chi and Y. Chen, “Compressive two-dimensional harmonic retrieval via atomic norm minimization,” IEEE Transactions on Signal Processing, vol. 63, no. 4, pp. 1030–1042, Feb 2015.