Deep unfolding for analysis Compressed Sensing: does redundancy affect the generalization ability?

Vasiliki Kouni
National & Kapodistrian University of Athens, Greece
Email: vicky-kouni@di.uoa.gr

***Abstract*—In this paper, we examine an ADMM-based deep unfolding network for analysis Compressed Sensing, dubbed U-ADMM-DAD net; the latter jointly learns a decoder for Compressed Sensing and a redundant analysis operator for sparsification. We compare U-ADMM-DAD net to a synthesis-sparsity-based unfolding network – serving as a baseline – on real-world speech data. Our experimental results demonstrate that the redundancy of the learnable sparsifier affects the generalization ability of U-ADMM-DAD net, which outperforms the baseline in terms of both reconstruction and generalization error.**

I. Introduction

Compressed Sensing (CS) [1] deals with reconstructing a vector $x \in \mathbb{R}^n$ from incomplete measurements $y = Ax + e \in \mathbb{R}^m$, $m < n$. To do so, we assume there exists a *redundant analysis operator* $\Phi \in \mathbb{R}^{N \times n}$ $(N > n)$, such that Φx is (approximately) sparse. Recently, approaches based on deep learning were also introduced [2], [3]. It seems promising to merge these two areas by considering what is called *deep unfolding* [4]. The latter pertains to unfolding the iterations of well-known optimization algorithms into layers of a deep neural network (from now called *unfolding network*), which reconstructs the signal of interest x.

A. Motivation and key results

Modern unfolding networks [5], [6], [7], [8], [9] jointly learn a decoder for CS and a dictionary that sparsely represents x. Inspired by [7], [9], we examine the generalization ability of the deep unfolding network presented in [9]. Our main contributions are the following a) we remove the clipping function that is applied on the learned decoder of [9], since we experimentally observe that there is no need for such a truncation operation b) we numerically compare the aforementioned unclipped version of ADMM-DAD – from now on called Unclipped ADMM-DAD (U-ADMM-DAD) – to a state-of-the-art ISTA-based unfolding network, on speech data. Our results show that the U-ADMM-DAD outperforms the baseline, indicating that the redundancy of the learned analysis operator leads to a smaller test MSE and generalization error as well.

II. Main Results

A. ADMM-DAD's derivation

According to [9], the iterative scheme of ADMM is formulated as a neural network with $L \in \mathbb{N}$ layers/iterations, defined as $f_1(y) = I_1 b(y) + I_2 \mathcal{S}_{\lambda/\rho}(b(y))$ and $f_k(v) = \tilde{\Theta} v + I_1 b + I_2 \mathcal{S}_{\lambda/\rho}(\Theta v + b)$, $k = 2, \ldots, L$. The trainable parameters are the entries of Φ, with the latter appearing in the formulation of the matrices $\tilde{\Theta}$, Θ and the vector $b = b(y)$. We denote the composition of L such layers (all having the same Φ) as $f_\Phi^L(y) = f_L \circ \cdots \circ f_1(y)$ and the final output $\hat{x}$ is obtained after applying an affine map T to the final layer L, i.e., $\hat{x} = T(f_\Phi^L(y))$. We introduce the hypothesis class $\mathcal{H}^L = \{h : \mathbb{R}^m \mapsto \mathbb{R}^n : h(y) = T(f_\Phi^L(y)), \Phi \in \mathbb{R}^{N \times n}, N > n\}$ consisting of all the decoders that U-ADMM-DAD can implement. Given the aforementioned class and a training set $\mathcal{S} = \{(y_i, x_i)\}_{i=1}^s$, U-ADMM-DAD yields a function/decoder $h_\mathcal{S} \in \mathcal{H}^L$ that aims at reconstructing x from $y = Ax + e$. In order to measure the difference between x_i and $\hat{x}_i = h_\mathcal{S}(y_i)$, $i = 1, \ldots, s$, we choose the training mean-squared error (train MSE) $\mathcal{L}_{train} = \frac{1}{s}\sum_{i=1}^s \|h_\mathcal{S}(y_i) - x_i\|_2^2$ as loss function. The test mean-squared error (test MSE) is defined as

$$\mathcal{L}_{test} = \frac{1}{d}\sum_{i=1}^{d} \|h_\mathcal{S}(\tilde{y}_i) - \tilde{x}_i\|_2^2, \tag{1}$$

where $\mathcal{D} = \{(\tilde{y}_i, \tilde{x}_i)\}_{i=1}^d$ is a set of d test data, not used in the training phase. The generalization error is then defined by

$$\mathcal{L}_{gen} = |\mathcal{L}_{test} - \mathcal{L}_{train}|. \tag{2}$$

III. Experiments

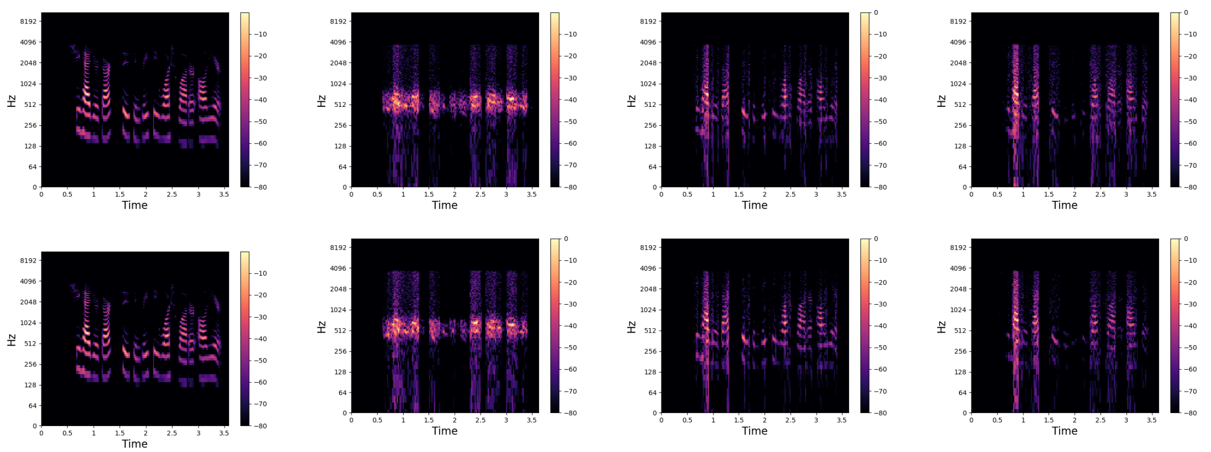
We train and test U-ADMM-DAD on the TIMIT [10] speech dataset. We compare the 10-layer ADMM-DAD, for different redundancy ratios N/n and number of measurements m, to the 10-layer ISTA-net proposed in [7] (which jointly learns a decoder for CS and an orthogonal sparsifier). We report in Table I the average test MSE and generalization error, as defined in (1) and (2), respectively. Both the test and generalization errors are always lower for our U-ADMM-DAD net than for ISTA-net. In fact, both the test MSE and generalization error of U-ADMM-DAD decrease, as N/n and m increase. On one hand, this behaviour of the test MSE seems reasonable, if one considers a standard analysis CS scenario: the reconstruction error provided by the analysis-l_1 algorithm typically benefits from the (high) redundancy offered by the involved analysis operator. On the other hand, this decrement of the generalization error looks very interesting and in need of mathematical investigation. Overall, the results from Table I indicate that the redundancy of the learned analysis operator improves the performance of U-ADMM-DAD. Furthermore, we extract the spectrograms of an example test raw audio file of TIMIT, reconstructed by either of the 10-layer decoders, for 40% and 60% CS ratios. The results are illustrated in Fig. 1, which indicates that our proposed decoder outperforms the baseline, since the former distinguishes many more frequencies than the latter. For both CS ratios, U-ADMM-DAD reconstructs a clearer version of the signal compared to ISTA-net; the latter recovers a significant part of noise, even for the 60% CS ratio.

IV. Conclusion

In the present paper, we examined U-ADMM-DAD, an unfolding network that jointly learns a decoder for Compressed Sensing and a redundant analysis operator, serving as sparsifier for the signals of interest. We compared our framework with a state-of-the-art ISTA-based unfolding network on speech data. Our experiments confirm improved performance: the redundancy provided by the learned analysis operator yields a lower average test MSE and generalization error of our method compared to the ISTA-net. Future work will include the derivation of generalization bounds for the hypothesis class defined in the previous section, similarly to [7].

10 layers	40% CS ratio				60% CS ratio			
Decoder \ Redundancy ratio	$N/n = 10$		$N/n = 60$		$N/n = 10$		$N/n = 60$	
	test MSE	gen. error	test MSE	gen. error	test MSE	gen. error	test MSE	gen. error
ISTA-net	$0.46 \cdot 10^{-2}$	$0.18 \cdot 10^{-2}$	$0.20 \cdot 10^{-3}$	$0.25 \cdot 10^{-4}$	$0.45 \cdot 10^{-2}$	$0.20 \cdot 10^{-2}$	$0.20 \cdot 10^{-3}$	$0.25 \cdot 10^{-4}$
ADMM-DAD	$\mathbf{0.52 \cdot 10^{-4}}$	$\mathbf{0.67 \cdot 10^{-5}}$	$\mathbf{0.43 \cdot 10^{-4}}$	$\mathbf{0.50 \cdot 10^{-5}}$	$\mathbf{0.24 \cdot 10^{-4}}$	$\mathbf{0.31 \cdot 10^{-5}}$	$\mathbf{0.23 \cdot 10^{-4}}$	$\mathbf{0.26 \cdot 10^{-5}}$

TABLE I: Average test MSE and generalization error for 10-layer decoders. Bold letters indicate the best performance between the two decoders.



(a) Original

(b) 10-layer ISTA-net reconstruction

(c) 10-layer ADMM-DAD reconstruction, $N/n = 10$

(d) 10-layer ADMM-DAD reconstruction, $N/n = 60$

Fig. 1: Spectrograms of reconstructed test raw audio file from TIMIT for 40% CS ratio (top) and 60% CS ratio (bottom).

References

[1] S. Foucart, and H. Rauhut, "An invitation to compressive sensing," in A mathematical introduction to compressive sensing, pp. 1-39, Birkhäuser, New York, NY.

[2] G. Yang, et al., "DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," IEEE transactions on medical imaging, 37(6), 2017, pp.1310-1321.

[3] A. Mousavi, G. Dasarathy, and R. G. Baraniuk, "Deepcodec: Adaptive sensing and recovery via deep convolutional neural networks," arXiv preprint arXiv:1707.03386, 2017.

[4] V. Monga, Y. Li and Y. C. Eldar, "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing," in IEEE Signal Processing Magazine, vol. 38, no. 2, pp. 18-44, March 2021, doi: 10.1109/MSP.2020.3016905.

[5] J. Zhang, and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," IEEE conference on computer vision and pattern recognition, pp. 1828-1837, 2018.

[6] J. Sun, L. Huibin, and X. Zongben, "Deep ADMM-Net for compressive sensing MRI," Advances in neural information processing systems, 29, 2016.

[7] A. Behboodi, H. Rauhut, and E. Schnoor, "Compressive Sensing and Neural Networks from a Statistical Learning Perspective." arXiv preprint arXiv:2010.15658, 2020.

[8] E. Schnoor, A. Behboodi, and H. Rauhut, "Generalization Error Bounds for Iterative Recovery Algorithms Unfolded as Neural Networks," arXiv preprint arXiv:2112.04364, 2021.

[9] V. Kouni, G. Paraskevopoulos, H. Rauhut and G. C. Alexandropoulos, "ADMM-DAD Net: A Deep Unfolding Network for Analysis Compressed Sensing," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 1506-1510, doi: 10.1109/ICASSP43922.2022.9747096.

[10] J. S. Garofolo, et al. "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," NASA STI/Recon technical report n 93: 27403, 1993.