EFFICIENT MODEL-FREE LEARNING TO OVERCOME HARDWARE NONIDEALITIES IN ANALOG-TO-INFORMATION CONVERTERS

Sam Buchanan[*†] *Tanbir Haque*[*#] *Peter Kinget*[*] *John Wright*[*†]

[*] Columbia University Dept. of Electrical Engineering
[†] Columbia University Data Science Institute
[#] InterDigital Communications Inc.

ABSTRACT

This paper considers compressed sensing (CS) in the context of RF spectrum sensing and presents an efficient approach for learning hardware nonidealities in an analog-to-information converter (A2IC). The proposed methodology is based on the learned iterative shrinkage-thresholding algorithm (LISTA), which enables co-optimization of the hardware and the reconstruction algorithm and leads to a model-free recovery approach that is optimally tuned for the unique computational constraints and hardware nonidealities present in the RF frontend. To achieve this, we devise a training protocol that employs a dataset and neural network of minimal sizes. We demonstrate the effectiveness of our methodology on simulated data from a model of a well-established CS A2IC in the presence of linear impairments and noise. The recovery process extrapolates from training on 1-sparse signals to recovering the support of signals whose sparsity runs up to the theoretical optimum for ℓ^1-based algorithms across a range of typical operating SNRs.

Index Terms— Compressed Sensing, Artificial Neural Network, LISTA, Spectrum Sensing, Analog-to-Information Converter

1. INTRODUCTION

Compressed sensing (CS) theory [1–4] asserts the existence of schemes for sensing structured signals that are vastly more efficient than the state-of-the-art in numerous applications, but translating these theorems into nontrivial improvements in key metrics has proved challenging. One particularly promising application area is spectrum sensing [5], where CS-based RF frontend receivers have surpasssed classical scan time vs. energy consumption tradeoffs [6–9] and could become a key enabling technology for cognitive-radio-based [10] dynamic spectrum access. Two challenges that have hindered general adoption of these designs are (i) the difficulty of efficiently implementing both greedy and iterative reconstruction algorithms in hardware, and (ii) the sensitivity of reconstruction algorithms' performance to deviations from the nominal design in the A2IC hardware. Point (ii) is a more general case of the well-studied issue of basis mismatch [11, 12], and thus signal reconstruction algorithms that are robust to linear and nonlinear impairments are essential. As to issue (i), we note that greedy algorithms for sparse recovery such as the orthogonal matching pursuit (OMP) [13, 14] require the computation of subspace projection operators on each iteration, which is impractical for IC implementation, whereas iterative algorithms such as basis pursuit can be incompatible with the non-negotiable constraints on slot times and computational resources that are a fundamental part of ICs for spectrum sensing. An ideal recovery algorithm would offer optimized performance subject to the particular constraints imposed by one's application and admit an efficient hardware implementation for use in real-time systems.

To address issues (i) and (ii) and move towards robust and hardware-feasible implementations of CS reconstruction algorithms, we propose an efficient signal reconstruction methodology based on the learned iterative shrinkage-thresholding algorithm (LISTA) [15]. In LISTA-type methods, a correspondence between sparse recovery algorithms and artificial neural networks motivates a learning problem in which input-output data is used to optimize the parameters of an iterative sparse recovery algorithm truncated to a fixed number of iterations. The resulting learning problem offers the chance to automatically adapt the recovery algorithm to hardware nonidealities through training, while simultaneously being well-justified from the perspective of numerical optimization: critically, it is in no way a black-box learning procedure. At the same time, this process produces a recovery algorithm that requires no matrix decompositions or inverses and is optimally tuned to the specific computational budget available in one's application. Existing studies of the LISTA method have focused on optimizing certain choices of parameters in general LISTA networks [16–18], extensions to more complex iterative methods than ISTA [19–21], and adaptations to domains where different structural priors are present [22]. In contrast, our work focuses on the development of a hardware-feasible methodology for applying LISTA to sparse signal reconstruction in the context of spectrum sensing, and it identifies and exploits LISTA's ability to enable model-free adaptivity through the process of optimizing performance over the training set. To this end, we propose and evaluate an efficient training protocol for use with reconstructing sparse RF signals—since *a priori* it is unclear how much input-output data is necessary to produce a robust signal reconstruction algorithm using the LISTA framework—and demonstrate the ability of the post-training signal reconstruction algorithm to operate robustly in the presence of linear circuit nonidealities in simulated data. In particular, we present numerical evidence that our methodology can recover near-uniform performance associated with certain ensembles of random matrices when the design matrix is an appropriate pseudorandom matrix and the system matrix is an impaired version of the design matrix.

2. SYSTEM MODELING

In spectrum sensing, one is generally interested in obtaining information about spectral occupancies of various parts of the RF spectrum. In this paper we consider a formulation of this problem that is amenable to CS techniques, where signals $x(t)$ in a class of signals $\mathcal{S}$ are assumed to be frequency-sparse. In particular, we define

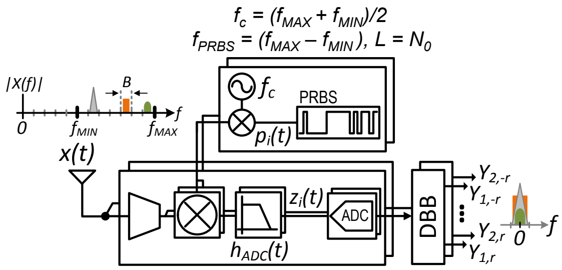

Fig. 1. Simplified schematic of the Direct RF-to-Information Converter.

$\mathcal{S}$ as the set of real-valued, square-integrable signals $x(t)$ satisfying two properties: (i) the Fourier transform $X(f)$ vanishes outside of a known interval $\mathcal{F} = (-f_{\max}, -f_{\min}] \bigcup [f_{\min}, f_{\max})$ and (ii) the support of $X(f)$ is a relatively small subset of $\mathcal{F}$. Furthermore, we partition $\mathcal{F}$ into into an even integer $N \in \mathbb{N}$ disjoint bands, each of width at most B, and assume that $X(f)$ is supported on only $K \ll N$ of these bands, where K is also even. It will be convenient to define additional parameters $\mathcal{F}_0 = [f_{\min}, f_{\max})$, $N_0 = (1/2)N$, and $K_0 = (1/2)K$ for working with single-sided spectra in the sequel. We focus on the case where all distinct bands of $X(f)$ have equal average power in our experiments, and for technical reasons we also require N_0 to be odd. Our goal is to recover the spectrum $X(f)$ from a small number of random linear measurements by exploiting sparsity.

Our receiver model and experiments target the Direct RF-to-Information Converter (DRF2IC) [6], which is illustrated in Fig. 1. We study this system due to its superior performance among receivers of the modulated wideband converter (MWC) type [8, 9] in terms of scan time and energy consumption, which makes it a candidate for inclusion in future dynamic spectrum access systems. Measurements are acquired in the DRF2IC using a pseudo-random bit sequence (PRBS)-modulated quadrature LO to create a very wide bandpass RF response with a 3dB bandwidth extending from $f_{\min}$ to $f_{\max}$. This mixing operation is performed with independent PRBSs on each of the m system branches to produce m shifted complex baseband spectrum measurements, which are subsequently low-pass filtered. All N_0 bins of width B from $\mathcal{F}_0$ are encoded in these measurements when the PRBSs are suitably chosen—for example, periodic Rademacher sequences of sufficient length are enough. The system is designed to oversample the m complex measurement branches at the digitizer so that the RF frontend's hardware complexity can be minimized in exchange for additional digital processing. To this end, the digitized measurements are multiplied with $2r$ orthogonal complex exponentials of frequencies that are integer multiples of B to extract $2r$ higher order intermediate frequency responses in addition to the response at DC [9]. The total number of measurements thus produced is $m(2r + 1)$, which we denote as R. The choice of r is flexible, so that the choice of m essentially determines the sparsity level of signals in $\mathcal{S}$ the system can reconstruct. Standard CS theory performance guarantees suggest a choice of $R \geq \lceil CK_0 \log(N_0/K_0) \rceil$, where $C > 0$ is an absolute constant, which has been shown experimentally to be a valid heuristic in this application [9, 23].

One can show in a straightforward manner [9, 23] that the model described above implies a linear relationship between the measured signal $\boldsymbol{y}(f) \in \mathbb{C}^R$ and the discretized RF signal $\boldsymbol{x}(f) \in \mathbb{C}^{N_0}$: for $f \in [-f_s/2, +f_s/2]$, one has $\boldsymbol{y}(f) = \tilde{\boldsymbol{A}}\boldsymbol{x}(f)$. Here the system matrix $\tilde{\boldsymbol{A}}$ captures the down-conversion, PRBS mixing, filtering, and branch expansion operations described above, and the number f_s denotes the ADC sampling frequency reduced by the factor $2r + 1$. Furthermore, the k-th row of $\boldsymbol{x}(f)$ has frequency support offset by $(k - (N_0 - 1)/2)f_s/(2r + 1)$. When impairments are present in the RF frontend, the map $\tilde{\boldsymbol{A}}$ is not equal to the ideal design matrix $\boldsymbol{A}$, and the resulting mismatch presents difficulties for sparse signal reconstruction algorithms. In this work, we study two types of realistic system impairments in a simulation environment. The first is deterministic relative phase shifts in the PRBSs by less than half of the PRBS switching frequency, which arise due to unavoidable circuit design limitations and pose problems for CS A2ICs that perform branch expansion. The second is frequency-independent gain and phase imbalances in the downconversion hardware. It can be shown that both of these types of impairments act linearly on the design matrix $\boldsymbol{A}$, so that

$$\boldsymbol{y}(f) = \mathcal{L}[\boldsymbol{A}]\,\boldsymbol{x}(f), \quad f \in [-f_s/2, +f_s/2] \tag{1}$$

accurately models the measurement process, where $\mathcal{L}$ is an unknown linear operator on the set of $R \times N_0$ matrices with complex entries that we assume is intrinsic to each instance of the DRF2IC. We will demonstrate that our LISTA-based reconstruction algorithm is robust to the action of $\mathcal{L}$ when the operator implements the nonidealities described above and signals are drawn from $\mathcal{S}$.

3. LISTA-BASED SIGNAL RECONSTRUCTION

We now present an efficient LISTA approach for finding a sparse solution to the linear system (1) given the design matrix but no knowledge of the operator $\mathcal{L}$. For a fixed discretized complex baseband signal $\boldsymbol{x} \in \mathbb{C}^{N_0}$, assumed stationary and corresponding to the complex baseband representation of a time signal $x(t)$ of the class $\mathcal{S}$, we collect p noisy time-domain samples $\boldsymbol{y}_j = \tilde{\boldsymbol{A}}\boldsymbol{x}_j + \boldsymbol{n}_j$, $j = 1, \ldots, p$, where $\boldsymbol{n}_j \sim_{i.i.d.} \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_R)$ and the samples take values in $\mathbb{C}^R$. It will be convenient to assume that $\max_{i\in[R], j\in[p]} |y_{ij}| = 1$, which fixes a σ^2-independent scale for our measurements and can be achieved in hardware with proper amplification. Given a receiver model $\boldsymbol{A}$, a natural optimization formulation for this inverse problem is the so-called basis pursuit denoising (BPDN) model [24]:

$$\underset{\boldsymbol{X}\in\mathbb{C}^{N_0\times p}}{\text{minimize}} \quad \tfrac{1}{2}\,\|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{Y}\|_F^2 + \lambda\,\|\boldsymbol{X}\|_{2,1}\,, \tag{2}$$

where the j-th column of $\boldsymbol{Y}$ is $\boldsymbol{y}_j$, $\boldsymbol{X} \mapsto \|\boldsymbol{X}\|_{2,1}$ is the sum of the ℓ^2 norms of the rows of $\boldsymbol{X}$, which encourages the rows of the solution to (2) to be sparse, and the regularization parameter $\lambda > 0$ allows to control the tradeoff between the regularization of noise and the enforcement of sparsity. This is a convex problem of type amenable to solution by the ISTA algorithm [25], whose iterative updates take the form

$$\boldsymbol{X}^{(k+1)} = \operatorname{prox}_{\frac{\lambda}{L}\|\cdot\|_{2,1}}\left(\tfrac{1}{L}\boldsymbol{A}^*\boldsymbol{Y} + (\boldsymbol{I} - \tfrac{1}{L}\boldsymbol{A}^*\boldsymbol{A})\boldsymbol{X}^{(k)}\right). \tag{3}$$

Here $L = \|\boldsymbol{A}\|^2$ is a Lipschitz constant for the gradient of the smooth term in (2) and the proximal mapping has the convenient form

$$\boldsymbol{e}_i^T \operatorname{prox}_{\frac{\lambda}{L}\|\cdot\|_{2,1}}(\boldsymbol{X}) = \left(1 - \tfrac{\lambda}{L\|\boldsymbol{e}_i^T\boldsymbol{X}\|_2}\right)_+ \boldsymbol{e}_i^T\boldsymbol{X}; \tag{4}$$

see e.g. [26]. When $\boldsymbol{A} = \tilde{\boldsymbol{A}}$ and the matrix $\tilde{\boldsymbol{A}}$ is drawn from an ensemble of matrices satisfying a condition such as the restricted isometry property [27], the (right) unitary invariance of the norms involved in (2) implies that solving (2) recovers the spectral occupancies of sparse signals $x(t)$ with high probability if p is sufficiently large, λ is set appropriately and the noise is not too strong [28].

In practice, the assumption that $\boldsymbol{A} = \tilde{\boldsymbol{A}}$ is often unreasonable, and thus the algorithm prescribed by (3) may fail to recover sparse measurements from $\tilde{\boldsymbol{A}}$. To address this, the LISTA algorithm identifies the iterations (3) with the layers of a feedforward neural network and proposes to truncate the iterations after a fixed number of rounds $D \geq 1$ and optimize the error over a training set $\{(\boldsymbol{Y}_i, \boldsymbol{X}_i)\}_{i=1}^T$ with respect to $\boldsymbol{A}$. If we choose to optimize the mean squared error and relax the parameterization in terms of $\boldsymbol{A}$ to one in terms of two matrices $\boldsymbol{W}$ and $\boldsymbol{S}$ with sizes commensurate with those of $\boldsymbol{A}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$, this procedure amounts to solving the optimization problem

$$\underset{\boldsymbol{W} \in \mathbb{C}^{N_0 \times R},\, \boldsymbol{S} \in \mathbb{S}^{N_0 \times N_0}(\mathbb{C})}{\text{minimize}} \quad \frac{1}{2} \sum_{i=1}^{T} \left\| \boldsymbol{X}^{(D)}(\boldsymbol{Y}_i) - \boldsymbol{X}_i \right\|_F^2 \qquad (5)$$

where $\mathbb{S}^{N_0 \times N_0}(\mathbb{C})$ denotes the set of self-adjoint complex matrices of size $N_0 \times N_0$, and where $\boldsymbol{X}^{(D)}(\boldsymbol{Y}_i)$ is the result of iterating (3) D times with $\boldsymbol{Y}$ set to $\boldsymbol{Y}_i$, $\boldsymbol{X}^{(0)}$ fixed as $\boldsymbol{0}$, and with $(1/L)\boldsymbol{A}^*$ replaced by $\boldsymbol{W}$ and $\boldsymbol{I} - (1/L)\boldsymbol{A}^*\boldsymbol{A}$ replaced by $\boldsymbol{S}$. The resulting problem is a recurrent neural network learning problem for which a critical point can be approximately computed using a package such as TensorFlow [29] that facilitates efficient gradient computations. Furthermore, hardware implementation of the learned recovery algorithm requires in principle little more than addition and multiplication primitives.

4. TRAINING PROTOCOL

To structure the LISTA protocol for recovering signals under the model (1), one must choose suitable values for the parameters D, λ, and the dataset $\{(\boldsymbol{Y}_i, \boldsymbol{X}_i)\}_{i=1}^T$. One must also give a recipe for initializing the parameters of the recurrent neural network. First we remark that the choice of the parameter D allows the designer to incorporate hard computational constraints directly and obtain an optimized recovery algorithm subject to those constraints. Similarly, coarse prior information in the form of the design map $\boldsymbol{A}$ can be used to initialize the gradient descent algorithm run on (5). Due to the infeasibility of tracking the operator norm of the matrix $\boldsymbol{S}$ across gradient descent iterations in hardware, we recommend scaling the initial value of $\boldsymbol{S}$ to have operator norm less than 1 and fixing L to this scaled value throughout the training.

Design of the composition of the dataset is necessarily more nuanced, since it is infeasible to collect a dataset of all possible support patterns for sparse signals from $\mathcal{S}$. Such a set would have size at least $\sum_k \binom{N_0}{k}$, where k runs up to approximately $R/\log(N_0)$. In general, however, it is unclear whether uniform performance guarantees can be maintained—or recovered, in the case where $\mathcal{L}$ is not the identity—on a structured set of signals such as the sparse signals without such an exhaustive training regimen. Our experiments demonstrate that, in the setting we consider, these guarantees can be recovered in large part using an efficient training strategy: we propose a dataset of size kN_0 for some $k \in \mathbb{N}$ consisting of measurements of 1-sparse signals $x_i(t) \in \mathcal{S}$, and ensure uniform composition of the dataset so that each of the possible 1-sparse signals supported on one of the N_0 subbands of $\mathcal{F}_0$ is included in the dataset with multiplicity k, with independent noise. We fix the SNR of the signals $x_i(t)$ to be 20 dB, and remark that in this regime it generally suffices to choose $k = 1$, which enhances the energy efficiency of the training process. Because the model (1) is linear and the matrix $\tilde{\boldsymbol{A}}$ is incoherent (under the model of [9, 23] used for our experiments), adding noise in this manner is roughly equivalent to that used in the definition of the samples $\boldsymbol{y}_j$ in section 3. Our protocol may seem to run contrary to the general intuition in deep learning, but our experiments will support the notion that for recovery of highly structured

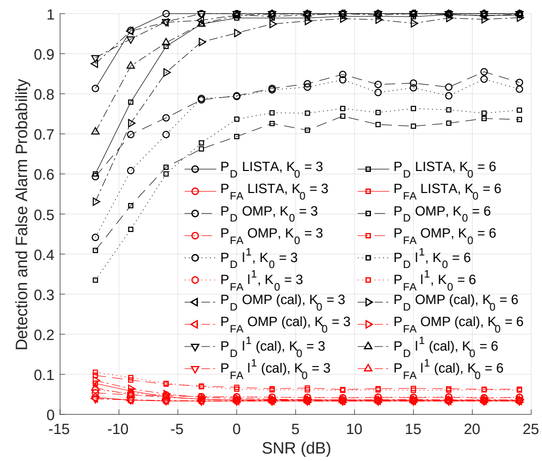

Fig. 2. Probability of detection and probability of false alarm plots for OMP, LISTA, and proximal gradient descent in the presence of linear nonidealities over signal SNR. For $K_0 \in \{3, 6\}$ we plot the mean P_D and P_{FA} averaged over 200 experiments on independently-generated signals with SNR swept from -12 dB to 21 dB in steps of 3 dB. Calibrated algorithms are run using unimpaired data.

signals such as sparse signals, a dataset of this size is sufficient at least in certain regimes of signal parameters.

We describe below some additional parameter selection that is linked to the geometric regularity of the problem (5). We observe empirically (c.f. Fig. 4) that performance measured in probability of detecting an occupied bin in the spectrum is nonincreasing over a band of values of D from 12 to 30, with deviations at most 3% from the value at $D = 12$, and our protocol thus employs a value of 12. We choose λ to be at the upper end of admissible parameter values for (2) given that we have fixed D to be relatively small compared to a typical number of iterations required to converge under the $O(1/k)$ guarantee of the ISTA algorithm; this decision is further justified by the rapid convergence of gradient methods for sparse regression problems [30]. We find experimentally that a value of 0.1 suffices for DRF2IC data under the specific case of the model (1) we study. All gradients are computed exactly over the full size T dataset, and in particular we employ a constant-stepping gradient method with step size chosen sufficiently small to avoid divergence in function value of iterates when solving (5).

5. EXPERIMENTAL RESULTS

We present experiments on synthetic data from a model of the DRF2IC, which is based on the derivation of the system equation (1) in [9, 23] with modifications for DRF2IC features [6]. We study the effects of downconverter gain and phase imbalance and PRBS phase mismatch on the recovery performance of LISTA and OMP across multiple levels of signal sparsity and signal SNR. All experiments employ $p = 80$ samples with $R = 18$ branch-expanded measurements and $N_0 = 63$ possible signal support bands. Our LISTA training protocol is implemented in TensorFlow, and we set the LISTA parameters to $D = 12$, $\lambda = 0.1$ and $k = 1$, with the SNR and sparsity level of the training signals set as 20 dB and 1, respectively. Our system resolves a swath of spectrum from $f_{\min} = 2.57$ GHz to $f_{\max} = 3.83$ GHz, which implies an RF-to-baseband oversampling ratio of 8, so that the PRBS relative phase shift must fall in the range

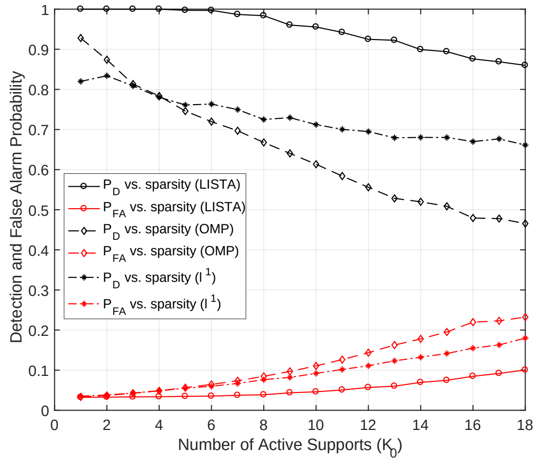

Fig. 3. Probability of detection and probability of false alarm plots for OMP, LISTA, and proximal gradient descent in the presence of linear nonidealities over signal sparsity. We fix the signal SNR at 20 dB and plot the mean P_D and P_{FA} averaged over 500 experiments on independently-generated signals with sparsity swept from 1 to R.

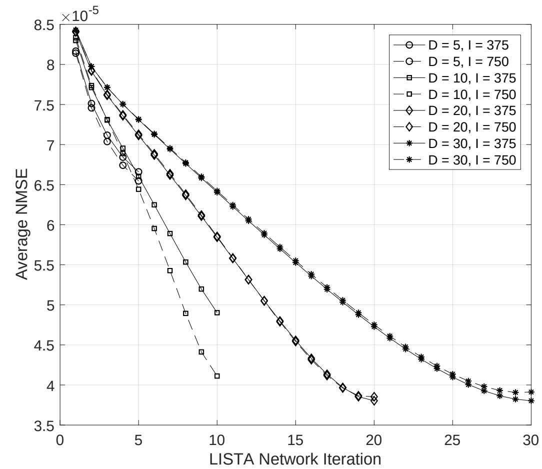

Fig. 4. Comparison of LISTA performance in NMSE against training parameters. We choose $D \in \{5, 10, 20, 30\}$ and evaluate models after $I \in \{375, 750\}$ epochs of training on (5) with independent signals with $K_0 = 6$ and 20 dB SNR. Data points represent NMSE averaged over a set of 500 evaluation signals.

$\{-1, -7/8, \ldots, -1/8, 0, 1/8, \ldots, 7/8, 1\}$. We choose a value of 6/8 for our experiments, and a gain-phase imbalance of (3dB, 30°) for the downconverter. Our dataset $\{(\boldsymbol{Y}_i, \boldsymbol{X}_i)\}_{i=1}^{N_0}$ is constructed according to the procedure in section 4, and in particular is measured by the impaired system $\tilde{\boldsymbol{A}}$.

We initialize the LISTA network appropriately using $\boldsymbol{A}$-derived parameters and iterate for 700 epochs of gradient descent. The resulting network parameters are then used to reconstruct further measurements from $\tilde{\boldsymbol{A}}$ of signals in $\mathcal{S}$ from outside the training set. We evaluate performance measured relative to that of OMP and a fast proximal gradient algorithm applied to (2) with $\lambda = 0.1$. For proximal gradient descent and LISTA, we evaluate probability of detection and probability of false alarm by taking the squared magnitude of the recovered signal and summing over each measurement's p samples to create a length-N_0 vector, and then taking the indices of the elements in this vector with the top $K_0 + 2$ amplitudes to be our estimate of the support.

Figs. 2 and 3 compare the performance of a LISTA model trained using input-output data from the system matrix to uncalibrated OMP and proximal gradient descent algorithms. The results in Fig. 2 show performance on signals from $\mathcal{S}$ having $K_0 \in \{3, 6\}$, and the results in Fig. 3 are evaluated on signals with SNR fixed at 20 dB. These experiments show that OMP and proximal gradient descent cannot sustain $P_D \geq 0.8$ across standard operating conditions in the presence of linear impairments, whereas LISTA is able to maintain performance at a level comparable to that of the former two algorithms run on an unimpaired system. These experiments also demonstrate the robustness of the LISTA recovery approach in that, similar to standard CS reconstruction algorithms in the absence of basis mismatch, the LISTA algorithm shows a controlled drop in performance as the signal SNR falls to a point where reliable detection is not possible and the signal sparsity decreases to the point where CS guarantees fail.

It is worth remarking that both OMP and the proximal gradient descent algorithm lack the calibration process inherent in the LISTA training process, so that their performance is necessarily diminished in the presence of the impairment operator $\mathcal{L}$. Nevertheless, the calibration process necessary to improve the performance of these two algorithms is laborious and inherently inexact, and furthermore any calibration process used must necessarily grow in time and complexity as new hardware nonidealities are introduced to the model. The LISTA approach does not suffer from these drawbacks: it is model-free, and we believe that it will extend readily to compensation of novel types of linear impairments. Thus, the performance disparities between these three algorithms should be analyzed from this methodological viewpoint, and not just a 'calibration time' perspective.

Fig. 4 offers insight into the sensitivity of the LISTA training protocol to the number of layers D and the number epochs of training I. These results demonstrate that the NMSE achieved by a 10 layer network is comparable to that of a 30 layer network. Nonetheless, the performance gap between the $D = 5$ network and the $D = 10$ network illustrates the benefit of choosing D sufficiently large, and the performance gap between the $I = 375$ and $I = 750$ curves for $D = 10$ illustrates that the length of training also plays a role.

6. CONCLUSIONS

We have presented an efficient training protocol that enables a LISTA-based signal reconstruction algorithm to recover general frequency-sparse signals in the presence of linear impairments. The protocol requires only a small set of 1-sparse signals to achieve this level of performance, and the robustness of the LISTA reconstruction algorithm as we further lower SNR and increase sparsity suggests applications in other CS systems where manual calibration times are prohibitive. In the future, it will be of interest to extend this methodology to cope with weak and strong nonlinear impairments, and to demonstrate its efficacy on signal data collected from deployed physical systems.

7. ACKNOWLEDGEMENTS

This work was partially supported by NSF CCF 1733857 and NSF EARS 1343282. SB acknowledges support from the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

8. REFERENCES

[1] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[3] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] David L Donoho, "For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[5] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio : State-of-the-art and recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, May 2012.

[6] T. Haque, M. Bajor, Y. Zhang, J. Zhu, Z. Jacobs, R. Kettlewell, J. Wright, and P. R. Kinget, "A direct RF-to-information converter for reception and wideband interferer detection employing pseudo-random lo modulation," in *Proc. IEEE Radio Frequency Integrated Circuits Symp. (RFIC)*, June 2017, pp. 268–271.

[7] D. Adams, Y. Eldar, and B. Murmann, "A mixer frontend for a four-channel modulated wideband converter with 62 dB blocker rejection," in *Proc. IEEE Radio Frequency Integrated Circuits Symp. (RFIC)*, May 2016, pp. 286–289.

[8] R. T. Yazicigil, T. Haque, M. R. Whalen, J. Yuan, J. Wright, and P. R. Kinget, "A 2.7-to-3.7GHz rapid interferer detector exploiting compressed sampling with a quadrature analog-to-information converter," in *Proc. IEEE Int. Solid-State Circuits Conf. - (ISSCC) Digest of Technical Papers*, Feb. 2015, pp. 1–3.

[9] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[10] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Proc. (MoMuC '99) 1999 IEEE Int Mobile Multimedia Communications Workshop*, 1999, pp. 3–10.

[11] Y Chi, L L Scharf, A Pezeshki, and A R Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[12] Marco F Duarte and Richard G Baraniuk, "Spectral compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 111–129, July 2013.

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

[14] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6215–6221, Sept. 2011.

[15] Karol Gregor and Yann LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 399–406.

[16] Ulugbek S Kamilov and Hassan Mansour, "Learning optimal nonlinearities for iterative thresholding algorithms," *IEEE Signal Process. Lett.*, , no. 5, pp. 747–751, Dec. 2015.

[17] Bo Xin, Yizhou Wang, Wen Gao, David Wipf, and Baoyuan Wang, "Maximal sparsity with deep networks?," in *Advances in Neural Information Processing Systems 29*, D D Lee, M Sugiyama, U V Luxburg, I Guyon, and R Garnett, Eds., pp. 4340–4348. Curran Associates, Inc., 2016.

[18] Debabrata Mahapatra, Subhadip Mukherjee, and Chandra Sekhar Seelamantula, "Deep sparse coding using optimized linear expansion of thresholds," *arXiv preprint arXiv:1705.07290*, May 2017.

[19] Mark Borgerding, Philip Schniter, and Sundeep Rangan, "AMP-Inspired deep networks for sparse linear inverse problems," Dec. 2016.

[20] Hao He, Bo Xin, and David Wipf, "From bayesian sparsity to gated recurrent nets," *arXiv preprint arXiv:1706.02815*, June 2017.

[21] H Palangi, R Ward, and L Deng, "Distributed compressive sensing: A deep learning approach," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4504–4518, 2016.

[22] P Sprechmann, A M Bronstein, and G Sapiro, "Learning efficient sparse and low rank models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1821–1833, Sept. 2015.

[23] T. Haque, R. T. Yazicigil, K. J. L. Pan, J. Wright, and P. R. Kinget, "Theory and design of a quadrature analog-to-information converter for energy-efficient wideband spectrum sensing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 2, pp. 527–535, Feb. 2015.

[24] Scott Chen and David L Donoho, "Examples of basis pursuit," in *Wavelet Applications in Signal and Image Processing III*. Sept. 1995, vol. 2569, pp. 564–575, International Society for Optics and Photonics.

[25] A Beck and M Teboulle, "A fast iterative Shrinkage-Thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[26] Neal Parikh and Stephen Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[27] Roman Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[28] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser Basel, 2013.

[29] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, and Others, "TensorFlow: A system for Large-Scale machine learning," in *OSDI*. 2016, vol. 16, pp. 265–283, usenix.org.

[30] Alekh Agarwal, Sahand Negahban, and Martin J Wainwright, "Fast global convergence of gradient methods for high-dimensional statistical recovery," *Ann. Stat.*, vol. 40, no. 5, pp. 2452–2482, Oct. 2012.