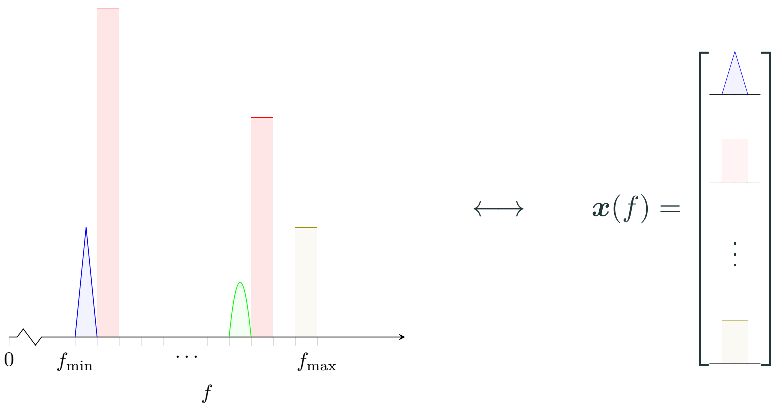
Efficient Model-Free Learning to Overcome Hardware Nonidealities in Analog-to-Information Converters

Sam Buchanan
Department of Electrical Engineering, Data Science Institute
Columbia University

Joint with **Tanbir Haque** (Columbia EE, InterDigital),
Peter Kinget (Columbia EE), **John Wright** (Columbia EE, Columbia DSI)

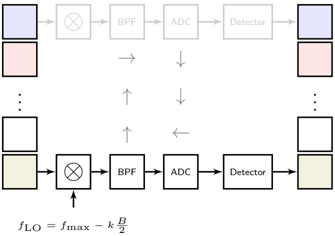
Multiband RF Spectrum Sensing

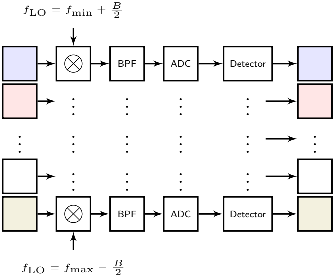
Spectrum sensing: efficiently determine the frequency content of an unknown real signal $x(t)$.



Assume there are n bands of width B in the partition.

Classical Spectrum Sensing Receiver Designs





With ideal components, a linear system: $\boldsymbol{y}(f) = \boldsymbol{A}\boldsymbol{x}(f)$

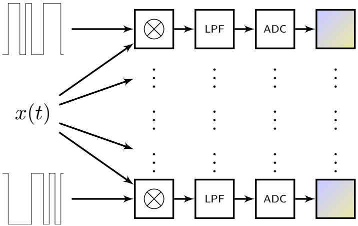
To first order, the receivers have **equal energy consumption**.

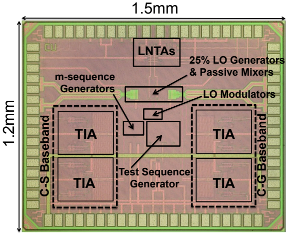
Compressed Sensing Receivers for Spectrum Sensing

Compressed sensing[1]: invert underdetermined $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ via

- **Structured** input $\boldsymbol{x}$,
- **Generic** sensor $\boldsymbol{A}$,
- Convex optimization for reconstruction.

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad \|\boldsymbol{x}\|_1 \qquad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \leq \delta$$





1: [Candès, Romberg, and Tao 2006; Candes and Tao 2005; Donoho 2006a; Donoho 2006b]

Reconstruction of the Spectrum via Proximal Gradient

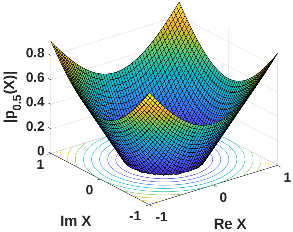
A recipe from numerical optimization: solve

$$\underset{\boldsymbol{x}\in\mathbb{C}^n}{\text{minimize}} \quad \lambda\|\boldsymbol{x}\|_1 + \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

via the **proximal gradient method**:

$$\boldsymbol{z}^{(k)} \leftarrow \boldsymbol{x}^{(k)} - \tfrac{1}{L}\boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{y})$$
$$\boldsymbol{x}^{(k+1)} \leftarrow e^{\hat{\imath}\arg(\boldsymbol{z}^{(k)})}\max\{\textstyle\sum_i |z_i^{(k)}|\boldsymbol{e}_i - \tfrac{\lambda}{L}\mathbf{1}, \mathbf{0}\}$$

for $k = 0, 1, \ldots$, $L = \|\boldsymbol{A}\|^2$, $\boldsymbol{x}^{(0)} = \mathbf{0}$ (say).



Characteristics of numerical optimization approaches to the recovery problem:

- Targeted towards **general** classes of problems.
- Accompanied by **worst-case** performance guarantees.
- **Hand-designed** using known models for the application.

$$\inf \underset{\text{(smooth)}}{f} + \underset{\text{(convex)}}{g} = h$$

$$h(\boldsymbol{x}^{(k)}) - h(\boldsymbol{x}_\star) = O(1/k)$$

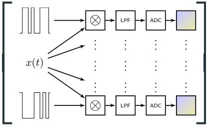
$$\boldsymbol{x} \text{ sparse} \implies g = \|\cdot\|_1$$

Reconstruction Algorithms—Desiderata

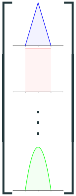
On the other hand, for spectrum sensing, we would like an algorithm that satisfies:

- Performance-optimized for a **specific** subclass of sparse recovery problems.
- Able to incorporate **hard constraints** on computational resources.
- **Adaptive** to deviations from the nominal design in a **model-free** manner.

$\boldsymbol{A} =$



$\boldsymbol{x}(f) =$

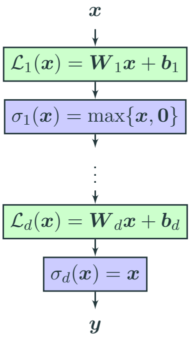


A neural network consists of:

1. An integer $d \geq 1$ (the **depth**)
2. Euclidean spaces $\mathbb{R}^{n_i}$, $i = 0, \ldots, d$
3. **Affine maps** $\mathcal{L}_i : \mathbb{R}^{n_{i-1}} \to \mathbb{R}^{n_i}$, $i \in [d]$
4. **"Nonlinearities"** $\sigma_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$, $i \in [d]$

Iteratively: define

$$f^{(k)} = \begin{cases} \sigma_k(\mathcal{L}_k(f^{(k-1)})) & k \geq 1 \\ \boldsymbol{I}_{n_0} & k = 0 \end{cases}$$



Then $\boldsymbol{y} = f^{(d)}(\boldsymbol{x})$.

For spectrum sensing: which architecture do we choose?

Recall the iteration

$$\boldsymbol{z}^{(k)} \leftarrow \boldsymbol{x}^{(k)} - \tfrac{1}{L}\boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{y})$$
$$\boldsymbol{x}^{(k+1)} \leftarrow e^{\hat{\imath}\arg(\boldsymbol{z}^{(k)})}\max\{\textstyle\sum_i |z_i^{(k)}|\boldsymbol{e}_i - \tfrac{\lambda}{L}\mathbf{1}, \mathbf{0}\}.$$

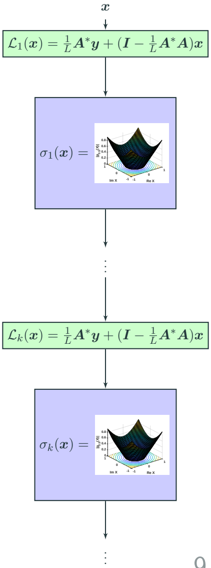
Define

$$p_\lambda(\boldsymbol{x}) = e^{\hat{\imath}\arg(\boldsymbol{x})}\max\{\sum_i |x_i|\boldsymbol{e}_i - \tfrac{\lambda}{L}\mathbf{1}, \mathbf{0}\}.$$

Then after rearranging:

$$\boldsymbol{x}^{(k+1)} \leftarrow p_\lambda\left(\tfrac{1}{L}\boldsymbol{A}^*\boldsymbol{y} + (\boldsymbol{I} - \tfrac{1}{L}\boldsymbol{A}^*\boldsymbol{A})\boldsymbol{x}^{(k)}\right),$$

which has the form of a neural network.

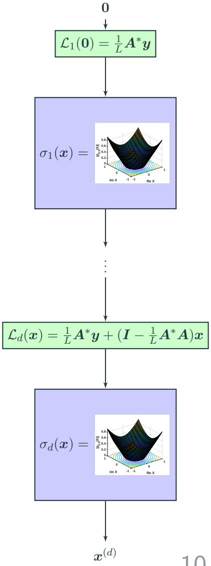


Learned Optimization Algorithms for Spectrum Sensing

Truncate the network, and **learn** its parameters using data. The result satisfies:

1. Optimized for a **specific** computational budget and sparse inverse problem.
2. **Adaptive** to nonidealities via the training procedure, without explicit modeling.
3. Leverages available **prior information** via the network topology and initializations.

This approach is called LISTA [Gregor and LeCun 2010].



Other works on learned optimization algorithms:

- (Near-)SoA in super-resolution [Wang et al. 2015], voice identification [Sprechmann, A. Bronstein, et al. 2013]
- Additional structural priors [Sprechmann, A. M. Bronstein, and Sapiro 2015]
- Other base recovery algorithms: AMP [Borgerding, Schniter, and Rangan 2016], SBL [He, Xin, and Wipf 2017]
- Co-optimization of nonlinearities and weights [Kamilov and Mansour 2015; Mahapatra, Mukherjee, and Seelamantula 2017]

Learned Optimization Training Procedure

Reparametrize the network. Then perform empirical risk minimization

$$\underset{\boldsymbol{W},\boldsymbol{S}}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^{T}\left\|\boldsymbol{x}^{(d)}(\boldsymbol{y}_i) - \overline{\boldsymbol{x}_i}\right\|_2^2$$

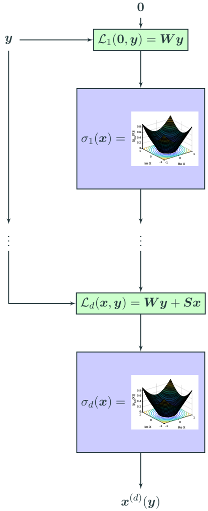
using

1. a dataset $\{\boldsymbol{y}_i, \overline{\boldsymbol{x}}_i\}_{i=1}^{T}$
2. initializations based on the design matrix $\widetilde{\boldsymbol{A}}$

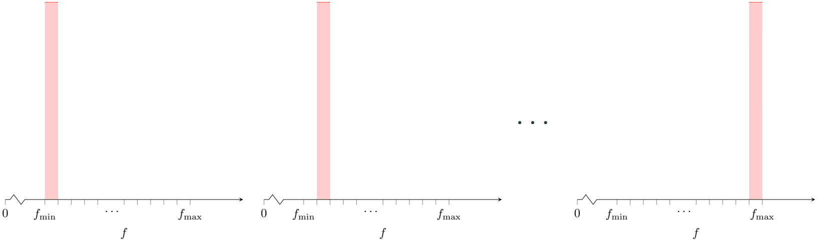
to obtain $(\boldsymbol{W}_\star, \boldsymbol{S}_\star)$.

We desire that:

1. T is not too large
2. Recovery guarantees associated with $\widetilde{\boldsymbol{A}}$ are associated with $(\boldsymbol{W}_\star, \boldsymbol{S}_\star)$



LISTA for Multiband Spectrum Sensing - Training Set



Proposed dataset: measurements $\boldsymbol{y}$ and corresponding spectra $\boldsymbol{x}$ for each possible 1-sparse $x(t)$, at moderate power.

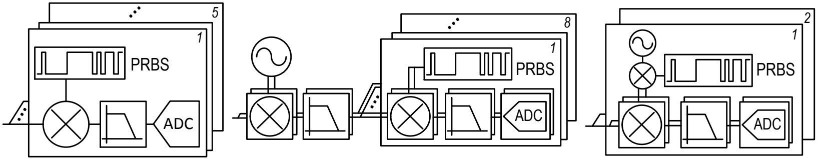
- Compare to

$$\sum_{k=1}^{\lceil m/\log n \rceil} \binom{n}{k}$$

  signals for exhaustive training.

Test on arbitrary k-sparse signals with equal support powers.

Experiments: Simulation Setup



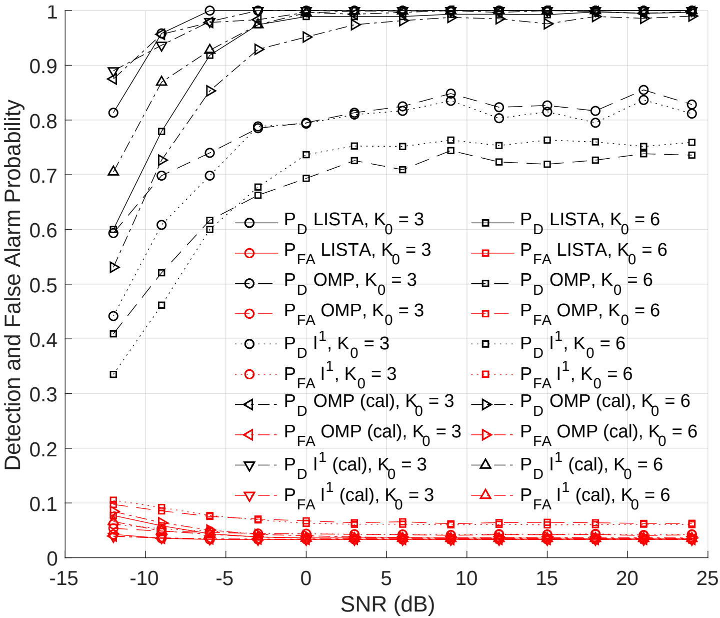
We simulate the DRF2IC receiver [Haque et al. 2018], and $\boldsymbol{A} \in \mathbb{C}^{18\times 63}$.

We consider two realistic receiver nonidealities:

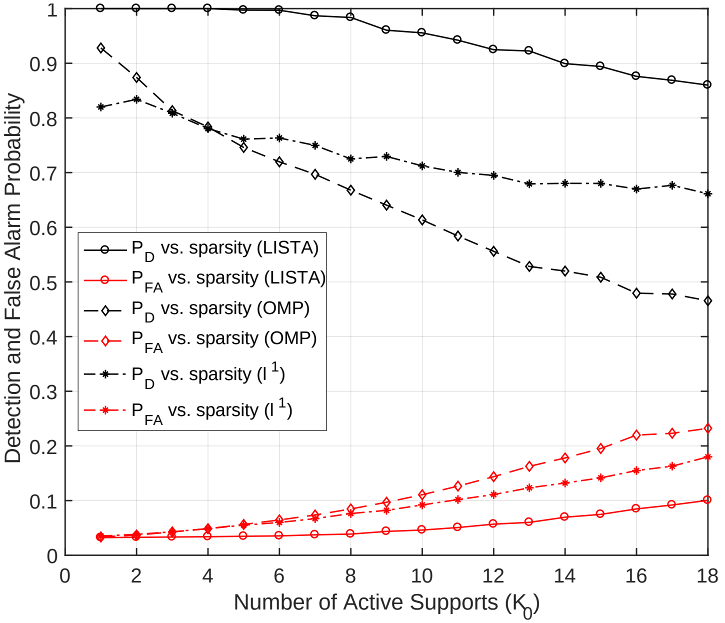
1. IQ downconversion frequency-independent mismatch
2. PRBS fractional phase offsets

Task: support recovery; evaluation metric: P_{D} and P_{FA}

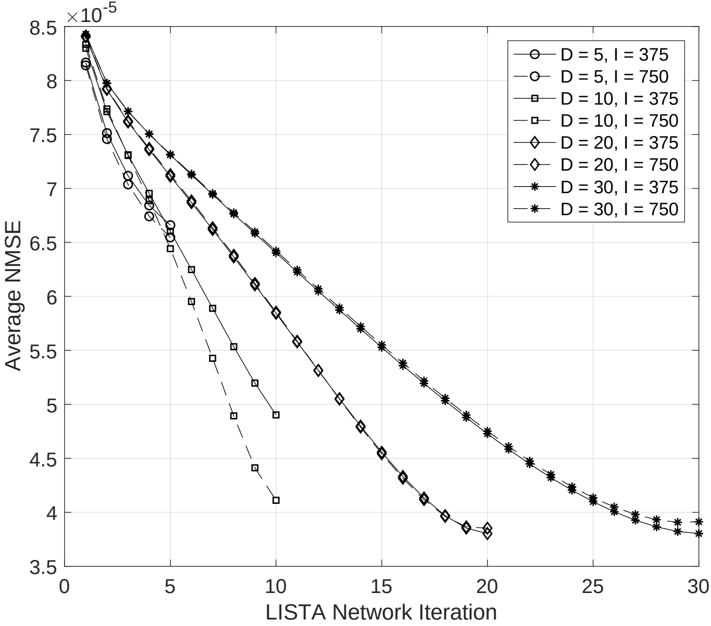
Experiments: Detection vs. SNR



Experiments: Detection vs. Sparsity



Experiments: Test Set NMSE vs. Training Parameters



Conclusions

1. Learned optimization-based signal recovery methodology was argued to be **ideal** for interferer detection in CS-based spectrum sensing receivers.
2. An efficient training protocol was developed and evaluated in the equal-power and linearly-impaired regime.
3. Extensions to more challenging signal models and receiver impairments will require additional innovations.

Tanbir Haque
Columbia U., InterDigital

Peter Kinget
Columbia U.

John Wright
Columbia U.

Thanks! Questions?

References I

Borgerding, Mark, Philip Schniter, and Sundeep Rangan (2016). "AMP-Inspired Deep Networks for Sparse Linear Inverse Problems". In: arXiv: `1612.01183 [cs.IT]`.

Candes, E J and T Tao (2005). "Decoding by linear programming". In: *IEEE Trans. Inf. Theory* 51.12, pp. 4203–4215.

Candès, Emmanuel J, Justin Romberg, and Terence Tao (2006). "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information". In: *IEEE Trans. Inf. Theory* 52.2, pp. 489–509.

Donoho, D L (2006a). "Compressed sensing". In: *IEEE Trans. Inf. Theory.*

Donoho, David L (2006b). "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution". In: *Commun. Pure Appl. Math.* 59.6, pp. 797–829.

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

Haque, T et al. (2018). "A Reconfigurable Architecture Using a Flexible LO Modulator to Unify High-Sensitivity Signal Reception and Compressed-Sampling Wideband Signal Detection". In: *IEEE J. Solid-State Circuits*, pp. 1–15.

He, Hao, Bo Xin, and David Wipf (2017). "From Bayesian Sparsity to Gated Recurrent Nets". In: arXiv: `1706.02815 [cs.LG]`.

Kamilov, Ulugbek S and Hassan Mansour (2015). "Learning Optimal Nonlinearities for Iterative Thresholding Algorithms". In: *IEEE Signal Process. Lett.* 5, pp. 747–751. arXiv: `1512.04754 [cs.LG]`.

Mahapatra, Debabrata, Subhadip Mukherjee, and Chandra Sekhar Seelamantula (2017). "Deep Sparse Coding Using Optimized Linear Expansion of Thresholds". In: arXiv: `1705.07290 [cs.LG]`.

Sprechmann, P, A M Bronstein, and G Sapiro (2015). "Learning Efficient Sparse and Low Rank Models". en. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 37.9, pp. 1821–1833.

Sprechmann, P., A. Bronstein, et al. (2013). "Learnable low rank sparse models for speech denoising". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 136–140.

References II

Wang, Zhaowen et al. (2015). “Deep Networks for Image Super-Resolution with Sparse Prior”. In: arXiv: 1507.08905 [cs.CV].

IRE Selection—Proximal Gradient (MMV)

A recipe from numerical optimization: solve

$$\underset{\boldsymbol{X}\in\mathbb{C}^{n\times p}}{\text{minimize}} \quad \lambda\sum_{i=1}^{n}\|\boldsymbol{x}_i\|_2 + \tfrac{1}{2}\|\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{X}\|_F^2$$

via the *proximal gradient method*:

$$\boldsymbol{Z}^{(k)} \leftarrow \boldsymbol{X}^{(k)} - \tfrac{1}{L}\boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{X}^{(k)} - \boldsymbol{Y})$$

$$\boldsymbol{X}^{(k+1)} \leftarrow e^{\hat{\imath}\arg(\boldsymbol{Z}^{(k)})}\max\{|\boldsymbol{Z}^{(k)}| - \tfrac{\lambda}{L}\sum_{i=1}^{n}\|\boldsymbol{x}_i\|\boldsymbol{e}_i\otimes\boldsymbol{1}_p, \boldsymbol{0}\}$$

for $k = 1, 2, \ldots$, $L = \|\boldsymbol{A}\|^2$, $\boldsymbol{X}^{(0)} = \boldsymbol{0}$ (say).

Top-Level Experiment Parameters

All experiments focus on the MMV problem

1. $f_{\min} = 2.57$ GHz, $f_{\max} = 3.83$ GHz
2. Sampling frequency set for an OSR of 8
3. DRF2IC run with 2 physical branches and expansion factor of 9
4. $m = 18$, $n = 63$, $p = 80$
5. IQ imbalance of 3 dB, 30°
6. PN sequence phase misalignment of $6/8$

LISTA parameters: $d = 12$, $T = 63$, $\lambda = 0.1$. Data is labeled via a pseudoinverse on the known support.

Support recovery details:

1. OMP: terminate after $k + 2$ iterations
2. ℓ^1 and LISTA: compute the average bin power in the predicted signal, and declare the top $k + 2$ bins to be active

First Experiment Details

1. Number of samples per point: 200
2. Sparsity levels: 3, 6
3. Samples are uniformly random k-sparse signals, plus independent noise
4. Noise power calculated over the entire band
5. Orthogonal matching pursuit run with only the previously-mentioned stopping criterion
6. ℓ^1 implementation is an accelerated proximal gradient algorithm
7. Calibrated traces run the corresponding algorithms on unimpaired data

Second Experiment Details

1. Number of samples per point: 500
2. SNR is fixed at 20 dB
3. Signal generation and algorithms as in the previous experiment

Third Experiment Details

1. Number of samples per point: 500
2. NMSE calculated similarly to the LISTA training objective (just with normalization)
3. SNR is fixed at 20 dB, sparsity is fixed at $k = 6$
4. Signal generation as in the previous experiment