Towards Neurally Augmented ALISTA

Freya Behrens (f.behrens@campus.tu-berlin.de), Jonathan Sauder (sauder@campus.tu-berlin.de), Peter Jung (peter.jung@tu-berlin.de)

https://arxiv.org/abs/2010.01930

Abstract

It is well-established that many iterative sparse reconstruction algorithms such as ISTA can be unrolled to yield a learnable neural network for improved empirical performance. Recently, ALISTA has been introduced, combining the strong empirical performance of a fully learned approach like LISTA, while retaining theoretical guarantees of classical compressed sensing algorithms and significantly reducing the number of parameters to learn. However, these parameters are trained to work in expectation, often leading to suboptimal reconstruction of individual targets. In this work we therefore introduce Neurally-Augmented-ALISTA, which computes step sizes and thresholds individually for each target vector during reconstruction. This adaptive approach is theoretically motivated by revisiting the recovery guarantees of ALISTA and is able to outperform existing algorithms in sparse reconstruction.

Problem

We consider the compressed sensing problem, i.e. reconstruction of a sparse vector from far fewer observations.

Given: $\Sigma_s^N := \{x \in \mathbb{R}^N \,|\, \|x\|_0 \leq s\}$, $x^* \in \Sigma_s^N$

$$\Phi \in \mathbb{R}^{M\times N}, \quad M \ll N, \quad y = \Phi x^*$$

Solve:

$$\underset{x}{\operatorname{argmin}} \|x\|_0 \quad \text{s.t.} \quad y = \Phi x$$

Observation

In ALISTA [1], it was proved that the thresholds of the l1-proximal operator must be larger than the l1-error of the current reconstruction. In previous literature, the thresholds are learned over the entire dataset, and must thus be larger than the supremum of all l1-errors. However, too-large thresholds also incur larger errors via the error bound in [1].

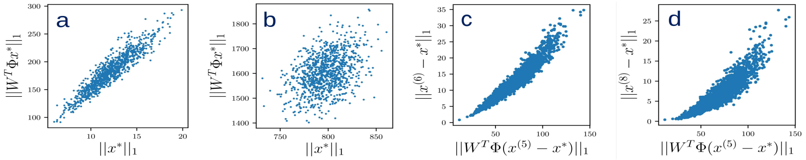
Idea

We can make thresholds adaptive to the current target using a good approximation of the l1 error. We find that:

$$r^{(k)} := \|\Phi x^{(k)} - y\|_1 = \|\Phi(x^{(k)} - x^*)\|_1$$

$$u^{(k)} := \|W^T(\Phi x^{(k)} - y)\|_1 = \|(W^T\Phi)(x^{(k)} - x^*)\|_1$$

Are good approximations and cheap to compute, as they are needed (besides norm computation) for the algorithm anyways.



Left: Correlation between $\|x^*\|_1$ and u for random Gaussian vectors $x^* \in \mathbb{R}^{1000}$ is strong for sparse and weak for non-sparse vectors $\|x^*\|_0=15$ in (a) and $\|x^*\|_0=1000$ in (b)). Also, the correlation between $u^{(i)}$ and the true error $\|x^{(j)} - x^*\|_1$ is even preserved over multiple layers for an instance of NA-ALISTA (c,d).

Neurally Augmented ALISTA

$$x^{(k+1)} = \eta_{\theta^{(k,x^*)}}\left(x^{(k)} - \gamma^{(k,x^*)} W^T(\Phi x^{(k)} - y)\right)$$

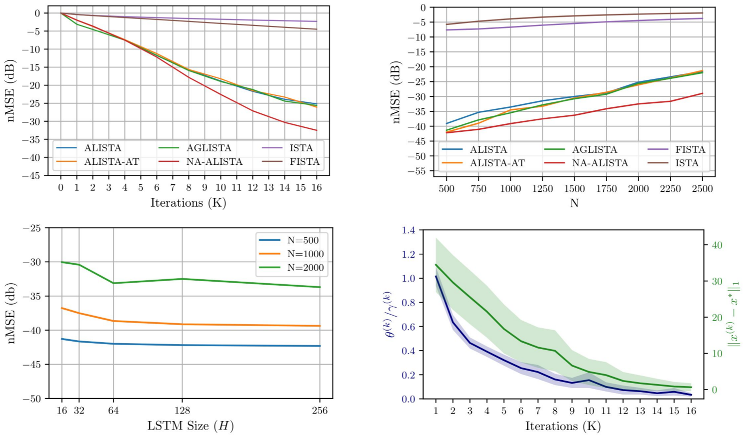
Top: Neurally Augmented ALISTA in mathematical form. $\gamma^{(k,x^*)}$ and $\theta^{(k,x^*)}$ are estimated by a neural network and adapt to x^*. W is computed to have optimal coherence with Φ, following the procedure in [1].
Bottom: Neurally Augmented ALISTA in algorithmic form.

Algorithm 1: Neurally Augmented ALISTA

Learnable Parameters:
initial cell state $c_0 \in \mathbb{R}^H$,
initial hidden state $h_0 \in \mathbb{R}^H$,
LSTM parameters
1-layer MLP parameters $U_1 \in \mathbb{R}^{H\times H}, U_2 \in \mathbb{R}^{2\times H}$
Input: y
$x \leftarrow 0$; $h \leftarrow h_0$; $c \leftarrow c_0$
for $\{1, \ldots, K\}$ **do**
  $r \leftarrow \|\Phi x - y\|_1$
  $u \leftarrow \|W^T(\Phi x - y)\|_1$
  $c, h \leftarrow \mathtt{LSTM}(c, h, [r, u])$
  $\theta, \gamma \leftarrow \mathtt{Softsign}(U_2(\mathtt{ReLU}(U_1 c)))$
  $x \leftarrow \eta_\theta\left(x - \gamma W^T(\Phi x - y)\right)$
end
Return x;

Experiments on Gaussian Synthetic Data

We compare NA-ALISTA with competitors (ISTA [2], FISTA[3], ALISTA-AT[4], AGLISTA[5]) in sparse reconstruction for M=250, S=50.
Top left: The reconstruction error over the number of iterations K for N=2000, SNR=40dB. *Top right* Reconstruction error over different compression ratios. *Bottom left:* Reconstruction error for LSTM size hidden layer size H for NA-ALISTA. *Bottom right:* Comparison of the ratio $\theta^{(k)}/\gamma^{(k)}$ with the true L1-error at each iteration for NA-ALISTA.



References

[1] Liu, Jialin, and Xiaohan Chen. "ALISTA: Analytic weights are as good as learned weights in LISTA." International Conference on Learning Representations (ICLR). 2019.
[2] Daubechies, Ingrid, Michel Defrise, and Christine De Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." Communications on Pure and Applied Mathematics. 2004.
[3] Beck, Amir, and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." SIAM journal on imaging sciences 2.1 (2009): 183-202.
[4] Kim, Dohyun, and Daeyoung Park. "Element-Wise Adaptive Thresholds for Learned Iterative Shrinkage Thresholding Algorithms." IEEE Access 8 (2020): 45874-45886.
[5] Wu, Kailun, et al. "Sparse Coding with Gated Learned ISTA." International Conference on Learning Representations. 2019.