Learned ISTA with Error-based Thresholding for Adaptive Sparse Coding

Ziang Li[1], Kailun Wu[2], Yiwen Guo[2] and Changshui Zhang[1]

[1] Department of Automation, Tsinghua University, Beijing, P.R.China, [2] Independent Researcher

Motivation

Iterative Shrinking Thresholding Algorithm (ISTA) [1]: Solve LASSO problem ($\min_x \|y - Ax\|_2 + \lambda\|x\|_1$).

Learned ISTA (LISTA) [2]: Achieve faster convergence by using deep neural network (DNN).

LISTA with Weight Coupling and Support Selection [3]: Achieve faster convergence.

Lack adaptability to the variety of training samples and robustness to outliers

This paper:

- Propose an error-based thresholding (EBT) mechanism to improve adaptivity.
- Require no extra learnable parameter.

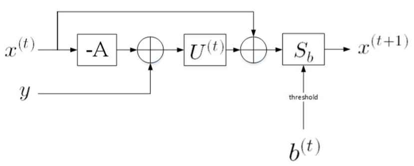
Methods

We propose to disentangle the reconstruction error term from the learnable part of the threshold and introduce adaptive thresholds for LISTA and related networks, i.e, something like $b^{(t)} = \rho^{(t)}\|x^{(t)} - x^*\|_\phi$.We found in noiseless cases, $Ax^{(t)} - y = A(x^{(t)} - x^*)$. Also, $U^{(t)}A$ approximates the identity matrix.

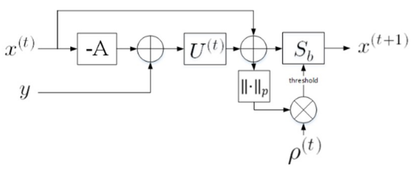
The update rule for EBT-LISTA:

$$\boldsymbol{x}^{(t+1)} = \boldsymbol{sh}_{\boldsymbol{b}^{(t)}}((\boldsymbol{I} - \boldsymbol{U}^{(t)}\boldsymbol{A})\boldsymbol{x}^{(t)} + \boldsymbol{U}^{(t)}\boldsymbol{y})$$

$$\boldsymbol{b}^{(t)} = \boldsymbol{\rho}^{(t)}\|\boldsymbol{U}^{(t)}(\boldsymbol{A}\boldsymbol{x}^{(t)} - \boldsymbol{y})\|_\phi$$



The t-th layer of LISTA



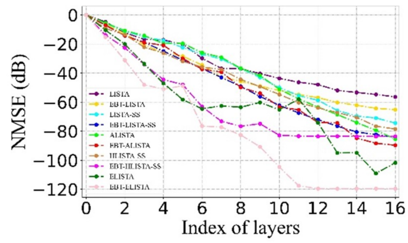
The t-th layer of EBT-LISTA

Theorem 1 EBT-LISTA converges similarly as the original LISTA. The convergence rate is faster and the reconstruction error is lower in a high probability.

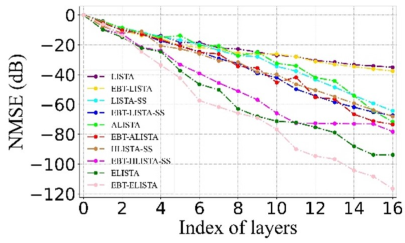
Theorem 2 LISTA with support selection [3] has two different convergence phases. EBT mechanism helps LISTA with support selection converge faster in the first phase and thus gets in the second phase faster.

Experiments

EBT can be combined with many LISTA-based methods and leads to significantly faster convergence, e.g., LISTA[3], LISTA-SS[3], ALISTA[4], ELISTA[5], and Hybrid LISTA-SS[6].

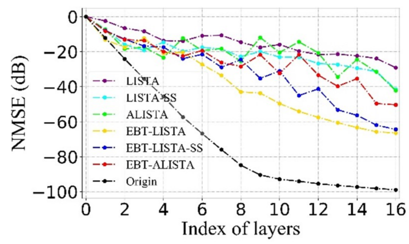


Sparsity = 0.95

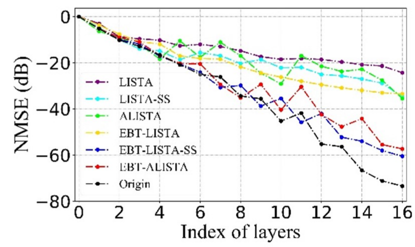


Sparsity = 0.9

We consider the adaptivity/generalization of the model. We let the test sparsity be different from the training sparsity. EBT has huge advantages in such scenario where the adaptivity to un-trained sparsity is required.

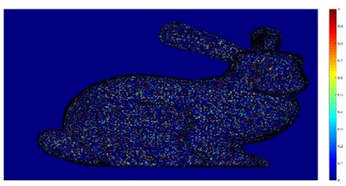


Sparsity changes from 0.8 to 0.99



Sparsity changes from 0.8 to 0.9

In photometric stereo analysis task, training sparsity is set as 0.8 and test sparsity is set as 0.9. The advantage of EBT-LISTA-SS is remarkable in such setting, which means EBT-based networks has better adaptivity.



LISTA−SS: ζ=0.0067

EBT−LISTA−SS: ζ=0.0026

Conclusion

- EBT mechanism helps LISTA and its variants converge faster and achieve superior final sparse coding performance.
- EBT mechanism endows deep unfolding models with higher adaptivity to different observations with a variety of sparsity.

Reference

[1] Ingrid Daubechies et al., “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.”
[2] Karol Gregor and Yann LeCun, “Learning fast approximations of sparse coding.”
[3] Xiaohan Chen et al., “Theoretical linear convergence of unfolded ista and its practical weights and thresholds.”
[4] Jialin Liu et al., “Alista: Analytic weights are as good as learned weights in lista.”
[5] Yangyang Li et al., “Learned extragradient ista with interpretable residual structures for sparse coding.”
[6] Ziyang Zheng et al., “Hybrid ista: unfolding ista with convergence guarantees using free-form deep neural networks.”

Scan for PDF