LISTA: Theoretical Linear Convergence, Practical Weights and Thresholds

Xiaohan Chen*, Jialin Liu†, Zhangyang Wang*, and Wotao Yin†

* TAMU, CSE
† UCLA, Math

NeurIPS'18
Recover \textbf{sparse} $x^*$ from

$$b := Ax^* + \text{white noise}$$

Our methods improve on LISTA (Gregor&LeCun’10) and related work by

- learning fewer parameters (\textbf{faster training})
- adding support detection (\textbf{faster recovery})
- proving linear convergence and robustness (\textbf{theoretical guarantee})
**ISTA** (iterative soft thresholding)

\[
x^{(k+1)} = \text{SoftThreshold}_\theta \left( x^{(k)} + \alpha A^T (b - Ax^{(k)}) \right).
\]

\(\alpha, \theta\) are chosen by hand or cross validation.

**LISTA** (Learned ISTA)

\[
x^{(k+1)} = \text{SoftThreshold}_{\theta^k} \left( W_{k1} b + W_{k2} x^{(k)} \right).
\]

\(\theta^k, W_{k1}, W_{k2}\) are chosen by stochastic optimization

\[
\min_{\theta^k, W_{k1}, W_{k2}} \left\{ E_{x^\star, b} \| x^K (b) - x^\star \|_2 \right\}
\]

using synthesized \((x^\star, b)\) obeying \(b = Ax^\star + \text{white noise}\).

Compare: ISTA is slow, no training. LISTA is fast, difficult-to-train.
**Review: ISTA and LISTA**

**ISTA** (iterative soft thresholding)

\[
x^{(k+1)} = \operatorname{SoftThreshold}_{\theta} \left( x^{(k)} + \alpha A^T (b - Ax^{(k)}) \right).
\]

\(\alpha, \theta\) are chosen by hand or cross validation.

**LISTA** (Learned ISTA)

\[
x^{(k+1)} = \operatorname{SoftThreshold}_{\theta^k} \left( W^k_1 b + W^k_2 x^{(k)} \right).
\]

\(\theta^k, W^k_1, W^k_2\) are chosen by stochastic optimization

\[
\minimize_{\{\theta^k, W^k_1, W^k_2\}} \left\{ \mathbb{E}_{x^*, b} \| x^K (b) - x^* \|^2 \right\}
\]

using synthesized \((x^*, b)\) obeying \(b = Ax^* + \text{white noise}\).
Review: ISTA and LISTA

**ISTA** (iterative soft thresholding)

\[ x^{(k+1)} = \text{SoftThreshold}_\theta \left( x^{(k)} + \alpha A^T (b - Ax^{(k)}) \right). \]

\( \alpha, \theta \) are chosen by hand or cross validation.

**LISTA** (Learned ISTA)

\[ x^{(k+1)} = \text{SoftThreshold}_{\theta_k} \left( W_1^k b + W_2^k x^{(k)} \right). \]

\( \theta^k, W_1^k, W_2^k \) are chosen by stochastic optimization

\[ \minimize_{\{\theta^k, W_1^k, W_2^k\}} \{ \mathbb{E}_{x^*, b} \| x^K (b) - x^* \|^2 \} \]

using synthesized \((x^*, b)\) obeying \(b = Ax^* + \text{white noise}\).

**Compare:** ISTA is slow, no training. LISTA is fast, difficult-to-train.
LISTA-CP: couple $W_1^k$ and $W_2^k$ via

$$W_1^k A + W_2^k = I.$$ 

We show: $x^{(k)} \rightarrow x^*$ implies this relation to hold asymptotically.
LISTA-CPSS: support selection

- Only the large coordinates pass activations to the next iteration.
- Ideas from Linearized Bregman iteration (kicking)\(^1\) and Fixed-Point Continuation method (FPC)\(^2\).

---

\(^1\)Stanley Osher et al. '2011
\(^2\)Elaine Hale, Wotao Yin, Yin Zhang '2008
Robust global linear convergence

**Theorem**

Fix $A$, sparsity level $s$, and noise level $\sigma$.

There exist $\{\theta^k, W_1^k\}$ such that LISTA-CP obeys

$$\|x^{(k)} - x^*\|_2 \leq sC_1 e^{-C_2 k} + C_3 \sigma, \quad k = 1, 2, \ldots$$

where $C_1, C_2, C_3 > 0$ are constants.

LISTA-CPSS improves the constants $C_2, C_3$. 


CP can stabilize intermediate results.

CP will not hurt final recovery performance.
Support selection test (no noise)

-70
-60
-50
-40
-30
-20
-10
0
LISTA
FISTA
AMP
LISTA
LAMP
LISTA-CP
LISTA-SS
LISTA-CPSS

LISTA and LISTA-CP
LISTA-SS
LISTA-CPSS
Thank you!

10:45 AM – 12:45 PM
Room 210 & 230 AB #163

Welcome to our poster for more details!