Learning step sizes for unfolded sparse coding

Pierre Ablin*, **Thomas Moreau***, **Mathurin Massias**, **Alexandre Gramfort**

Univ. Paris-Saclay, INRIA, Parietal team, Saclay, France. * Contributed Equally

Solving sparse linear inverse problems

- **Objective:** Solve the Lasso for a fixed D and $x \sim \mathbb{P}$.

$$z^*(x) = \underset{z}{\operatorname{argmin}} F_x(z) = \frac{1}{2}\|x - Dz\|_2^2 + \lambda \|z\|_1$$

- **Iterative algorithm:** Use ISTA for each x
- **Deep Learning:** Use a NN to learn a mapping

$$\Phi_\Theta^T : x \mapsto z^*(x); \quad \text{for } x \sim \mathbb{P}$$

Learned ISTA [Gregor & Le Cun 2010]

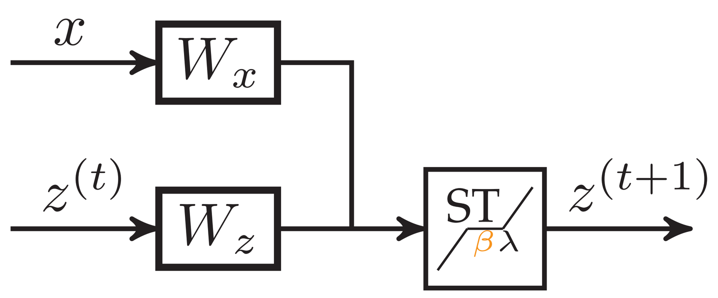
ISTA: $z^{(t+1)} = \mathrm{ST}(z^{(t)} - \gamma D^\top(Dz^{(t)} - x), \gamma\lambda)$,
where γ is the step size, usually chosen as $1/L$.

LISTA: Let $W_z = I_m - \gamma D^\top D$; $W_x = \gamma D^\top$ and $\beta = \gamma$

$$z^{(t+1)} = \mathrm{ST}(W_z z^{(t)} + W_x x, \lambda\beta)$$

Re-parametrization:

$W_z = I_m - \alpha W^\top D$; $W_x = \alpha W^\top$



- Learn parameters $\Theta = \{W^{(t)}, \alpha^{(t)}, \beta^{(t)}\}$

Supervised learning	Semi-supervised learning	Unsupervised learning
Ground truth available $s_1, \dots, s_N$	Compute $s_i = \operatorname{argmin}_z F_{x_i}(z)$	Learn to solve the Lasso
$\sum_{i=1}^{N} (\Phi_\Theta^T(x_i) - s_i)^2$	$\sum_{i=1}^{N} (\Phi_\Theta^T(x_i) - s_i)^2$	$\sum_{i=1}^{N} F_{x_i}(\Phi_\Theta^T(x_i))$

Local smoothness constants

- $L = \max \|Dz\|_2^2$ subject to $\|z\|_2 = 1$

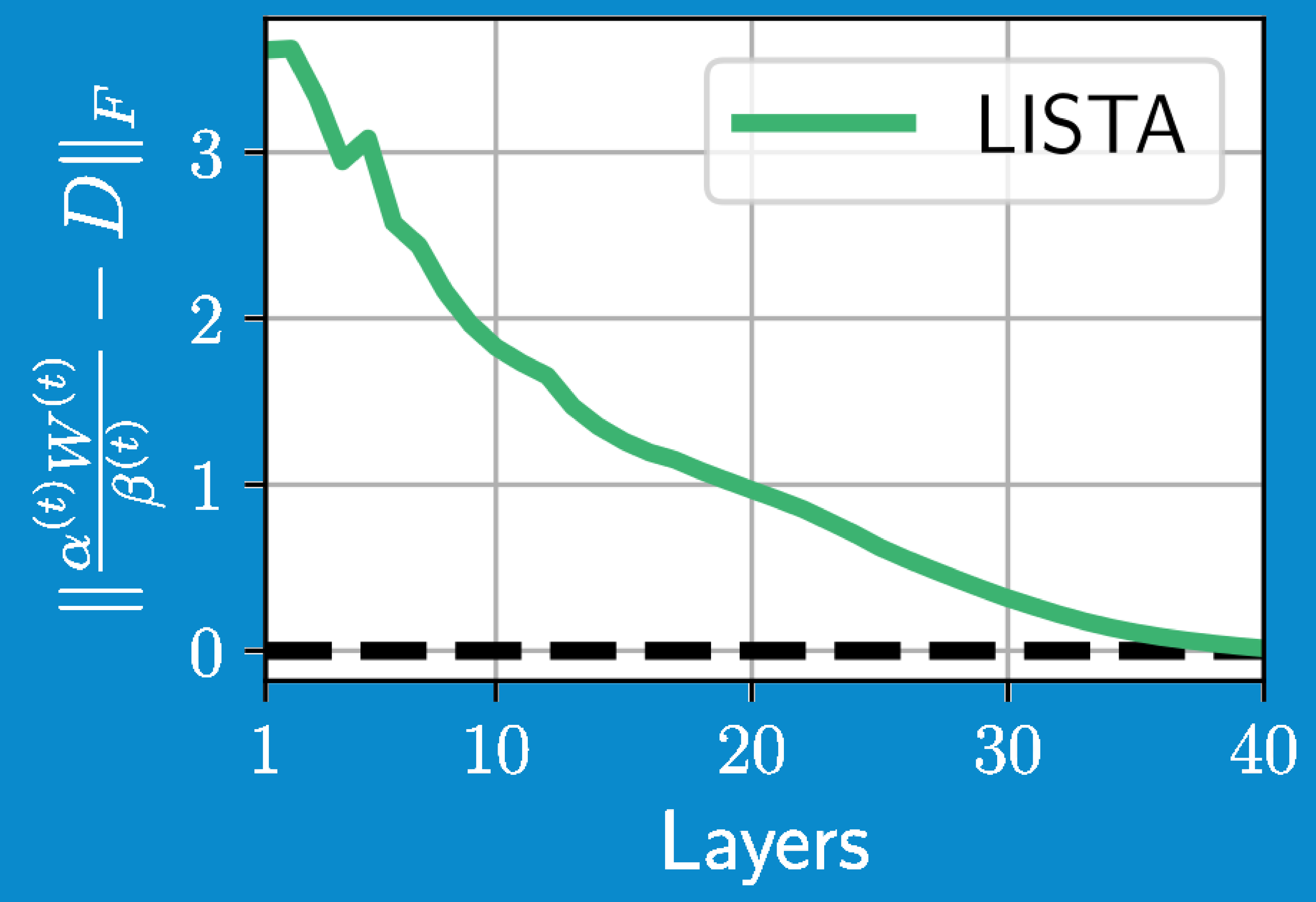
$$F_x(z) = f_x(z^{(t)}) + \langle \nabla f_x(z^{(t)}), z - z^{(t)} \rangle + \frac{1}{2}\|D(z - z^{(t)})\|_2^2 + \lambda\|z\|_1$$
$$\leq f_x(z^{(t)}) + \langle \nabla f_x(z^{(t)}), z - z^{(t)} \rangle + \frac{L}{2}\|z - z^{(t)}\|_2^2 + \lambda\|z\|_1$$

- $L_S = \max \|Dz\|_2^2$ subject to $\|z\|_2 = 1, Supp(z) \subset S$.

ISTA with **greater step-size:** $\gamma = 1/L_s$

Theorem: The weights of a neural network trained to solve the lasso asymptotically only learn a step size.

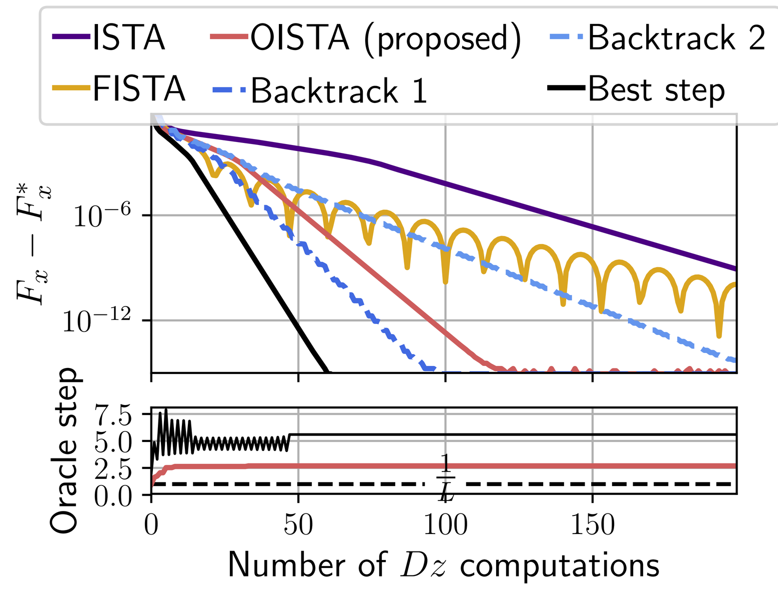
$$\frac{\alpha^{(t)}}{\beta^{(t)}} W^{(t)} \xrightarrow[t \to \infty]{} D$$

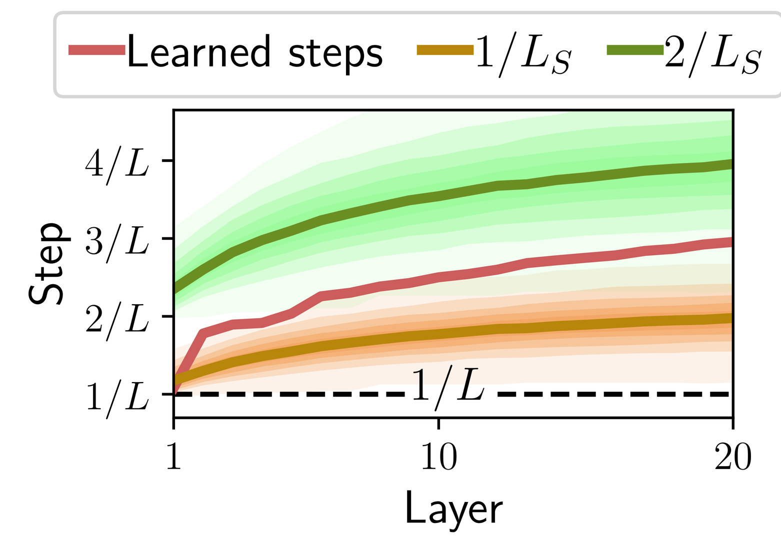


Improving ISTA step-size

Better step-sizes for ISTA

- Back-tracking line-search
- **OISTA:** Adapt step-sizes to **Local smoothness constants** L_S
- **SLISTA:** Learn only step-sizes with LISTA.

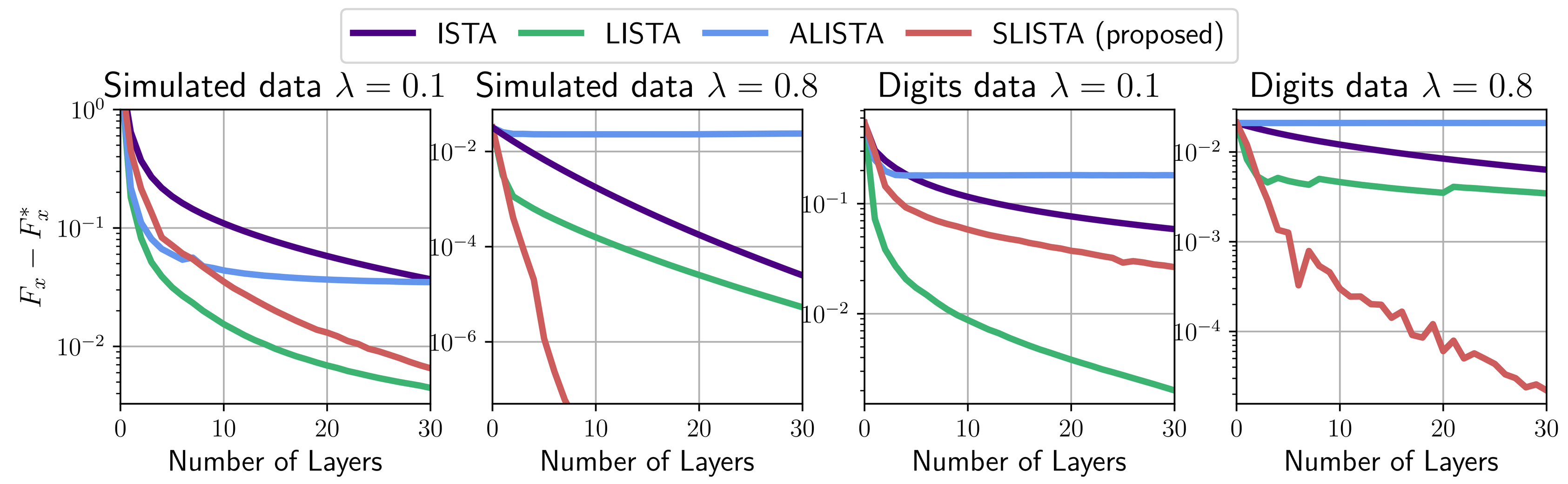




The step-sizes learned by SLISTA tend to be in $[\frac{1}{L_s}, \frac{2}{L_S}]$.

Varying the sparsity

SLISTA works better when z^* is sparse as this reduces L_S.



References

- Gregor, K. & Le Cun, Y. (2010) Learning Fast Approximations of Sparse Coding. ICML.
- Chen, X., Liu, J., Wang, Z. & Yin, W. (2018) Theoretical Linear Convergence of Unfolded ISTA and its Practical Weights and Thresholds. NeurIPS.
- Liu, J., Chen, X., Wang, Z. & Yin, W (2019). ALISTA: Analytic Weights are as good as Learned weigths in LISTA. ICLR.
- Moreau, T. & Bruna, J. (2017). Understanding Trainable Sparse Coding with Matrix Factorization. ICLR.