Learning step sizes for unfolded sparse coding

Pierre Ablin*, Thomas Moreau*, Mathurin Massias, Alexandre Gramfort
Univ. Saclay, INRIA, Parietal team, Saclay, France. * Contributed Equally

Solving sparse linear inverse problems

- Objective: Solve the Lasso for a fixed $D$ and $x \sim P$.
  \[ z^*(x) = \arg\min_z F_x(z) = \frac{1}{2} \| x - Dz \|_2^2 + \lambda \| z \|_1 \]
- Iterative algorithm: Use ISTA for each $x$
- Deep Learning: Use a NN to learn a mapping
  \[ \Phi^*_t : x \mapsto z^*(x) ; \text{ for } x \sim P \]

Learned ISTA

[Gregor & Le Cun 2010]

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

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

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

Re-parametrization:

$W_z = I_m - \alpha W D^T D$; $W_z = \alpha W^T$

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

- Supervised learning
  - Ground truth available $s_1, \ldots, s_N$
  - $\sum_{i=1}^N (\Phi^T_0(x_i) - s_i)^2$
- Semi-supervised learning
  - Compute $s_i = \arg\min_z F_{x_i}(z)$
  - $\sum_{i=1}^N (\Phi^T_0(x_i) - s_i)^2$
- Unsupervised learning
  - Learn to solve the Lasso
  - $\sum_{i=1}^N F_{x_i}(\Phi_0^T(x_i))$

Local smoothness constants

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

- $F_x(z) = f_x(z^{(0)}) + (\nabla f_x(z^{(0)}), z - z^{(0)}) + \frac{1}{2} \| D(z - z^{(0)}) \|_2^2 + \lambda \| z \|_1$

  $\leq f_x(z^{(0)}) + (\nabla f_x(z^{(0)}), z - z^{(0)}) + \frac{1}{2} \| z - z^{(0)} \|_2^2 + \lambda \| z \|_1$

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

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.

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

Improving ISTA step-size

Better step-sizes for ISTA

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

Varying the sparsity

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

Local smoothness constants

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

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

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

References