Distributed soft thresholding for sparse signal recovery

Chiara Ravazzi, Sophie M. Fosson, Enrico Magli

Politecnico di Torino, ITALY
http://calvino.polito.it/~ravazzi

GTTI – Ancona – June 24-26, 2013
Compressed sensing in sensor networks

Compressed Sensing
- technique for nonadaptive compressed acquisition;
- reconstruction from few linear measurements (exploiting signal’s sparsity)

Distributed Compressed Sensing paradigm
- distributed compressed acquisition in a sensor network
- centralized reconstruction in a fusion center
- drawbacks: energy utilization, delays, robustness, privacy

Our goal: distributed reconstruction (no fusion center)
- distribute the reconstruction task over the network
- cope with sensors’ limited computational power and memory
Problem Statement

Model

- a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - nodes in $\mathcal{V} \rightsquigarrow$ sensors
  - edges in $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V} \rightsquigarrow$ available communication links

- set of observations
  $$y_v = A_v x_0 + \xi_v \quad v \in \mathcal{V}$$
  - $x_0 \in \Sigma_k = \{x \in \mathbb{R}^n : |\text{supp}(x)| \leq k\}$ unknown signal
  - $A_v \in \mathbb{R}^{m \times n}$ with $m |\mathcal{V}| \ll n$
  - $\xi_v$ independent gaussian noise

Classical Approach: LASSO regression problem

- all data $(y_v, A_v)$ in the network collected in a fusion center
- reconstruction solving the convex optimization problem
  $$\hat{x}_{LASSO} = \arg\min_{x} \sum_{v \in \mathcal{V}} \|y_v - A_v x\|_2^2 + 2\alpha \|x\|_1, \quad \alpha > 0$$
Algorithms for distributed reconstruction

**DSM - Distributed subgradient methods**

+ Averaging the estimation of neighbouring nodes
+ Low-memory requirement
  - Convergence not guaranteed
  - 'Stopped model': number of iterations for subgradient computation critical

**ADMM - Alternating direction of multipliers**

+ Averaging the estimation of neighbouring nodes
+ Convergence is guaranteed to the LASSO solution
  - High-memory requirement: at each node need to store $n \times n$ matrix ($n$ signal length)
Distributed Iterative Soft Thresholding

**DISTA**: Given \( x_v(0) = 0 \), iterate

- for \( t \in 2\mathbb{N} \)
  \[
  \bar{x}_v(t + 1) = \sum_{w \in \mathcal{V}} P_{v,w} x_w(t), \quad x_v(t + 1) = x_v(t)
  \]
  consensus step

- for \( t \in 2\mathbb{N} + 1 \)
  \[
  \bar{x}_v(t + 1) = \bar{x}_v(t)
  \]
  \[
  x_v(t + 1) = \eta_\alpha \left[ (1 - \gamma) \sum_{w \in \mathcal{V}} P_{v,w} \bar{x}_w(t) + \gamma \left( x_v(t) + \tau A_v^T(y_v - A_v x_v(t)) \right) \right]
  \]
  consensus step + gradient step

**Consensus-based subgradient algorithm:**

- **consensus** step: doubly-stochastic matrix \( P = P(\mathcal{G}) \)
- **gradient** step balanced by \( \gamma \in (0, 1) \)
- sparsity-inducing operator:
  \[
  \eta_\alpha[x] = \begin{cases} 
  \text{sgn}(x)(|x| - \alpha) & \text{if } |x| > \alpha \\
  0 & \text{otherwise.}
  \end{cases}
  \]
Theoretical results

Hp: $d$-regular and connected graphs, uniform weights

\[
F_{\gamma}(X) = \sum_{v \in V} \left( \gamma \|A_v x_v - y_v\|^2 + \frac{2\alpha}{\tau} \|x_v\|_1 + \frac{1 - \gamma}{d\tau} \sum_{w \in V} \|x_v - x_w\|^2 \right)
\]

Th: For any initial choice $x_v(0), \tau < \|A_v\|^{-2}$ for all $v \in V$, there exists $X^\gamma \in \mathbb{R}^{m \times |V|}$ such that

1. DISTA produces a sequence \( \{X(t) = (x_1(t), ..., x_{|V|}(t))\}_{t \in \mathbb{N}} \) such that
   \[
   \lim_{t \to \infty} \|X(t) - X^\gamma\|_F = 0
   \]

2. the limit point $X^\gamma = (x_1^\gamma, ..., x_{|V|}^\gamma)$ is a minimizer of $F_{\gamma}$.

3. if $\hat{x}_{LASSO}$ is the solution of the centralized LASSO, then
   \[
   \lim_{\gamma \to 0} x_v^\gamma = \hat{x}_{LASSO}
   \]
Numerical results

Experiments

$n$ signal length, $m$ number of measurements per node, $|\mathcal{V}|$ number of nodes

- signal: $\text{supp}(x)$ choosing $k$ components uniformly, then $x_i \sim \mathcal{N}(0,1), \forall i \in \text{supp}(x)$
- sensing matrix: $A_v(i,j) \sim \mathcal{N}(0,1/\sqrt{m})$
- consensus matrix: uniform weights

Declare success if

$$\text{MSE} = \sum_{v \in \mathcal{V}} \|x_0 - x^\gamma_v\|_F^2/(n|\mathcal{V}|) < 10^{-4}$$
Distributed vs centralized reconstruction

\[ n = 150, \ k = 15. \]
Reconstruction probability

\[ n = 150, \ k = 15. \]
Convergence times

$n = 150, k = 15.$
Concluding remarks

DISTA

• blending gradient methods + consensus techniques;
• variational characterization of provided estimate;
• convergence guaranteed
  ▶ suboptimal algorithm
  ▶ optimal for small temperature parameter ($\gamma \to 0$)
• good tradeoff memory/performance
  ▶ faster than DSM;
  ▶ low-memory requirement (STM32W microcontrollers with Contiki operating systems: ADMM $n \approx 60$, DSM and DISTA $n \approx 1360$)

Future developments

• asynchronous algorithms for distributed reconstruction (gossip randomized algorithms, ...)
• distributed algorithms for anomaly detection and classification from compressed measurements