Distributed soft thresholding for sparse signal recovery

Chiara Ravazzi, Sophie M. Fosson, Enrico Magli

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

GLOBECOM – Atlanta – December 9-13, 2013
Outline

- Compressed sensing
- Distributed compressed sensing
- Distributed algorithms
- Our proposed solution: DISTA
  - convergence
  - conditions for optimality
  - numerical comparisons
- Concluding remarks
- Future developments
Compressed sensing

- technique for nonadaptive compressed acquisition;

\[ A \in \mathbb{R}^{m \times n} \] random projections

\[ x_0 \in \mathbb{R}^n \] signal

\[ \xi \in \mathbb{R}^{m \times 1} \] noise

\[ y \in \mathbb{R}^{m \times 1} \] observation

- reconstruction from few linear measurements (exploiting signal’s sparsity)

\[ x_0 \in \Sigma_k = \{ x \in \mathbb{R}^n : |\text{supp}(x)| \leq k \}, k \ll n, m \ll n \]
Compressed sensing in sensor networks

Model

- a 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

$$y_v = A_v x_0 + \xi_v \quad \forall v \in \mathcal{V}$$

- $x_0 \in \Sigma_k = \{ x \in \mathbb{R}^n : |\text{supp}(x)| \leq k \}$ unknown signal
- $A_v \in \mathbb{R}^{m \times n}$ with $m|\mathcal{V}| \ll n$
- $\xi_v$ bounded noise
Distributed Compressed Sensing

Classical approach

• distributed compressed acquisition in a sensor network
• all data \((y_v, A_v)\) in the network collected in a fusion center
• least-absolute shrinkage and selection operator

\[
\hat{x}_{\text{LASSO}} = \arg\min\limits_{x} \sum_{v \in V} \| y_v - A_v x \|_2^2 + 2\alpha \| x \|_1, \quad \alpha > 0
\]

• convex optimization tools: interior-point methods, subgradient methods, iterative hard and soft thresholding, ...
• 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
Distributed algorithms

**Distributed Approach:** use the cooperation among nodes

- **DSM - Distributed subgradient methods**
  - Averaging the estimation of neighboring nodes
  - Low-memory requirement: \( O(n) \) real values per node
  - Convergence not guaranteed
  - ’Stopped model’: number of iterations for subgradient computation critical

- **ADMM - Alternating direction of multipliers**
  - Averaging the estimation of neighboring nodes
  - Convergence is guaranteed to the LASSO solution
  - High-memory requirement: \( O(n^2) \) real values per node

**Our goal:** distributed algorithm with

- convergence guarantees
- low-memory requirements as close as possible to that of DSM
- performance as close as possible to that of ADMM
Distributed Iterative Soft Thresholding

**DISTA**: Given \( x_v(0) = 0, \gamma \in (0, 1) \), iterate

- for \( t \in 2\mathbb{N} \)

\[
\bar{x}_v(t + 1) = \frac{1}{|\mathcal{N}_v|} \sum_{w \in \mathcal{N}_v} x_w(t), \quad x_v(t + 1) = x_v(t)
\]

**consensus step**

- for \( t \in 2\mathbb{N} + 1 \)

\[
x_v(t + 1) = \eta_\alpha \left[ (1 - \gamma) \frac{1}{|\mathcal{N}_v|} \sum_{w \in \mathcal{N}_v} \bar{x}_w(t) + \gamma \left( x_v(t) + \tau A_v^T (y_v - A_v x_v(t)) \right) \right]
\]

**gradient step**

**Low-memory requirements**: \( O(n) \) real values per node
- **consensus** step: averaging the estimations of neighboring nodes
- **gradient** step: minimizing the least square function
- **soft-thresholding step**: inducing sparsity in the estimation

\[
\eta_\alpha [x] = \begin{cases} 
\text{sgn}(x)(|x| - \alpha) & \text{if } |x| > \alpha \\
0 & \text{otherwise.}
\end{cases}
\]
Theoretical results

**Hp:** $d$-regular and connected graphs

$$
\mathcal{F}_\gamma(X) = \sum_{v \in \mathcal{V}} \left( \gamma \| A_v x_v - y_v \|_2^2 + \frac{2\alpha}{\tau} \| x_v \|_1 + \frac{1 - \gamma}{d \tau} \sum_{w \in \mathcal{N}_v} \| x_v - x_w \|_2^2 \right)
$$

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

1. DISTA produces a sequence $\{X(t) = (x_1(t), ..., x_{|\mathcal{V}|}(t))\}_{t \in \mathbb{N}}$ such that
   $$\lim_{t \to \infty} \| X(t) - X^\gamma \|_F = 0$$
2. the limit point $X^\gamma = (x_1^\gamma, ..., x_{|\mathcal{V}|}^\gamma)$ is a minimizer of $\mathcal{F}_\gamma$.
3. if $\hat{x}_{\text{LASSO}}$ is the solution of the centralized LASSO, then
   $$\lim_{\gamma \to 0} x_v^\gamma = \hat{x}_{\text{LASSO}}$$
Numerical results

Experiments

- $n$ signal length, $m$ number of measurements per node, $|\mathcal{V}|$ number of nodes
  - signal: $\text{supp}(x_0)$ choosing $k$ components uniformly, then $x_0^i \sim \mathcal{N}(0, 1), \forall i \in \text{supp}(x)$
  - sensing matrix: $A_v(i, j) \sim \mathcal{N}(0, 1/m)$
  - temperature parameter: $\gamma = 1/2$
  - topology: complete graphs

Declare success if

$$\text{MSE} = \sum_{v \in \mathcal{V}} \|x_0 - x_0^\gamma\|_2^2/(n|\mathcal{V}|) < 10^{-4}$$
Distributed vs centralized reconstruction

\[ n = 150, \ k = 15 \]

\[ m|\mathcal{V}| = 70. \]
Reconstruction probability

\[ n = 150, \quad k = 15. \]
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 compared to ADMM
Future developments

Distributed algorithms for $\ell_0/\ell_1$-regularized inverse problems

- localization problems
- anomaly detection and robust recovery from compressed measurements

Randomized algorithms for distributed sparse signal recovery

- consensus-based asynchronous algorithms (gossip, ...)
- diffusion-based algorithms (incremental methods, ...)

Sparsity constrained nonlinear optimization

- centralized vs distributed computation
- conditions for optimality
Recommended reading

This paper and companion papers

- C. Ravazzi, S. M. Fosson, E. Magli, *Distributed soft thresholding for sparse signal recovery*, Globecom 2013;
- C. Ravazzi, S. M. Fosson, E. Magli, *Distributed iterative thresholding for $\ell_0/\ell_1$-regularized linear inverse problems*, submitted to Trans. on Information Theory, 2013;

Other works on distributed algorithms for sparse signal recovery

- S. Patterson, Y.C. Eldar, I. Keidar, *Distributed compressed sensing for static and time-varying networks*, submitted to Trans. on Signal Processing, 2013