BAYESIAN TUNING FOR SUPPORT DETECTION AND SPARSE SIGNAL ESTIMATION VIA ITERATIVE SHRINKAGE-THRESHOLDING

POLITECNICO DI TORINO
Department of Electronics and Telecommunications

CRISP
Towards Compressive Information Processing Systems
erc

Chiara Ravazzi and Enrico Magli
Politecnico di Torino (DET), Italy
41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20th – 25th March 2016

1 -- A Bayesian view of sparse signal recovery via Lasso

Problem: sparse signal recovery from compressed data

- unknown sparse signal: $x^\star \in \mathbb{R}^n$ with $|\{i : x_i^\star \neq 0\}| \leq k \ll n$
- set of observations:

$$y = Ax^\star + \eta$$

- sensing matrix: $A \in \mathbb{R}^{m\times n}$ with $m \ll n$
- white gaussian noise: $\eta \in \mathbb{R}^m$
- the signal prior is not known

Lasso solution:

$$\min \tfrac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1, \ \lambda > 0$$

- equivalent to a ML estimation, assuming $x_i^\star$ i.i.d. Lap$(0, \lambda^{-1})$:
- is invariant to the signal prior

Aim: new methods for learning sparsity models able to
- ▷ reduce the number of measurements required for reconstruction
- ▷ improve the computational complexity in the estimation

2 -- This work: weighted Lasso via 2LMM

Our solution: new parametric model for the signal able to:
- ▷ **combine support detection and estimation**
- ▷ **preserve the simplicity coming from Laplace assumption**

2LMM: x is a random variable with i.i.d. entries

$$x_i = z_i u_i + (1 - z_i)v_i,$$
$$u_i \sim \text{Lap}(0, \alpha), v_i \sim \text{Lap}(0, \beta), \quad 0 \approx \alpha < \beta$$
$$z_i \sim \text{Ber}(1 - p) \quad p \ll 1/2$$

- ML from complete data: weighted Lasso solution

$$\min \frac{1}{2}\|y - Ax\|_2 + \sum_{i=1}^{n}\left[\frac{\pi_i|x_i| + \epsilon/n}{\alpha} + \frac{(1-\pi_i)|x_i| + \epsilon/n}{\beta} - \pi_i \log\frac{1-p}{\alpha} - (1-\pi_i)\log\frac{p}{\beta}\right]$$

z_i hidden data, $\pi_i = \mathbb{P}(z_i = 1|x, \alpha, \beta, p)$, x_i visible data, α, β, p mixture parameters

3 -- 2LMM tuned iterative shrinkage-thresholding

Aim: Improve existing methods for Lasso minimization based on shrinkage-thresholding via 2-LMM tuning

ISTA (Daubechies & al., 2004) Start from $x(0) = 0$ and compute

$$x^{(t+1)} = \eta_{\lambda\tau}[x^{(t)} + \tau A^{\mathsf{T}}(y - Ax^{(t)})].$$

where $\tau \in (0, 2\|A\|_2^{-2})$, η_γ thresholding function

$$\eta_\gamma[x] = \text{sgn}(x)\max(|x| - \gamma, 0)$$

+ Convergence: analytical conditions for convergence to a minimum
− Rate: speed is sensitive to local well-conditioning of the matrix A

Other alternatives: better rate of convergence and phase transitions, i.e. FISTA (Beck&Teboulle, 2010), AMP (Donoho & al., 2009).

Proposed solution: 2LMM-tuning

- ▶ incorporate support detection at each iteration
- ▶ iterative minimization of weighted Lasso
- ▶ use Expectation Maximization to learn the mixture parameters

2LMM-ISTA

Initialization: set K, $p = K/n$ $\alpha^{(0)} = \alpha_0$, $\pi^{(0)} = \mathbb{1}$, $\epsilon^{(0)} = 1$;
for $t = 1, \ldots, StopIter$
- ℓ_1-weights: $\omega_i^{(t+1)} = \pi_i^{(t)}/\alpha^{(t)} + (1 - \pi_i^{(t)})/\beta^{(t)}$
- Gradient/Thresholding step:

$$x^{(t+1)} = \eta^{\mathsf{S}}_{\lambda\tau\omega^{(t+1)}}(x^{(t)} + \tau A^\top(y - Ax^{(t)}))$$

- Posterior distribution evaluation:

$$\pi^{(t+1)} = \sigma_{n-K}\left[\mathbb{P}(z_i = 1|x^{(t+1)}, \alpha^{(t)}, \beta^{(t)}, p)\right]$$

- Regularization parameter:

$$\epsilon^{(t+1)} = \min\left(\frac{1}{\log(t+1)} + c\|x^{(t+1)} - x^{(t)}\| + \frac{r(x^{(t+1)})_{K+1}}{n}, \epsilon^{(t)}\right)$$

- Parameters estimation:

$$\alpha^{(t+1)} = \frac{\sum_i \pi_i^{(t+1)}|x_i^{(t+1)}| + \epsilon^{(t+1)}}{\|\pi^{(t+1)}\|_1} \qquad \beta^{(t+1)} = \frac{\sum_i (1-\pi_i^{(t+1)})|x_i^{(t+1)}| + \epsilon^{(t+1)}}{\|\mathbb{1} - \pi^{(t+1)}\|_1}$$

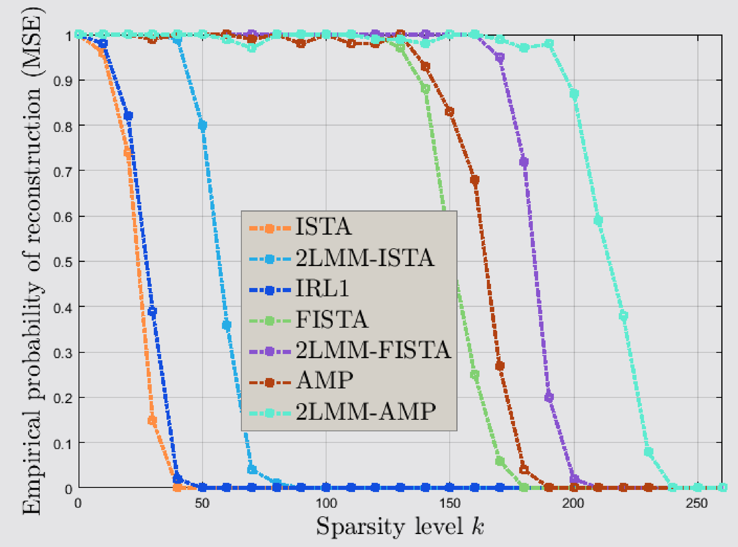
end for

+ Convergence: analytical conditions for convergence to a limit point
+ Rate: faster than classical methods

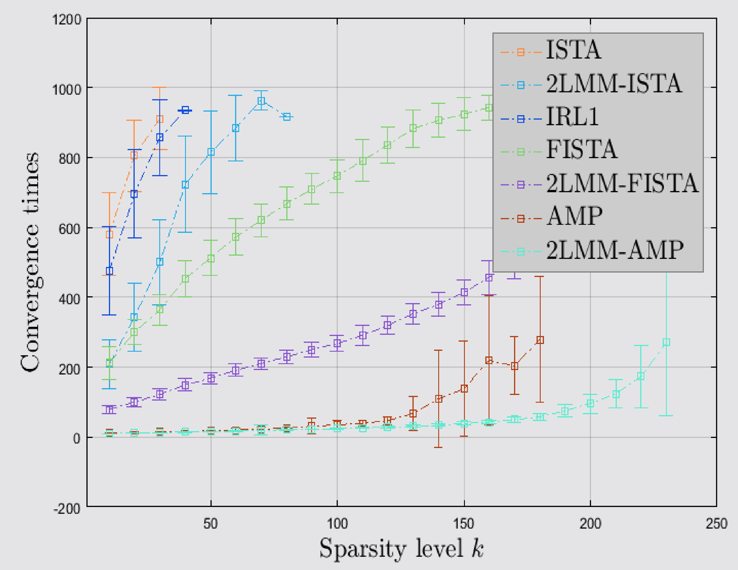
2LMM-tuning can also be applied to other shrinkage-thresholding methods, e.g. 2LMM-FISTA, 2LMM-AMP.

4 -- Numerical experiments

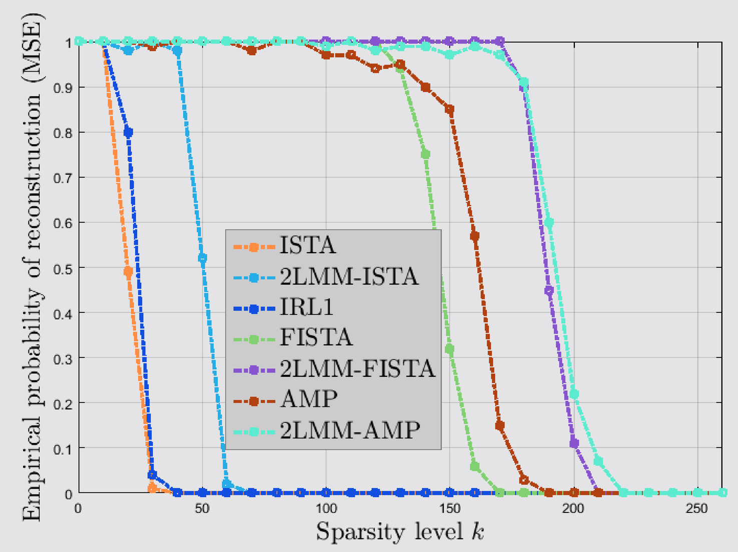
Reconstruction from noise-free measurements:
- ▶ Sparse signals of length $n = 512$, number of measurements $m = 350$
- ▶ Gaussian sensing matrices: $A_{ij} \sim N(0, 1/m)$



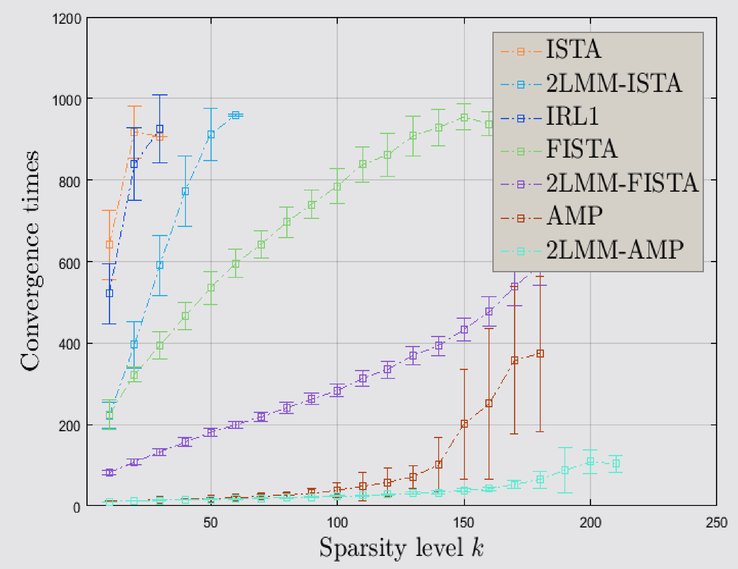
Sparse Bernoulli-Gaussian signals: success (MSE $< 10^{-4}$)



Sparse Bernoulli-Gaussian signals: average running times (CPU Times in [s])

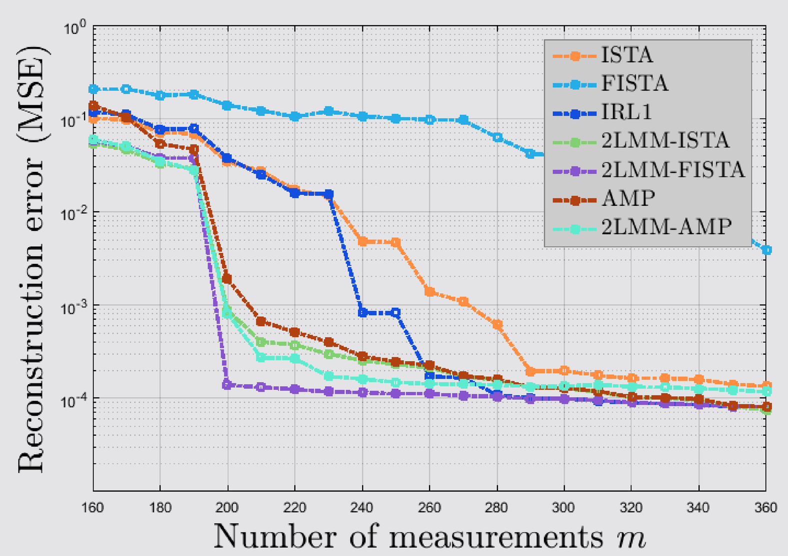


Sparse Bernoulli-Binomial signals: success (MSE $< 10^{-4}$)



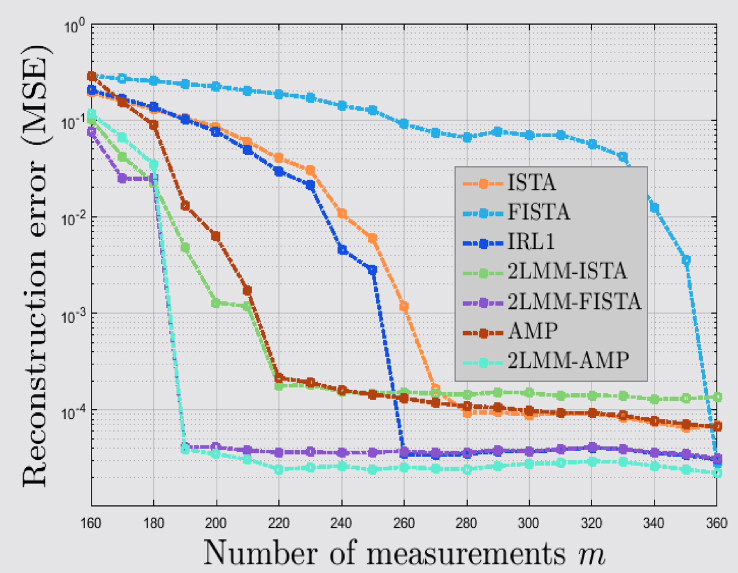
Sparse Bernoulli-Binomial signals: average running times (CPU Times in [s])

Reconstruction in imperfect scenarios:



Not exactly sparse signals: $x^\star = x_0 + \xi$

- x_0 exactly 56-sparse Bernoulli-Gaussian signal of length $n = 512$;
- ξ random noise with $\sigma = 0.01$



Reconstruction from noisy measurements

- $x^\star$ exactly 56-sparse Bernoulli-Gaussian signal of length $n = 512$;
- η random noise with $\sigma = 0.01$

Additional material: C. Ravazzi and E. Magli, "Laplace mixture models for efficient and accelerated compressed sensing", submitted to IEEE Trans. on Signal Processing.

WEBSITE: www.crisp-erc.eu
CONTACTS: chiara.ravazzi@polito.it, enrico.magli@polito.it