Quadratically fast IRLS for sparse signal recovery

Chiara Ravazzi, and Enrico Magli

Politecnico di Torino - DET, Italy

Signal Processing with Adaptive Sparse Structured Representations (SPARS) 2015

1 -- Classical IRLS for sparse signal recovery

Aim: Designing algorithms for sparse recovery problems with simple implementation and fast rate of convergence

Model: compressed data acquisition

- set of observations: $y = Ax^\star + \eta$
- $x^\star \in \Sigma_k = \{x \in \mathbb{R}^n : |\mathrm{supp}(x)| \leq k << n\} \rightsquigarrow$ unknown sparse signal
- $A \in \mathbb{R}^{m \times n}$ with $m \ll n \rightsquigarrow$ sensing matrix
- $\eta \in \mathbb{R}^m$ bounded noise with $\|\eta\|_2 \leq \delta$
- $\mathcal{F}(y, \delta) = \{z \in \mathbb{R}^n : \|Az - y\|_2 \leq \delta\}$

IRLS for ℓ_τ minimization: Let $\tau \in (0, 1]$

$$\min \|x\|_{\ell_\tau} \quad \text{s.t.} \quad x \in \mathcal{F}(y, \delta)$$

Given $\epsilon > 0$ and an initial guess $x^{(0)}$, compute

$$x^{(t+1)} = x^{(t+1)}(\tau) = \underset{x \in \mathcal{F}(y,\delta)}{\operatorname{argmin}} \|x\|^2_{w^{(t+1)}} = \underset{x \in \mathcal{F}(y,\delta)}{\operatorname{argmin}} \sum_{i=1}^n w_i^{(t+1)} x_i^2$$

$$w_i^{(t+1)} = w_i^{(t+1)}(\tau) = (\epsilon^2 + (x_i^{(t)})^2)^{\tau/2-1} \qquad i \in \{1, \ldots, n\}$$

+ **Convergence:** analytical conditions for convergence (Daubechies & al., 2010), (Ba & al., 2014)
+ **Rate:** (Daubechies & al., 2010), (Ba & al., 2014)
  - ▷ $\tau = 1$ globally linearly fast
  - ▷ $\tau \in (0, 1)$ locally super linearly fast with rate $2 - \tau$
- Local superlinear convergence: the algorithm gets trapped in local minima (e.g. $\tau < 1/2$)
- Open issue: heuristic methods to avoid local minima

2 -- This work: GSM-IRLS for sparse signal recovery

Aim: Designing faster IRLS (with quadratic rate of convergence)

GSM-IRLS for support detection and estimation:

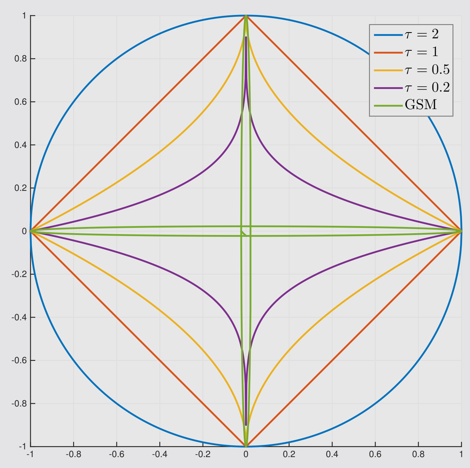
Given $0 \approx \alpha^{(0)} \ll \beta^{(0)}$, $\pi^{(0)} \in [0, 1]^n$, iterate

$$x^{(t+1)} = \underset{x \in \mathcal{F}(y,\delta)}{\operatorname{argmin}} \sum_{i=1}^n \left(\frac{\pi_i^{(t)}}{\alpha^{(t)}} + \frac{1 - \pi_i^{(t)}}{\beta^{(t)}} \right) x_i^2$$

$$\pi^{(t+1)} = f(x^{(t+1)}, \alpha^{(t)}, \beta^{(t)})$$

$$\alpha^{(t+1)} = \frac{\sum_{i=1}^n \pi_i^{(t+1)} (x_i^{(t+1)})^2 + \epsilon}{\sum_{j=1}^n \pi_j^{(t+1)}},$$

$$\beta^{(t+1)} = \frac{\sum_{i=1}^n (1 - \pi_i^{(t+1)})(x_i^{(t+1)})^2 + \epsilon}{\sum_{j=1}^n (1 - \pi_j^{(t+1)})}$$



The ℓ_τ-balls and the weighted ℓ_2-ball.

Three different implementations:

1. ML-IRLS: $\pi^{(t)} \in \{0, 1\}^n$ (*hard* support detection)
2. EM-IRLS: $\pi^{(t)} \in [0, 1]^n$ (*soft* support detection)
3. K-EM-IRLS: $\pi^{(t)} \in [0, 1]^n$ (*soft* with *thresholding*)

+ **Convergence:** analytical conditions for convergence
+ **Rate:**
  - ▷ noise-free case ($\delta = 0$): analytical conditions for locally quadratically fast convergence (with rate equal to 2)
  - ▷ noisy case ($\delta > 0$): open theoretical problem

3 -- Relating classical IRLS (1) and GSM-IRLS (2)

Interpretation as a constrained maximum log-likelihood estimation under a GSM distribution

Classical IRLS Proxy: $x^\star$ is a random variable with i.i.d. entries

$$f_{x_i}(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \kappa(x_i^2) \right)$$

$$k(z) = \kappa(0) + \int_0^z \frac{1}{(\epsilon^2 + t)^{1-\tau/2}} \mathrm{d}t.$$

ML from visible data (Ba & al., 2014): minimization of

$$L_\tau^-(x) := \sum_{i=1}^n (x_i^2 + \epsilon^2)^{\tau/2} \quad \text{s.t.} \quad x \in \mathcal{F}(y, \delta)$$

GSM-IRLS Proxy: $x^\star$ is a random variable with i.i.d. entries

$$f_{x_i}(x) = \frac{1-p}{\sqrt{2\pi\alpha}} \mathrm{e}^{-\frac{x^2}{2\alpha}} + \frac{p}{\sqrt{2\pi\beta}} \mathrm{e}^{-\frac{x^2}{2\beta}}$$

ML from complete data: minimization of

$$\mathrm{L}^-(x, z, \alpha, \beta) := \frac{1}{n} \sum_{i=1}^n \left[\frac{z_i x_i^2 + \epsilon^2/n}{2\alpha} + \frac{z_i}{2} \log\alpha - z_i \log(1-p) + \frac{(1-z_i)x_i^2 + \epsilon^2/n}{2\beta} + \frac{(1-z_i)}{2} \log\beta - (1-z_i)\log p \right] \quad \text{s.t.} \quad x \in \mathcal{F}(y, \delta)$$

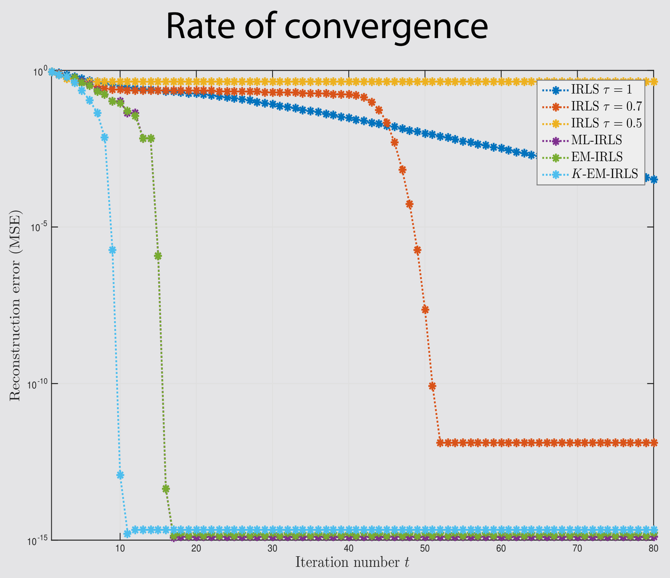
z_i hidden data, x_i visible data, α, β, p mixture parameters

4 -- Numerical comparison: classical IRLS (1) vs GSM-IRLS (2)

Setup: sparse uniform signals (nonzero components $x_i^\star \sim \mathrm{U}([-10, 10])$), Gaussian sensing matrices: $A_{ij} \sim \mathrm{N}(0, 1/m)$

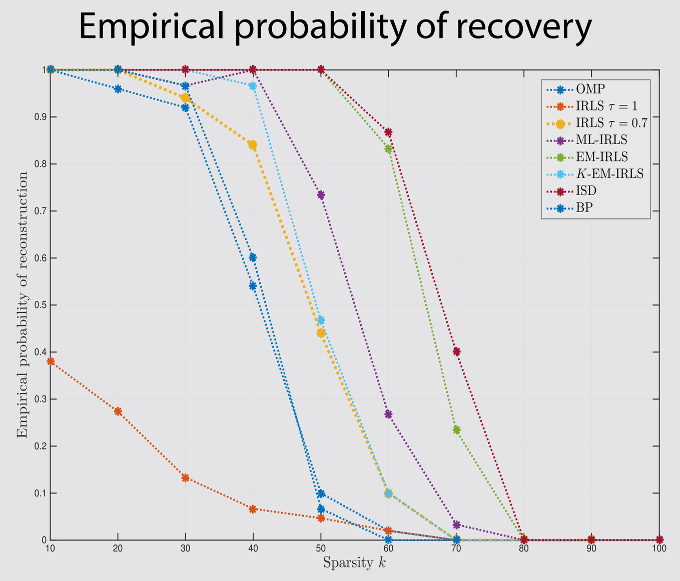
Reconstruction from noise-free measurements: $\delta = 0$

Rate of convergence



$k = 45, n = 1500, m = 250$

Empirical probability of recovery

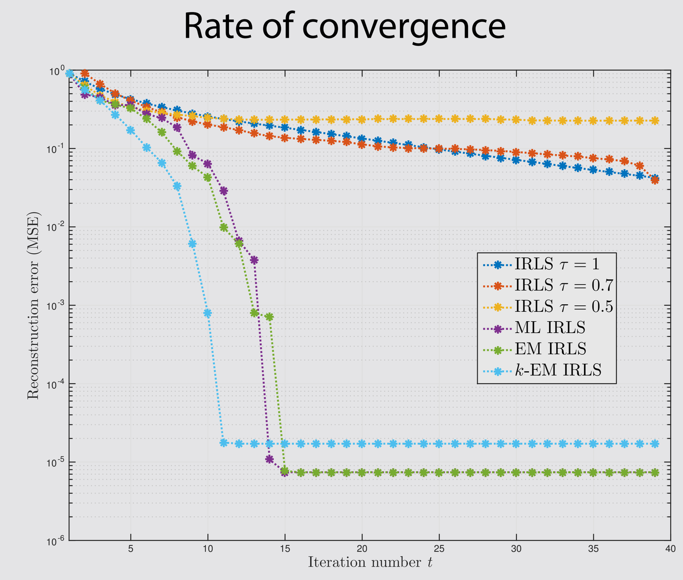


Success: $MSE < 10^{-4}$: $n = 512$ and $m = 160$

Other tests (Ravazzi & Magli, 2015): sparse Gaussian/Bernoulli signals, Image reconstruction

Reconstruction from noisy measurements

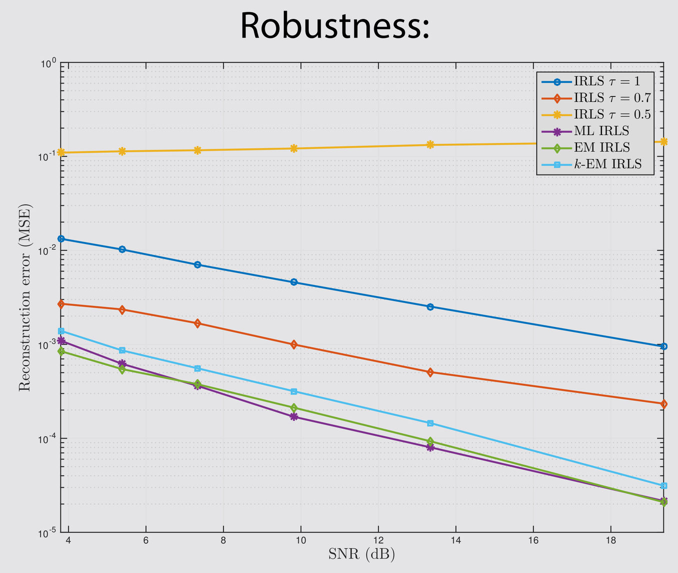
Rate of convergence



$k = 45, n = 1500, m = 250$,

white noise with std $\sigma = 0.01$, $\delta = \sqrt{m}\sigma$

Robustness:



$k = 45, n = 1500, m = 250$

Essential bibliography

- I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, no. 1, pp. 1-38, Jan. 2010.
- D. E. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 183-195, 2014.
- C. Ravazzi and E. Magli, "Gaussian mixtures based IRLS for sparse recovery with quadratic convergence," *IEEE Transactions on Signal Processing*, in press.

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