Fast and robust EM-based IRLS algorithm for sparse signal recovery from noisy measurements

Chiara Ravazzi and Enrico Magli
Politecnico di Torino (DET), Italy
40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19th – 24th April 2015

1 – Framework: 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: $y = Ax + \eta$
- $x^* \in \Sigma = \{x \in \mathbb{R}^n : \| \text{supp}(x) \| \leq k \ll n \} \sim$ unknown sparse signal
- $A \in \mathbb{R}^{m \times n}$ with $m \ll n \sim$ sensing matrix
- $\eta \in \mathbb{R}^m$ bounded noise with $\| \eta \| \leq \delta$
- $\mathcal{F}(y, \delta) = \{ x \in \mathbb{R}^n : \| Az - y \| \leq \delta \}$

IRLS for $\ell_1$ minimization: Let $\tau \in [0, 1]$

$$\min_{x \in \mathcal{F}(y, \delta)} \| x \|_1 \quad \text{s.t.} \quad x \in \mathcal{F}(y, \delta)$$

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

$$x^{(1)} = x^{(1)}(\tau) = \arg\min_{x \in \mathcal{F}(y, \delta)} \| x \|_1 + \tau \| x \|_2^2$$

$$w_i^{(1)}(\tau) = (\epsilon^2 + \tau_0^2)^{\tau/2-1} \quad i \in \{1, \ldots, n\}$$

Convergence: analytical conditions for convergence (Daubechies et al., 2010; Ba & al., 2014)

- Rate: (Daubechies et al., 2010; Ba & al., 2014)
  - $\tau = 1$ globally linearly fast to that sparse solution
  - $\tau \in (0, 1)$ locally super linearly fast with rate $2 \sim \tau$
- Local superlinear convergence: the algorithm trapped in local minima
- Open issue: heuristic methods to avoid local minima

4 – Relating classical IRLS (1) and EM-IRLS (2)

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

Classical IRLS

Proxy: $x^*$ is a random variable with i.i.d. entries

$$f_k(x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \| x \|_k^2 \right)$$

$$k(x) = k(0) + c x^2 \quad \text{for } 2^{\tau-1} \ll 1 \text{ or } 2^{\tau-1} \ll x^2$$

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

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

EM-IRLS

Proxy: $x^*$ is a random variable with i.i.d. entries

$$f_k(x) = \frac{1 - p}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{x^2}{\sqrt{2\pi}} \right) + \frac{p}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{x^2}{\sqrt{2\pi}} \right)$$

ML from complete data: minimization of

$$L(x, z, \alpha, \beta, \delta) = \frac{1}{2} \sum_{i=1}^n \left( x_i^2 + c_i^2 \right)^{\tau/2} + \frac{n}{2} \log \alpha - z_i \log(1 - p) + \frac{(1 - z_i) x_i^2 + c_i^2}{2\beta} + \frac{(1 - z_i) \log \beta - (1 - z_i) \log p}{2}$$

s.t. $x \in \mathcal{F}(y, \delta)$

4 hidden data, $x$, visible data, $\alpha$, $\beta$, $p$ mixture parameters

2 – This work: EM-IRLS for sparse signal recovery

Aim: Designing faster IRLS able to reach the region of guaranteed convergence

EM-IRLS for support detection and estimation:

Given $0 \approx \epsilon(0) \ll \beta(0), \alpha(0)$, compute

$$\hat{x}^{(1)} = \arg\min_{x \in \mathcal{F}(y, \delta)} \sum_{i=1}^n \left( x_i^2 + c_i^2 \right)^{\tau/2} + \frac{n}{2} \log \alpha - z_i \log(1 - p) + \frac{(1 - z_i) x_i^2 + c_i^2}{2\beta} + \frac{(1 - z_i) \log \beta - (1 - z_i) \log p}{2}$$

Three different implementations:

1. ML-IRLS: $\pi_t(0) \in [0, 1]^n$ (hard support detection)
2. EM-IRLS: $\pi_t(0) \in [0, 1]^n$ (soft support detection)
3. K-EM-IRLS: $\pi_t(0) \in [0, 1]^n$ (supp($\pi_t(0)$) $\leq n - K$) with K sparsity guess

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 – Numerical comparison: classical IRLS (1) vs EM-IRLS (2)

Setup:

- Sparse uniform signals: nonzero components $x^* \sim U([-10, 10])$
- Gaussian sensing matrices: $A_n \sim N(0, 1/m)$

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

Rate of convergence: evolution of MSE

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

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

Reconstruction from noisy measurements: $\delta = \sqrt{\text{SNR}}$

Rate of convergence: evolution of MSE

$k = 45, n = 1500, m = 250, n = 0.01$

Robustness: log-linear dependence of the MSE as a function of SNR and $\delta$

Other tests (Ravazzi & Magli, 2015): sparse Gaussian/Bernoulli signals, Shepp-Logan Phantom

Essential bibliography