FAST AND ROBUST EM-BASED IRLS ALGORITHM FOR SPARSE SIGNAL RECOVERY FROM NOISY MEASUREMENTS

C. Ravazzi and E. Magli

Department of Electronics and Telecommunications (DET), Politecnico di Torino, Italy

ABSTRACT

In this paper, we analyze a new class of iterative re-weighted least squares (IRLS) algorithms and their effectiveness in signal recovery from incomplete and inaccurate linear measurements. These methods can be interpreted as the constrained maximum likelihood estimation under a two-state Gaussian scale mixture assumption on the signal. We show that this class of algorithms, which performs exact recovery in noiseless scenarios under suitable assumptions, is robust even in presence of noise. Moreover these methods outperform classical IRLS for ℓ_τ-minimization with $\tau \in (0, 1]$ in terms of accuracy and rate of convergence.

***Index Terms*—** Compressed sensing, constrained maximum likelihood, Gaussian scale mixtures, ℓ_τ-minimization, sparsity

1. INTRODUCTION

We consider the compressive sensing (CS) problem that concerns the recovery of a sparse or compressible signal (*i.e.* it contains many coefficients close or equal to zero, when represented in some domain) from incomplete and inaccurate measurements [1]. This topic has attracted a lot of attention in recent years for potential applications in various areas such as communication theory [2], imaging sciences [3], radar technology [4], sensor networks [5], and MR angiography [6] and tomography [7].

The reconstruction problem can be addressed in different ways and the literature proposes a large number of approaches such as optimization-based methods [8], pursuit strategies [9, 10], coding-theoretic tools [11, 12], and Bayesian methods [13], to mention just a few. In particular, optimization-based estimation, such as ℓ_τ-minimization with $\tau \in (0, 1]$, Dantzig selector and the LASSO, require to solve a convex or non-convex program whose minimizer is known to approximate the target signal. Under certain conditions, the provided estimates have been proven to be exact in absence of noise and robust in the presence of inaccurate observations [14]. This means that small perturbations in the observations should cause small perturbations in the reconstruction.

Because of their simplicity and their theoretical guarantees, iteratively re-weighted least square methods (IRLS) have been proposed in [15] as an appealing strategy for ℓ_τ-minimization problems in sparse recovery with $\tau \in (0, 1]$. More precisely, under certain conditions, these methods converge to a minimizer globally linearly fast when $\tau = 1$ and locally superlinearly fast with rate $2 - \tau$ for $\tau \in (0, 1)$. Although IRLS algorithms appear very robust and super-linearly fast with rate close to 2 for τ approaching 0, such guarantees of rate of convergence are valid only in a neighborhood of the solution [16]. More precisely, the algorithm seems to converge properly to the desired solution when τ is not too small (*e.g.* $\tau > 1/2$) and tends to fail to reach the region of guaranteed convergence when $\tau < 1/2$ [8]. Heuristic methods to avoid local minima are still an open issue.

In this paper, we propose a new class of IRLS algorithms. These new procedures can be interpreted as an Expectation Maximization (EM) algorithm [17] for constrained maximum likelihood estimation where a two-state Gaussian mixture [18] with Bernoulli prior is used as a proxy for the sparse signal model. These strategies consider the complete log-likelihood function based on the unknown data and are methods of estimating the parameters of the mentioned statistical model. More precisely, in the first step the current values for the parameters are used to estimate the signal and to evaluate the posterior distribution of the signal coefficients; in the second step the mixture parameters are updated based on their probabilities.

If there is a sparse solution to the inverse problem, then the reconstruction is exact in absence of noise and the algorithm is quadratically fast in a neighborhood of that solution. The interested reader can find additional material on the application of these algorithms in the noise-free case in [19] and a discussion on the relation to prior literature.

We prove here that these reconstruction schemes converge even in presence of noise to a fixed point of the map that rules their dynamics. Moreover, we show via numerical experiments that the proposed algorithms are robust against noise and outperform classical IRLS for sparse recovery in terms of speed of convergence and accuracy.

This work is supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n.279848.

2. RECOVERY FROM NOISY MEASUREMENTS

In a typical CS setting one is concerned with the estimation of a k-sparse vector $x^\star \in \mathbb{R}^n$, (*i.e.*, the number of its nonzero components is not larger than $k \ll n/2$) starting from *few* linear measurements $y \in \mathbb{R}^m$ obtained through the following acquisition model

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

where $A \in \mathbb{R}^{m\times n}$ is the sensing matrix, $m < n$, and $\eta \in \mathbb{R}^m$ is some unknown perturbation bounded by $\|\eta\| \leq \delta$. Under certain assumptions on the sensing matrix and for a sufficient low level of the signal sparsity [14] , *robust signal recovery* is achieved by solving

$$(P_{0,\delta}): \qquad \min \|x\|_{\ell_0} \quad \text{subject to } x \in \mathcal{F}(y,\delta),$$

with $\mathcal{F}(y,\delta) = \{x \in \mathbb{R}^n : \|Ax - y\| \leq \delta\}$. This means that the solution $\widehat{x}_{0,\delta}$ of $(P_{0,\delta})$ obeys $\|\widehat{x}_{0,\delta} - x^\star\| \leq \kappa\delta$ where κ is a proportionality constant. Solving $(P_{0,\delta})$ is known to be an NP-hard problem. However, a practical solution is given by the following convex or non-convex surrogate of $(P_{0,\delta})$

$$(P_{\tau,\delta}): \qquad \min \|x\|_{\ell_\tau} \quad \text{subject to } x \in \mathcal{F}(y,\delta),$$

with $\tau \in (0,1]$. Given sufficient regularity conditions on the matrix A and assuming that δ is small enough [14], the global minimizer of $(P_{\tau,\delta})$ is expected to be close to $x^\star$.

The minimization in $(P_{\tau,\delta})$ can be carried out by the IRLS algorithm. More precisely, defining the w-weighted norm of $x \in \mathbb{R}^n$ as $\|x\|_w^2 = \sum_{i=1}^n w_i x_i^2$ with $w \in \mathbb{R}^n$ and $w_i > 0$, and given an initial guess $x^{(0)}$, this algorithm generates a sequence of estimates for the signal $x^\star$ as follows:

$$x^{(t+1)} = \underset{x\in\mathcal{F}(y,\delta)}{\arg\min} \|x\|_{w^{(t+1)}(\tau)}$$

with

$$w_i^{(t+1)} = ((\epsilon^{(t)})^2 + (x_i^{(t)})^2)^{\tau/2-1}$$

for $i \in \{1,\dots,n\}$ and a suitable non-increasing sequence $\epsilon^{(t)}$. At each iteration the IRLS algorithm corresponds to a constrained weighted least-squares problem and can be efficiently solved using standard convex optimization tools. As observed in [16], if the output SNR is greater than 1, then, at each iteration, 0 is not a feasible solution and the solutions lie on the boundary of $\mathcal{F}(y,\delta)$. IRLS appears very robust and stable with a linear (if $\tau = 1$) or super-linear (when $\tau \in (0,1)$) convergence in a neighborhood of the global minimizer of $(P_{\tau,\delta})$. In next section we introduce a new class of IRLS (with a different choice of weights w_i) that outperform the classical ones in terms the speed of convergence and accuracy.

3. EM-BASED IRLS

We now describe the basic principles of a new class of IRLS algorithms.

3.1. Algorithms description

Let us define the non increasing rearrangement of x by $r(x) := (|x_{i_1}|, |x_{i_2}|, \dots, |x_{i_n}|)^\top$, where $|x_{i_\ell}| \geq |x_{i_{\ell+1}}|$, $\forall \ell = 1,\dots,n-1$.

Given $K = K_0$, $p = K_0/n$, $\alpha^{(0)} = \alpha_0$, $\beta^{(0)} = \beta_0$, $\pi^{(0)} = \pi_0$, $\epsilon^0 = 1$, where K_0 and $\pi_i^{(0)}$ are an initial guess of the sparsity level of the signal and of the probability that $i \in \mathrm{supp}(x^\star)$, respectively. The EM-based IRLS generates a sequence $\{x^{(t)}\}_{t=1}^\infty$ of estimations of $x^\star$ as follows

$$w_i^{(t+1)} = \pi_i^{(t)}/\alpha^{(t)} + (1-\pi_i^{(t)})/\beta^{(t)} \tag{1}$$

$$x^{(t+1)} = \underset{x\in\mathcal{F}(y,\delta)}{\arg\min} \|x\|^2_{w^{(t+1)}} \tag{2}$$

$$\pi_i^{(t+1)} = \frac{g(x_i^{(t+1)}, \alpha^{(t)}, 1-p)}{g(x_i^{(t+1)}, \alpha^{(t)}, 1-p) + g(x_i^{(t+1)}, \beta^{(t)}, p)}, \tag{3}$$

with

$$g(s,\sigma,q) = \exp\left(-\frac{s^2}{2\sigma} - \frac{\log(\sigma)}{2} + \log(q)\right),$$

$$\epsilon^{(t+1)} = \min\left(\epsilon^{(t)}, r(x^{(t+1)})_{K+1}/n\right)$$

$$\alpha^{(t+1)} = \frac{\sum_{i=1}^n \pi_i^{(t+1)} |x_i^{(t+1)}|^2 + |\epsilon^{(t+1)}|^2}{\sum_{i=1}^n \pi_i^{(t+1)}}, \tag{4}$$

$$\beta^{(t+1)} = \frac{\sum_{i=1}^n (1-\pi_i^{(t+1)}) |x_i^{(t+1)}|^2 + |\epsilon^{(t+1)}|^2}{\sum_{i=1}^n (1-\pi_i^{(t+1)})}. \tag{5}$$

As in classical IRLS algorithms, (2) requires the solution of a weighted least squares problem constrained to the closed quadratic convex set $\mathcal{F}(y,\delta)$.

The EM-based IRLS can be seen as an EM algorithm for a constrained maximum likelihood estimation. More precisely, let us assume that $x^\star$ is a random variable with components modeled as a two-state GSM

$$x_i^\star = z_i\sqrt{\alpha}u_i + (1-z_i)\sqrt{\beta}u_i \quad i \in [n]$$

where u_i are identically and independently distributed (i.i.d.) zero-mean Gaussians and z_i are i.i.d. Bernoulli variables with probability mass function $\mathbb{P}(z_i = 1) = 1-p$, $p = k/n$, $\alpha \approx 0$, and $\beta \gg 0$.

With these assumptions on the signal $x^\star$, the proposed IRLS algorithm can be viewed as an iterative solution for the constrained minimization of the ϵ-smoothed negative log-likelihood function L, defined as follows

$$L(x,z,\alpha,\beta,\epsilon) = \sum_{i=1}^n \left[\frac{z_i x_i^2 + \epsilon^2/n}{2\alpha} + \frac{z_i}{2}\log\frac{\alpha}{(1-p)^2} + \frac{(1-z_i)x_i^2 + \epsilon^2/n}{2\beta} + \frac{(1-z_i)}{2}\log\frac{\beta}{p^2}\right] + c \tag{6}$$

for some fixed constant $c > 0$. The EM-based IRLS strategy can be summarized as follows: (a) set an initial estimate K for the sparsity level, $p = K/n$, a small variance $\alpha^{(0)} \approx 0$ (*e.g.* $\alpha^{(0)} = 0.1$), the initial distribution on the zero elements $\pi^{(0)} = \mathbb{1}$, and $\epsilon^{(0)} = 1$; (b) given the observed data y and pretending for the moment that the current values of the parameters are correct, estimate the signal $x^{(t)}$ and evaluate the posterior distribution of the signal coefficients $\pi^{(t)}$ (according to (2), (1), and (3)); (c) given the probabilities, use them to re-estimate the mixture parameters $\alpha^{(t)}$ and $\beta^{(t)}$ (as in (4) and (5)); (d) and iterate until a stopping criterion is satisfied.

We refer the reader to [19] for the technical derivation on how the constrained ML estimation (6) leads to the updates (1)-(5).

3.2. An acceleration via thresholding: K-EM based IRLS

Finally, we consider a modification of EM-based IRLS, which we call K-EM-based IRLS Algorithm. This algorithm is a thresholded version of EM-based IRLS, taking into account that we are seeking a K-sparse signal. More precisely, at each step a weighted least squares problem is solved (see (2)) with weights given by

$$w_i^{(t+1)} = \widehat{\pi}_i^{(t)}/\alpha^{(t)} + (1-\widehat{\pi}_i^{(t)})/\beta^{(t)} \qquad (7)$$

and $\widehat{\pi}_i^{(t+1)} = \mathsf{H}_{N-K}(\pi^{(t+1)})$, where the thresholding operator H acts on π keeping its $n-K$ biggest elements and setting the others to zero, and $\pi^{(t+1)}$ is updated as in (3).

3.3. Theoretical guarantees

Let us denote the iterations given by (2)-(4) as follows: $\theta^{(t)} = (x^{(t)}, \pi^{(t)}, \alpha^{(t)}, \beta^{(t)}, \epsilon^{(t)})$. The following theorem ensures that, under certain conditions, the sequence $\{\theta^{(t)}\}_{t\in\mathbb{N}}$ of estimates provided by the proposed algorithms converges to the set of fixed points of the map that rules their dynamics.

Theorem 1 (Convergence)**.** *Let $\mathcal{F}(y,\delta)$ be a non-empty closed convex subset of $\mathbb{R}^n$. Then the sequence $(\theta^{(t)})_{t\in\mathbb{N}}$ converges to a fixed point of the algorithm.*

Sketch of the proof. The algorithms have been designed in such a way that there exists a function V which is nonincreasing and convergent along the sequence of iterates:

$$V(\theta^{(t)}) \geq V(\theta^{(t+1)}).$$

where $\theta^{(t)} = (x^{(t)}, \pi^{(t)}, \epsilon^{(t)}, \alpha^{(t)}, \beta^{(t)})$. Moreover the following facts can be proved: (a) the sequence $(x_i^{(t)})_{t\in\mathbb{N}}$ is upper bounded; (b) two successive iterations of these algorithms become closer and closer: $\lim_{t\to\infty} \|x^{(t+1)} - x^{(t)}\| = 0$.

Finally, the convergence is obtained using arguments of variational calculus. The interested reader can deduce the rigorous proof of these facts using similar arguments devised in Theorem 1 of [19]. □

4. NUMERICAL EXPERIMENTS

In this section we discuss a series of experiments in order to assess the performance of the proposed EM-based IRLS methods in terms of convergence time and accuracy. We also show that these algorithms yield exact reconstruction in the noiseless scenario and are robust in presence of noise, in that small errors on the measurements produce small perturbation in the reconstruction.

4.1. Experiment Setup

For all experiments the signal $x^\star$ to be estimated is generated by choosing $k = 45$ nonzero components uniformly at random among $n = 1500$ elements and drawing the amplitude of each nonzero component from a uniform distribution $\mathsf{U}([-10, 10])$ in order to introduce a mismatch in the signal model. The sensing matrix $A \in \mathbb{R}^{m\times n}$ with $m = 250$ is sampled with i.i.d. Gaussian entries with zero mean and variance equal to $1/m$. We have initialized the parameters $\alpha^0 = 0.1$, $\pi^{(0)} = \mathbb{1}$ and $K = 55$.

4.2. Reconstruction from Noise-free Measurements

In Fig. 1, the convergence rate of classical IRLS for different choices of $\tau = 1, 0.7, 0.5$ and EM-based methods are compared. In particular, the mean square error (MSE) $E(t) = \left\|x^{(t+1)} - x^\star\right\|^2/n$ is depicted as a function of the iterations. For classical IRLS, the case with $\tau = 1$ shows linear convergence and for the smaller values of it (*e.g.*, $\tau = 0.7$), the error decay initially follows a linear, transient regime. However, once the iterates get sufficiently close to the sparse solution vector, the convergence speeds up dramatically resulting in super-linear convergence. For smaller values of τ (*e.g.*, $\tau = 0.5$), we often do not observe convergence to the desired solution. In fact, if $\tau \leq 0.5$, then the algorithm tends to fail to reach the region of guaranteed convergence. The proposed EM-based IRLS and k-EM-IRLS are faster than classical IRLS methods: the transient linear regime lasts less and the local region of super-linear convergence is larger than classical IRLS methods based on ℓ_τ-minimization with $\tau < 0.5$. In [19] we compare the performance of GSM-IRLS with classical IRLS methods, Basis Pursuit (BP, [20]), Iterative support detection (Threshold-ISD, [21]) and Orthogonal Matching Pursuit (OMP, [22]), in terms of the empirical recovery success rate as a function of the sparsity level and number of measurements.

4.3. Reconstruction from Noisy Measurements

As a second example, we have considered the noisy scenario where the vector η is an additive white Gaussian noise with standard deviation $\sigma = 0.01$. We assume the standard deviation of the noise is known in advance and, as $\mathbb{E}[\|\eta\|^2] = m\sigma^2$, we set $\delta = \sqrt{m}\sigma$. As expected from Theorem 1, Figure 2

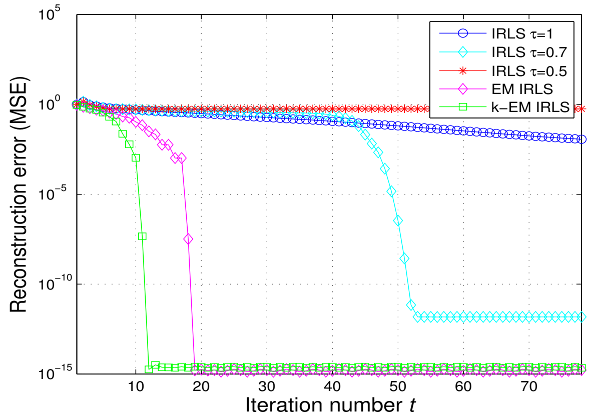

Fig. 1. Noise-free scenario: A typical evolution of MSE as a function of the iterations for classical IRLS algorithms (with $\tau = 1, 0.7, 0.2$) and IRLS based on EM algorithm. The nonzero components of the signal $x^\star$ are drawn from a uniform distribution $\mathsf{U}([-10, 10])$.

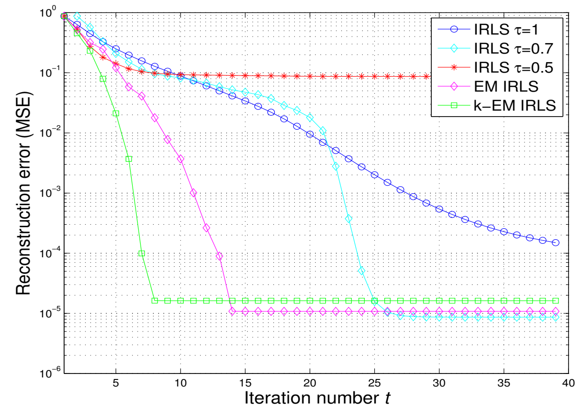

Fig. 2. Noisy scenario: A typical evolution of MSE as a function of the iterations for classical IRLS algorithms (with $\tau = 1, 0.7, 0.2$) and IRLS based on EM algorithm. The nonzero components of the signal $x^\star$ are drawn from a uniform distribution $\mathsf{U}([-10, 10])$ and the vector η is an additive white noise with standard deviation $\sigma = 0.01$.

shows that all the tested algorithms converge to a fixed point of the algorithm in few iterations. For smaller $\tau = 0.5$, we often do not observe convergence to the desired solution. In fact, since the ℓ_τ-norm is not a convex function, the algorithm easily gets trapped in local minima. It is worthwhile noting that the estimations obtained by the proposed EM-based IRLS are significantly more accurate in terms of the mean square error compared to those obtained with the classical IRLS methods.

4.4. Robustness

In Figure 3 we show that the proposed methods are robust against noise. More precisely, the mean square error, averaged over 50 runs and obtained after 50 iterations, is depicted as a function of Signal-to-Noise ratio (SNR). It should be noted that only few iterations are required to reach a satisfactory degree of accuracy. In all curves we can clearly identify the log-linear dependence of the MSE as a function of the SNR and, consequently, of the parameter δ. Moreover the MSE of the proposed algorithms are smaller than those obtained via classical IRLS algorithms with $\tau = 1, 0.7$ and $\tau = 0.5$. As already observed, the MSE of classical IRLS with $\tau = 0.5$ is very high compared to the other methods. Moreover, it does not decrease as the SNR increases and the algorithm turns out to be not robust against noise. This is due to the fact that the algorithm gets trapped into local minima and is unable to reach the global minima of $(P_{\tau,\delta})$.

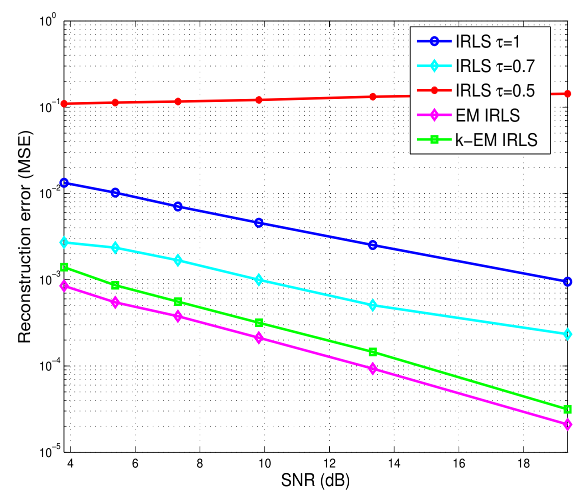

Fig. 3. Mean square error after 40 iterations as a function of the SNR for classical IRLS (with $\tau = 1, \tau = 0.7$ and $\tau = 0.5$) and the proposed EM-based algorithms.

5. CONCLUSIONS

This paper has proposed and explored a new class of IRLS algorithms for sparse recovery. These iterative procedures, obtained by modeling the signal as a two-state gaussian scale mixture and performing constrained log-likelihood maximization, are very powerful for signal recovery from both noise-free and inaccurate measurements. The main theoretical contribution includes the proof of convergence of the algorithm to a fixed point of the algorithm. Numerical results confirm that the proposed algorithms outperform classical IRLS schemes in terms of convergence rate and are robust against noise.

6. REFERENCES

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[2] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[3] E. Chouzenoux, A. Jezierska, J. Pesquet, and H. Talbot, "A majorize-minimize subspace approach for $\ell_2 - \ell_0$ image regularization," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 563–591, 2013. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/11085997X

[4] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *Trans. Sig. Proc.*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009. [Online]. Available: http://dx.doi.org/10.1109/TSP.2009.2014277

[5] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," 2005 (revised 2009), http://arxiv.org/abs/0901.3403.

[6] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, Feb 2006.

[7] H. Erdogan and J. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Medical Imaging*, vol. 18, no. 9, pp. 801–814, 1999.

[8] M. Fornasier, *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Radon Series on Computational and Applied Mathematics, 2010.

[9] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation: part i: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[10] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, Apr. 2008.

[11] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *Trans. Sig. Proc.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[12] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *Information Theory, IEEE Transactions on*, vol. 55, no. 9, pp. 4299–4308, Sept 2009.

[13] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[14] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6–18, Jan 2006.

[15] 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.

[16] D. Ba, B. Babadi, P. Purdon, and E. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *Signal Processing, IEEE Transactions on*, vol. 62, no. 1, pp. 183–195, Jan 2014.

[17] R. A. Redner and H. F. Walker, "Mixture Densities, Maximum Likelihood and the Em Algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, 1984. [Online]. Available: http://dx.doi.org/10.2307/2030064

[18] D. Titterington, A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York, 1985.

[19] C. Ravazzi and E. Magli, "Gaussian mixtures based IRLS for sparse recovery with quadratic convergence," *submitted to IEEE Trans. Signal Proc, (last revision Dec. 2014), http://calvino.polito.it/~ravazzi*, 2014.

[20] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences. Paris, France, ser. I*, vol. 346, pp. 589 – 592, 2008.

[21] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM J. Img. Sci.*, vol. 3, no. 3, pp. 462–491, aug 2010.

[22] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information theory*, vol. 53, pp. 4655–4666, 2007.