

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

C. Ravazzi and E. Magli

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

ABSTRACT

Iterative shrinkage-thresholding algorithms provide simple methods to recover sparse signals from compressed measurements. In this paper, we propose a new class of iterative shrinkage-thresholding algorithms which preserve the computational simplicity and improve iterative estimation by incorporating a soft support detection. Indeed, at each iteration, by learning the components that are likely to be nonzero from the current signal estimation using Bayesian techniques, the shrinkage-thresholding step is adaptively tuned and optimized. Unlike other adaptive methods, we are able to prove, under suitable conditions, the convergence of the proposed methods. Moreover, we show through numerical experiments that the proposed methods outperform classical shrinkage-thresholding in terms of rate of convergence and of sparsity-undersampling tradeoff.

Index Terms— Compressed sensing, MAP estimation, mixture models, reweighted ℓ_1 -minimization, sparsity

1. INTRODUCTION

In this paper, we consider the standard compressed sensing (CS) setting [1, 2], where we are interested in recovering a signal $x^* \in \mathbb{R}^n$ from $m \leq n$ measurements of the form

$$y = Ax^* + \xi \quad (1)$$

where $y \in \mathbb{R}^m$ is the observation vector and $A \in \mathbb{R}^{m \times n}$ is the measurements matrix, and ξ is zero-mean additive Gaussian noise with standard deviation σ . The linear system in (1) is underdetermined and has infinitely many solutions. Particular interest has been focused on parsimonious solutions to (1), selecting as desired solution the sparsest one, *i.e.* the one with the smallest number of nonzero components. A natural optimization formulation of this problem involves the minimization of ℓ_0 -psudonorm [1]. However, ℓ_0 -minimization is NP-hard and the search of the solution requires an exponential time in the length n .

A practical alternative is provided by least-absolute shrinkage and selection operator (Lasso) problem [3] which

has the following form

$$\min_{x \in \mathbb{R}^n} \lambda \|x\|_1 + \frac{1}{2} \|Ax - y\|_2^2 \quad (2)$$

where $\lambda > 0$ is a regularization parameter. From a probabilistic point of view, the Lasso problem may be interpreted as a Bayesian maximum a posteriori (MAP) estimate [4] when the signal coefficients x_i^* have independent and identical double exponential (Laplace) priors. The optimization in (2) can be solved by iterative shrinkage-thresholding algorithms (ISTA, [5, 6, 7]) that are generally first order methods followed by a shrinkage-thresholding step. Due to its implementation simplicity and suitability for high-dimensional problems, a large effort has been spent to improve their speed of convergence [8, 9], asymptotic performance in the large system limit [10, 11] and ease of use [12].

Although the Laplace probability density function may be a good probability model for the distribution of the signal coefficients, since it captures the main feature that is usually revealed in compressible signals (heavy tails and peaks at zero), this assumption fails to incorporate prior information on the support of the signal. In some cases, in fact, one might have some prior information about the support of the sparse signal, or estimate the signal support during the reconstruction [13, 14]. Such support estimates could be employed to reduce the number of measurements needed for good reconstruction via Lasso, *e.g.* by combining support information with weighted ℓ_1 -minimization [15].

In order to address this issue, in this paper, we propose an iterative support detection and estimation method for CS, that can be applied to improve a number of existing methods based on shrinkage-thresholding. More precisely, we consider iterative reweighted Lasso-minimization methods that incorporate a probabilistic model of the signal support. The fundamental idea is to use a Laplace mixture model as the parametric representation of the prior distribution of the signal coefficients. Because of the partial symmetry of the signal sparsity we know that each coefficient should have one out of only two distributions: a Laplace with small variance with high probability and a Laplace with large variance with low probability. In this work, the expectation maximization [4] algorithm is combined with iterative shrinkage-thresholding to (a) estimate the distribution of the components that are

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.

likely to be nonzero from signal estimations at each iteration (support detection); (b) tune and optimize the shrinkage-thresholding step allowing better estimation. As opposed to other adaptive methods [9], we are able to prove, under suitable conditions, the convergence of the proposed *Bayesian tuned* method. We apply this method to several algorithms based on shrinkage-thresholding, showing by numerical simulation that it improves recovery in terms of both speed of convergence and sparsity-undersampling tradeoff, while preserving the implementation simplicity.

2. ITERATIVE SHRINKAGE-THRESHOLDING

Iterative shrinkage-thresholding algorithms can be understood as a special proximal forward backward iterative scheme [16] reported as follows. Let $x(0) = x_0$, $\{\tau^{(t)}\}_{t \in \mathbb{N}}$ be a sequence in $(0, \infty)$ such that $\inf_{t \in \mathbb{N}} \tau^{(t)} > 0$ and $\sup_{t \in \mathbb{N}} \tau^{(t)} < 2\|A\|_2^{-2}$, and let $\{u^{(t)}\}_{t \in \mathbb{N}}$ be a sequence in \mathbb{R}^n . Then for every $t \in \mathbb{N}$ let

$$x^{(t+1)} = \eta_{\lambda\tau^{(t)}}^S [x^{(t)} + \tau^{(t)} A^T (y - Ax^{(t)} + u^{(t)})]. \quad (3)$$

where η_γ^S is a thresholding function to be applied element-wise, *i.e.*

$$\eta_\gamma^S[x] = \begin{cases} \text{sgn}(x)(|x| - \gamma) & \text{if } |x| > \gamma \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

ISTA, whose original version [5] is of the above form with $u^{(t)} = 0$ and $\tau^{(t)} = \tau < 2\|A\|_2^{-2}$ for all $t \in \mathbb{N}$, is guaranteed to converge to a minimizer of the Lasso. Moreover, as shown in [17], if A fulfills the so-called finite basis injectivity condition, the convergence is linear. However, the factor which determines the speed within the class of linearly-convergent algorithms depends on local well-conditioning of the matrix A , meaning that ISTA can converge arbitrarily slowly in some sense, which is also often observed in practice.

In order to speed up ISTA, alternative algorithms have exploited preconditioning techniques or adaptivity, combining a decreasing thresholding strategy with adaptive descent parameter. In [10] the thresholding and descending parameters are optimally tuned in terms of phase transitions, *i.e.*, they maximize the number of nonzeros at which the algorithm can successfully operate. However, preconditioning can be very expensive and there is no proof of convergence for adaptive methods. Finally, other variations update of the next iterate using not only the previous estimation, but two or more previously computed iterates. Among all the proposed techniques with a significantly better rate of convergence and phase transitions, we recall (a) Fast Iterative Shrinkage-Thresholding Algorithm (FISTA, [8]) obtained by (3) choosing $\tau^{(t)} = \tau < 2\|A\|_2^{-2}$ and

$$\begin{aligned} u^{(t)} &= \frac{\zeta^{(t-1)} - 1}{\zeta^{(t)}} (\tau^{-1} I - A^T A) (x^{(t)} - x^{(t-1)}) \\ \zeta^{(0)} &= 1, \quad \zeta^{(t+1)} = \frac{1 + \sqrt{1 + 4(\zeta^{(t)})^2}}{2} \end{aligned} \quad (5)$$

(b) Approximate Message Passing (AMP, [11]) with threshold recursion proposed in [18]

$$\begin{aligned} u^{(t)} &= \frac{1 - \tau^{(t)}}{\tau^{(t)}} A^T (Ax^{(t)} - y) - \frac{\|x^{(t)}\|_0}{m\tau^{(t)}} A^T r^{(t-1)} \\ \tau^{(t)} &= \chi \frac{\|r^{(t)}\|_2}{\sqrt{m}}, \quad r^{(t)} = y - Ax^{(t)} + r^{(t-1)}. \end{aligned} \quad (6)$$

3. BAYESIAN TUNED ITERATIVE SHRINKAGE-THRESHOLDING ALGORITHM

In this section we propose to incorporate, at each iteration of (3), support soft detection in order to improve signal recovery. The fundamental idea is to use a mixture model as the parametric representation that describes our prior knowledge about the sparsity of the solution. Because of the partial symmetry of the sparsity of the signal, we consider the case in which x is a random variable with components of the form

$$x_i = z_i u_i + (1 - z_i) w_i \quad i \in [n]$$

where $u_i \sim \text{Laplace}(0, \alpha)$, $w_i \sim \text{Laplace}(0, \beta)$ and $z_i \sim \text{Ber}(1 - p)$, where $p \ll 1/2$, $\alpha \approx 0$, and $\beta \gg 0$, in order to ensure that x has few large coefficients. Given y, A, z, α, β , we consider MAP estimation which prescribes to minimize over the variables x_i the following cost function

$$\begin{aligned} G(x; \alpha, \beta) &:= \frac{1}{2} \|y - Ax\|^2 + \lambda \sum_{i=1}^n \left[\frac{\pi_i |x_i| + \epsilon/n}{\alpha} \right. \\ &\quad \left. + \pi_i \log \frac{\alpha}{1-p} + \frac{(1 - \pi_i) |x_i| + \epsilon/n}{\beta} + (1 - \pi_i) \log \frac{\beta}{p} \right] \end{aligned} \quad (7)$$

where $\pi_i = \mathbb{E}[z_i | x_i, \alpha, \beta, p]$ is the posterior distribution and $\epsilon > 0$ is a regularization parameter introduced to avoid singularities when $\alpha \rightarrow 0$ or $\beta \rightarrow 0$. As can be easily seen, taking $\alpha = \beta$ the minimizers of (7) are minimizers of (2) and viceversa. Compared to (2), if $\alpha \neq \beta$ (7) is able to penalize the coefficients of the solution vector in different ways. Since information about the locations of the nonzero coefficients of the original signal is not available a priori, the task of selecting the parameters α, β and π is performed iteratively. We propose an alternating method for the minimization of (7), inspired by the so-called EM algorithm. The strategy is summarized in Algorithm 1.

In Bayesian tuned iterative shrinkage-thresholding three main tasks are iterated until a stopping criterion is satisfied: (a) given an initial guess K of the sparsity and current parameters π_i, α, β , the estimation of the signal is obtained by minimizing the weighted lasso

$$x^{(t+1)} = \arg \min \frac{1}{2} \|Ax - y\|^2 + \lambda \sum_{i=1}^n \omega_i^{(t+1)} |x_i| \quad (8)$$

with $\omega_i^{(t+1)} = \pi_i/\alpha + (1 - \pi_i)/\beta$; (b) the posterior distribution of the signal coefficients is evaluated and thresholded

Algorithm 1 Bayesian tuned iterative shrinkage-thresholding

Require: Data (y, A) , set $K = \tilde{K}$, $p = K/n$
1: Initialization: $\alpha^{(0)} = \alpha_0$, $\pi^{(0)} = \mathbb{1}$, $\epsilon^{(0)} = 1$
2: **for** $t = 1, \dots, \text{StopIter}$ **do**
3: Computation of ℓ_1 -weights:

$$\omega_i^{(t+1)} = \frac{\pi_i^{(t)}}{\alpha} + \frac{1 - \pi_i^{(t)}}{\beta}$$

4: Gradient/Thresholding step:

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

5: Posterior distribution evaluation:

$$\gamma_i^{(t+1)} = \frac{\frac{(1-p)}{\alpha^{(t)}} \exp(-\frac{|x_i^{(t+1)}|}{\alpha^{(t)}})}{\frac{(1-p)}{\alpha^{(t)}} \exp(-\frac{|x_i^{(t+1)}|}{\alpha^{(t)}}) + \frac{p}{\beta^{(t)}} \exp(-\frac{|x_i^{(t+1)}|}{\beta^{(t)}})}$$
$$\pi^{(t+1)} = \sigma_{n-K}(\gamma^{(t+1)})$$

6: Regularization parameter:

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

7: Parameters estimation:

$$\alpha^{(t+1)} = \frac{\langle \pi^{(t+1)}, x^{(t+1)} \rangle + \epsilon^{(t+1)}}{\|\pi^{(t+1)}\|_1}$$
$$\beta^{(t+1)} = \frac{\langle \mathbb{1} - \pi^{(t+1)}, x^{(t+1)} \rangle + \epsilon^{(t+1)}}{\|\mathbb{1} - \pi^{(t+1)}\|_1}$$

8: **end for**

by keeping its $n - K$ biggest elements and setting the others to zero; (c) the mixture parameters α and β are updated. It should be noticed that $\epsilon^{(t)}$ is a regularization parameter used to avoid singularities when $\alpha \rightarrow 0$ or $\beta \rightarrow 0$.

3.1. Relation to prior literature

As already observed, Algorithm 1 belongs to the more general class of methods for weighted ℓ_1 -norm minimization [19, 20, 21] (see (8)). Common strategies for iterative reweighting ℓ_1 -minimization (IRL1, [19]) that have been explored in literature re-compute weights at every iteration using the estimate at the previous iteration $\omega_i^{(t+1)} = \chi / (|x_i^{(t)}| + \epsilon)$ where χ and ϵ are appropriate positive constants. In Algorithm 1 the weights $\omega_i^{(t)}$ are chosen to jointly fit the signal prior and, consequently, depend on all components of the signal and not exclusively on the value $x_i^{(t)}$. Our strategy is also related to Threshold-ISD [22] that incorporates support detection in the weighted ℓ_1 -minimization and runs as fast as the basis pursuit. Given a support estimate, the estimation is performed by solving a truncated basis pursuit problem. Compared to Threshold-ISD, Bayesian tuned iterative shrinkage-thresholding does not use binary weights and is more flexible. Moreover, in Threshold-ISD, like CoSaMP, the identification of the support is based on greedy rules and not chosen to optimally fit the prior distribution of the signal.

A prior estimation based on EM was incorporated within the AMP framework also in [20] where a Gaussian mixture model is used as the parametric representation of the signal. The key difference in our approach is that fitting the signal prior is used to estimate the support and to adaptively select the best thresholding function with the least mean square error. The necessity of selecting the best thresholding function is also proposed in Parametric SURE AMP [23] where a class of parametric denoising functions is used to adaptively choose the best-in-class denoiser. However, at each iteration, Parametric SURE AMP needs to solve a linear system and the number of parameters determines heavily both performance and complexity.

3.2. Theoretical results

Under suitable conditions, we are able to guarantee the convergence of Bayesian tuned ISTA (B-ISTA), obtained by setting $\tau^{(t)} = \tau < 2\|A\|_2^{-2}$ and $u^{(t)} = 0$ in Algorithm 1. Let $\zeta^{(t)} = (x^{(t)}, \pi^{(t)}, \alpha^{(t)}, \beta^{(t)}, \epsilon^{(t)})$. B-ISTA is designed in such a way that there exists a function $V : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ which is nonincreasing and convergent along the sequence of iterates: $V(\zeta^{(t)}) \geq V(\zeta^{(t+1)})$. The next theorem ensures that also the sequence $(\zeta^{(t)})$ converges to a limit point which is also a fixed point of the algorithm.

Theorem 1 (B-ISTA convergence). *Let $\tau^{(t)} = \tau < 2\|A\|_2^{-2}$, $u^{(t)} = 0$. Then for any $y \in \mathbb{R}^m$, the sequence $\zeta^{(t)}$ generated by Algorithm 1 converges to $(x^\infty, \pi^\infty, \alpha^\infty, \beta^\infty)$ such that*

$$x^\infty = \eta_{\omega_\infty \lambda \tau}(x^\infty + \tau A^\top (y - Ax^\infty)), \quad (9)$$

$$\omega_i^\infty = \frac{\pi_i^\infty}{\alpha^\infty} + \frac{1 - \pi_i^\infty}{\beta^\infty}, \quad (10)$$

$$\pi_i^\infty = \frac{(\alpha^\infty)^{-1} \exp(-\frac{|x_i^\infty|}{\alpha^\infty})}{(\alpha^\infty)^{-1} \exp(-\frac{|x_i^\infty|}{\alpha^\infty}) + (\beta^\infty)^{-1} \exp(-\frac{|x_i^\infty|}{\beta^\infty})}, \quad (11)$$

$$\alpha^\infty = \sum_{i=1}^n \frac{\pi_i^\infty |x_i^\infty|}{\sum_{j=1}^n \pi_j^\infty}, \quad \beta^\infty = \sum_{i=1}^n \frac{(1 - \pi_i^\infty) |x_i^\infty|}{\sum_{j=1}^n (1 - \pi_j^\infty)}. \quad (12)$$

4. NUMERICAL EXPERIMENTS

In this section iterative shrinkage-thresholding methods are compared with their versions augmented by Bayesian tuning, in terms of convergence times and empirical probability of reconstruction. We consider the problem in (1) in absence of noise and present experiments for signals with different priors.

As a first experiment we consider Bernoulli-Uniform signals [24]. More precisely, the signal to be recovered has length $n = 560$ with $k = 56$ nonzero elements drawn from a $\mathcal{U}([-2, -1] \cup [1, 2])$, respectively. The sensing matrix A with $m = 350$ rows is sampled from the Gaussian ensemble with zero mean and variance $1/m$. The mixture parameters have been initialized as follows: $\lambda = 10^{-3}$, $\tau = 0.2$, $\alpha^{(0)} =$

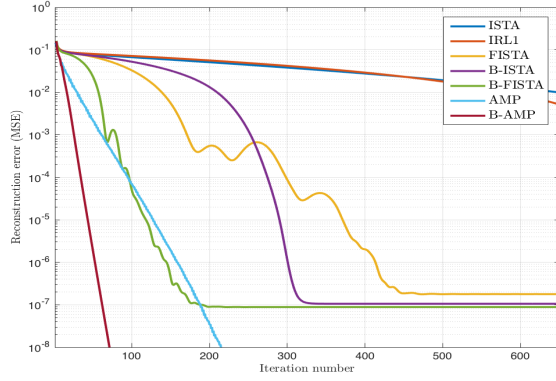


Fig. 1. Sparse Bernoulli signals: Evolution of the MSE for classical thresholding methods algorithms and the corresponding versions with Bayesian tuning.

$1, \pi^{(0)} = \mathbf{1}$ and $K = 61$. In Fig. 4, we compare the convergence rate of ISTA, FISTA, IRL1, AMP with the corresponding methods with Bayesian tuning (B-ISTA, B-FISTA, and B-AMP). In particular, the mean square error (MSE) of the iterates $\text{MSE}(t) = \|x^{(t+1)} - x^*\|^2/n$ averaged over 100 instances is depicted as a function of the iteration number. It should be noted that incorporating the soft support detection improves the reconstruction and the resulting algorithms are much faster than classical iterative shrinkage-thresholding methods.

In the second experiment we take the fraction of the nonzero coefficients fixed to $\rho = k/n$ and we study the effect of the nonzero coefficients distribution on the rate of convergence. More precisely, $x_i^* \sim (1 - \rho)\delta_0(x_i^*) + \rho g(x_i^*)$ where δ_0 is the Dirac delta function and g is a probability distribution function. In Table 1 the acronyms of the considered distributions are summarized (see also [25]).

Notation	g
5P1	$\mathbb{P}(x = -1) = \mathbb{P}(x = 1) = 0.3$ $\mathbb{P}(x = -5) = \mathbb{P}(x = +5) = 0.2$
U1	$\mathcal{U}[0, 4]$
G1	$\mathcal{N}(0, 4)$

Table 1. Nonzero coefficients distribution

Figures 2-4 (left) show the empirical recovery success rate, averaged over 100 experiments, as a function of the signal sparsity, for different signal priors. For all recovery algorithms, the convergence tolerance has been fixed to 10^{-4} . Also in this case the elements of matrix A with $m = 350$ are sampled from a normal distribution with variance $1/m$. It should be noticed that the Bayesian tuning improves the performance of iterative shrinkage-thresholding methods in terms of sparsity-undersampling tradeoff. In Figures 2-4 (right) the average running times of the algorithms computed

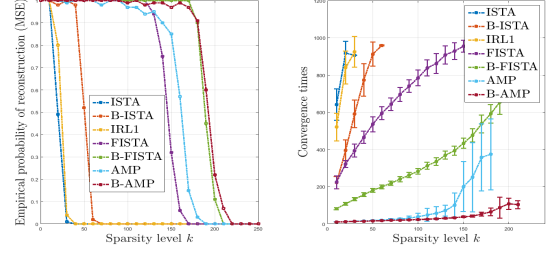


Fig. 2. Sparse 5P1-signals: Performance (left) and speed of convergence (right).

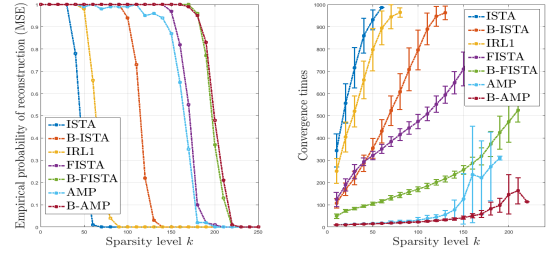


Fig. 3. Sparse U1-signals: Performance (left) and speed of convergence (right).

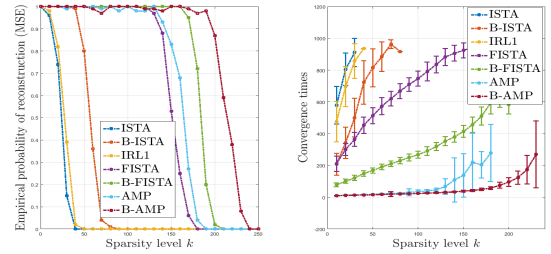


Fig. 4. Sparse G1-signals: Performance (left) and speed of convergence (right).

over the successful experiments are shown; the error bar represents the standard deviation of uncertainty. In all tested cases, the gain of the Bayesian tuning ranges from 2 to over 6 times, depending on the signal prior.

5. CONCLUDING REMARKS

In this paper, we proposed a Bayesian tuning for the class of iterative shrinkage-thresholding algorithms. Iterative procedures have been designed by combining MAP estimation with classical iterative thresholding methods, and they allow to perform both support detection and estimation of sparse signals. The main theoretical contribution includes the proof of convergence of B-ISTA to a fixed point. Numerical simulations show that these new algorithms are faster than classical ones and outperform related algorithms based on iteratively reweighted ℓ_1 -minimization in terms of phase transitions.

6. REFERENCES

- [1] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207 – 1223, 2006.
- [2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289 – 1306, 2006.
- [3] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267 – 288, 1994.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [5] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413 – 1457, 2004.
- [6] E. T. Hale, W. Yin, and Y. Zhang, “A Fixed-Point Continuation Method for ℓ_1 -Regularized Minimization with Applications to Compressed Sensing,” Rice University, Tech. Rep., Jul. 2007.
- [7] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” *Trans. Sig. Proc.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.
- [8] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Img. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.
- [9] M. Fornasier, *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Radon Series on Computational and Applied Mathematics, 2010.
- [10] A. Maleki and D. Donoho, “Optimally tuned iterative reconstruction algorithms for compressed sensing,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 330–341, April 2010.
- [11] D. L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: I. motivation and construction,” *CoRR*, vol. abs/0911.4219, 2009.
- [12] T. Goldstein, C. Studer, and R. Baraniuk, “A field guide to forward-backward splitting with a FASTA implementation,” *arXiv eprint*, vol. abs/1411.3406, 2014.
- [13] C. Hegde, M. F. Duarte, and V. Cevher, “Compressive sensing recovery of spike trains using structured sparsity,” in *Proceedings of the Workshop on Signal Processing with Adaptive Sparse Representations (SPARS)*, Saint Malo, France, Apr. 2009.
- [14] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Trans. Inf. Theor.*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [15] H. Mansour and R. Saab, “Weighted one-norm minimization with inaccurate support estimates: Sharp analysis via the null-space property,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015.
- [16] P. L. Combettes and V. R. Wajs, “Signal Recovery by Proximal Forward-Backward Splitting,” *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [17] K. Bredies and D. Lorenz, “Linear convergence of iterative soft-thresholding,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 813–837, 2008.
- [18] A. Montanari, “Graphical models concepts in compressed sensing,” *CoRR*, vol. abs/1011.4328, 2010.
- [19] E. Candès, M. Wakin, and S. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [20] M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz, “Recovering compressively sampled signals using partial support information,” *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1122–1134, 2012.
- [21] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, “Weighted ℓ_1 minimization for sparse recovery with prior information,” in *Proceedings of the 2009 IEEE International Conference on Symposium on Information Theory - Volume 1*, 2009, pp. 483–487.
- [22] 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.
- [23] C. Guo and M. E. Davies, “Near optimal compressed sensing without priors: Parametric sure approximate message passing,” *IEEE Transactions on Signal Processing*, vol. 63, no. 8, pp. 2130–2141, 2015.
- [24] C. Ravazzi and E. Magli.
- [25] A. Maleki, “Approximate message passing algorithms for compressed sensing,” Ph.D. dissertation, Stanford University, 2010.