

Non-convex approach to binary compressed sensing

Original

Non-convex approach to binary compressed sensing / Fosson, Sophie M.. - 2018-:(2019), pp. 1959-1963. (Intervento presentato al convegno 52nd Asilomar Conference on Signals, Systems and Computers, ACSSC 2018 tenutosi a Asilomar, CA, USA nel 2018) [10.1109/ACSSC.2018.8645293].

Availability:

This version is available at: 11583/2729893 since: 2019-04-03T10:38:27Z

Publisher:

IEEE Computer Society

Published

DOI:10.1109/ACSSC.2018.8645293

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

IEEE postprint/Author's Accepted Manuscript

©2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

Non-convex approach to binary compressed sensing

Sophie M. Fosson

November 8, 2018

Abstract

We propose a new approach for the recovery of binary signals in compressed sensing, based on the local minimization of a non-convex cost functional. The desired signal is proved to be a local minimum of the functional under mild conditions on the sensing matrix and on the number of measurements. We develop a procedure to achieve the desired local minimum, and, finally, we propose numerical experiments that show the improvement obtained by the proposed approach with respect to classical convex methods.

1 Introduction

Binary compressed sensing (BCS, [1]) refers to compressed sensing (CS, [2]) in the case of sparse binary signals. We can mathematically formulate BCS as follows: recover the sparsest $x \in \{0, 1\}^n$ from $y = Ax$, given $A \in \mathbb{R}^{m,n}$ with $m < n$. This is the simplest formulation of the most general problem of CS for finite-valued or discrete signals, that is, signals whose components are known to belong to a finite or discrete alphabet. The first theoretical analyses on this topic are very recent [3], despite the relevance in practical applications is widespread, e.g., in digital communications, wireless communications, sensor networks, digital image processing, spectrum sensing, localization, and quantized systems.

It is worth noticing that the classical recovery algorithms for CS can be applied to finite-valued signals as well, but they are not always prone to embed prior information on the discrete nature of the signal, which is expected to improve the accuracy of the solution. On the other hand, the discrete nature might

lead to combinatorial approaches, which turn out to be computationally burdensome. In this paper, we instead show that solving a non-convex problem over the convex hull of the alphabet can lead to the exact recovery. In [3], the problem is tackled by solving the basis pursuit [4] on the convex hull of the alphabet, and theoretical results on phase transition and stability are provided under null space properties.

In this paper, we propose a new formulation of the problem, introducing a cost functional tailored for binary signals in $\{0, 1\}$. For simplicity, we restrict the problem to the noise-free setting. Generalizations to large alphabets and noisy settings are under study. We however notice that the binary case is itself relevant, and BCS has drawn some attention in the last years, see, e.g., [1, 5–10]. In these papers, methods that take into account the prior information on the binary nature of the signal are proposed, and are shown to perform better than classical CS strategies in numerical experiments. Moreover, the problem of recovering binary signals is widespread in different frameworks, e.g., localization [11], hybrid systems [12], and jump linear systems [13], with possible applications to fault detection.

Our main contribution is the introduction and analysis of a cost functional suitable for BCS. In particular, we prove that a minimum is exactly the desired signal. A drawback of our approach is the non-convexity of the functional; however, we develop a suitable iterative procedure, which can be used to look for the the desired minimum. This method is shown to improve the recovery with respect to classical CS convex approach.

The paper is organized as follows. In Section 2, we introduce the problem. In Section 3 we prove

our main theoretical guarantees. In Section 4, we describe the recovery algorithm and show some simulation results. Finally, we draw some conclusions.

2 Problem statement

In this work, we propose a new efficient approach to BCS. Specifically, we consider the following cost functional:

$$\mathcal{F} : [0, 1]^n \rightarrow \mathbb{R}^+$$

$$\mathcal{F}(x) := \frac{1}{2} \|y - Ax\|_2^2 + \lambda \sum_{i=1}^n \left(x_i - \frac{1}{2} x_i^2 \right), \quad \lambda > 0. \quad (1)$$

$\mathcal{F}(x)$ is similar to the popular Lasso functional [4, 14], but it presents a concave penalty $g(x) = \sum_{i=1}^n (x_i - \frac{1}{2} x_i^2)$, which belongs to the family of minimax concave penalties (MCP). In the last years, the MCP family has been used and analyzed for sparse recovery and variable selection [15–17], and for in-network recovery of jointly sparse signals [18].

Even though other concave penalties are more popular in the literature (for example, the log penalty [19]), $g(x)$ is preferable for its quadratic polynomial structure, which makes it mathematically more tractable. For example, it is possible to compute its global minimum by quadratic programming or other techniques of polynomial optimization [20], even though this is computationally burdensome. This point is not examined in this paper, but may be considered for future developments. As we explain in the next section, in this work we rather focus on fast algorithms to find the desired local minima of \mathcal{F} .

3 Theoretical guarantees

In this section, we show that finding a (local) minimum of (1) can lead to the exact recovery of the unknown binary, sparse signal, under mild assumptions. Specifically, our assumptions are weaker than usual conditions required in CS: neither coherence conditions nor restricted isometry property nor null space property [4] are involved in our analysis. The

main results can be summarized as follows: the true $\tilde{x} \in \{0, 1\}^n$ is a local minimum of (1), and it is the unique signal in $\{0, 1\}^n$ which is a local minimum of \mathcal{F} over the convex hull $[0, 1]^n$. Therefore, using any descent algorithm that achieves a local minimum, if the achieved value is binary, then it is the exact signal.

The non-convex \mathcal{F} can have multiple minima, thus finding the right solution may be not easy. However, the following points should be considered.

- Fast algorithms are available to get local minima, that can be run multiple times starting from different initial points. This increases the chance of getting the right solution.
- The non-convexity of \mathcal{F} is described by its Hessian matrix $A^T A - \lambda I$, which has $n - m$ negative eigenvalues (assuming λ sufficiently small). Increasing the number of measurements m we reduce the number of negative eigenvalues, thus we somehow mitigate the non-convexity of \mathcal{F} (which sometimes is even convex in $[0, 1]^n$).

These observations suggest that suitable design of the system and implementation of the recovery method reinforce the possibility of exact reconstruction. In this paper, these points are verified via numerical simulations, while a theoretical investigation is left for future extended work.

Let us now prove the theoretical guarantees.

Let $\tilde{x} \in \{0, 1\}^n$, with support S and sparsity level k , and let $y = A\tilde{x}$. We indicate by $A_{[m]}$ any submatrix of A obtained by selecting m columns, while we write $Q \succ 0$ to say that matrix Q is positive definite. We recall that a family of vectors $\{v_1, \dots, v_n\} \subset \mathbb{R}^m$, $m < n$, are said to be in general position if any $\sum_i \sigma_i v_i$, with $\sigma_i \in \{0, \pm 1\}$, is different from $\pm v_j$, for any $j = 1, \dots, n$ [21].

Assumption 1.

- $A_{[m]}^T A_{[m]} - \lambda I \succ 0$;
- the sparsity level is not larger than the number of measurements: $k \leq m$;
- the columns of A are in general position.

Conditions (a) and (c) are usual: for example, any random matrix with entries generated according to a continuous distribution satisfy them with probability one [21]. Instead, condition (b) is a necessary requirement to recover a sparse signal (in CS theory, at least $m \geq ck \log(\frac{n}{m})$ for some $c > 0$ is required [4]).

Theorem 1. *Let $\tilde{x} \in \{0, 1\}^n$, with support S and sparsity level k , and let $y = A\tilde{x}$. Under assumptions 1.(a)-(b), \tilde{x} is an isolated local minimum of $\mathcal{F}(x)$ as defined in (1).*

Proof. Let $\tilde{x} \in \{0, 1\}^n$ and let $h \in \mathbb{R}^n$ be any small increment such that $\tilde{x} + h \in [0, 1]^n$. Our goal is to show that $\mathcal{F}(\tilde{x} + h) > \mathcal{F}(\tilde{x})$ for any $h \neq 0$. Let S be the support of \tilde{x} , and $S^C = \{1, \dots, n\} \setminus S$.

$$\begin{aligned} \mathcal{F}(\tilde{x} + h) - \mathcal{F}(\tilde{x}) &= \\ &= \frac{1}{2} \|Ah\|_2^2 + \lambda \sum_{i=1}^n h_i - \frac{\lambda}{2} \|h\|_2^2 - \lambda \langle \tilde{x}, h \rangle \\ &= \frac{1}{2} \|Ah\|_2^2 + \lambda \sum_{i=1}^n h_i - \frac{\lambda}{2} \|h\|_2^2 - \lambda \sum_{i \in S} h_i \\ &= \frac{1}{2} \|Ah\|_2^2 + \lambda \sum_{i \in S^C} h_i - \frac{\lambda}{2} \|h\|_2^2. \end{aligned}$$

In particular, let us notice that $h_i \geq 0$ when $i \in S^C$ (while $h_i \leq 0$ when $i \in S$).

We now prove that $f(h) = \frac{1}{2} \|Ah\|_2^2 + \lambda \sum_{i \in S^C} h_i - \frac{\lambda}{2} \|h\|_2^2 > 0$ for any $h \neq 0$ with sufficiently small magnitude. This is sufficient to obtain the thesis. Let us say that $|h_i| < \epsilon$, where ϵ will be assessed in a while.

Since f is differentiable, we compute its gradient to look for minima in the interior part of its domain. We easily obtain a unique stationary point $-(A^T A - \lambda I)^{-1} \lambda v$ where $v \in \{0, 1\}^n$ has entries equal to 1 in S^C and zero elsewhere. Assuming that λ is smaller than the positive eigenvalues of $A^T A$, the Hessian matrix $A^T A - \lambda I$ has both positive and negative eigenvalues (in particular, the negative ones are equal to $-\lambda$). Therefore, the stationary point is a saddle. In order to find minima, we then move on the boundaries. Let $A_{[m]}$ be any selection of m columns of A . We assume that $A_{[m]}^T A_{[m]} - \lambda I \succ 0$ (this is always false for any $A_{[j]}$, $j > m$) according

to Assumption 1.(a). Therefore, $n - m$ entries must be on boundaries ($\{0, \epsilon\}$ for h_{S^C} , $\{-\epsilon, 0\}$ for h_S) to have a candidate minimum. Fixed, these $n - m$ entries, let Ω be the set of the remaining m entries: we compute the gradient on Ω to look for the minimum. We obtain $h_\Omega = (A_\Omega^T A_\Omega - \lambda I)^{-1} (A_\Omega^T A_{\Omega^C} h_{\Omega^C} - \lambda v)$ where $v \in \{0, 1\}^m$ is 1 on $S^C \cap \Omega$. By Assumption 1.(a), $A_\Omega^T A_\Omega - \lambda I \succ 0$, therefore this corresponds to a minimum. However, since the entries of h_{Ω^C} are in $\{0, \pm\epsilon\}$, we can choose an ϵ small enough so that the entries of h_Ω all have magnitude greater than ϵ . More precisely, if ϵ is much smaller than λ , h_Ω can be approximated by $(A_\Omega^T A_\Omega - \lambda I)^{-1} (-\lambda v) = -(\lambda A_\Omega^T A_\Omega - I)^{-1} v$, which in turn can be approximated by v if λ is small. In conclusion the so-computed candidate minimum is outside the domain, and also the entries over Ω should be on the boundaries. In conclusion, we are observing that a candidate minimum for f is in $\{0, \pm\epsilon\}^n$.

At this point, we notice that if $h_{S^C} = 0$, then $f(h) = \frac{1}{2} \|A_S h_S\|_2^2 - \frac{\lambda}{2} \|h_S\|_2^2$, which has minimum for $h_S = 0$. On the other hand, if there exists a $j \in S^C$ such that $h_j = \epsilon$, this is sufficient to have $f(h) > \lambda(\epsilon - \frac{k}{2}\epsilon^2)$, which is positive for any $\epsilon < \frac{2}{k}$ (which again can be assumed for the arbitrariness of ϵ). This proves that $f(h) \geq 0$ for any sufficiently small h , and $f(h) = 0$ if and only if $h = 0$.

As a consequence, $\mathcal{F}(\tilde{x} + h) > \mathcal{F}(\tilde{x})$ for any $h \neq 0$. \square

Theorem 2. *Let $\tilde{x} \in \{0, 1\}^n$, with support S and sparsity level k , and let $y = A\tilde{x}$. Under assumptions 1.(a)-(b)-(c), if λ is sufficiently small, then \tilde{x} is the global minimum of $\mathcal{F}(x)$ defined in (1) over $\{0, 1\}^n$.*

Proof. Let us consider $f(h)$ as defined in the proof of Theorem 1. In order to investigate the minima of \mathcal{F} over $\{0, 1\}^n$, we consider $h_S \in \{0, -1\}^k$ and $h_{S^C} \in \{0, 1\}^{n-k}$. Therefore, $f(h) \geq \frac{1}{2} \|Ah\|_2^2 - \lambda \frac{k}{2}$. Now, for any $h \neq 0$, if the columns of A are in general position, then $\frac{1}{2} \|Ah\|_2^2$ has a positive value. Therefore, we can always assume $\lambda < \frac{\|Ah\|_2^2}{k}$, which proves that $f(h) > 0$. This yields to $\mathcal{F}(z) > \mathcal{F}(\tilde{x})$ for any $z \in \{0, 1\}^n$. \square

The following theorem reinforces this result by proving that \tilde{x} is the unique minimum over $\{0, 1\}^n$.

Theorem 3. *Let us consider any $z \in \{0, 1\}^n$, $z \neq \tilde{x}$. Under assumptions 1.(a)-(b)-(c), if λ is sufficiently small, then, z is not a minimum of \mathcal{F} .*

Proof. Let us consider any $z \in \{0, 1\}^n$, $z \neq \tilde{x}$. We prove that there exists a direction h along which $\mathcal{F}(z + h) < \mathcal{F}(z)$. First, we have:

$$\begin{aligned} \mathcal{F}(z + h) - \mathcal{F}(z) &= \\ &= \frac{1}{2} \|Ah\|_2^2 + \langle Ah, A(z - \tilde{x}) \rangle + \lambda \sum_{i \in S^c} h_i - \frac{\lambda}{2} \|h\|_2^2. \end{aligned} \quad (2)$$

Now, let us define $h = -\epsilon(z - \tilde{x})$, where $\epsilon > 0$. This h is an admissible increment, since if $z_i = 1$, then $h \leq 0$, while if $z_i = 0$, then $h \geq 0$; therefore, for $\epsilon \leq 1$, $z + h \in [0, 1]^n$. Substituting this h in (2), we obtain:

$$\begin{aligned} \mathcal{F}(z + h) - \mathcal{F}(z) &\leq \frac{\epsilon^2}{2} \|A(z - \tilde{x})\|_2^2 - \epsilon \|A(z - \tilde{x})\|_2^2 + \\ &+ \lambda \epsilon \|z - \tilde{x}\|_2^2 - \frac{\lambda}{2} \epsilon^2 \|z - \tilde{x}\|_2^2 = \\ &= \epsilon \left(\frac{\epsilon^2}{2} - 1 \right) \left(\|A(z - \tilde{x})\|_2^2 - \lambda \|z - \tilde{x}\|_2^2 \right) \end{aligned} \quad (3)$$

where we use the fact that $\|\cdot\|_1$ is equal to $\|\cdot\|_2^2$ if the argument is binary. As the columns of A are in general position, then $\|A(z - \tilde{x})\|_2^2$ is a positive value, and we can always choose λ so that $\|A(z - \tilde{x})\|_2^2 - \lambda \|z - \tilde{x}\|_2^2 > 0$. Thus, the last expression in (3) is negative for any $\epsilon < \sqrt{2}$. Since this is true for any arbitrarily small $\epsilon > 0$, we conclude that $\mathcal{F}(z + h) - \mathcal{F}(z) < 0$ if the direction of h is $z - \tilde{x}$. This proves that z cannot be a minimum. \square

4 Numerical results

In this section, we propose some numerical simulations that support the efficiency of our approach.

4.1 Algorithm

An algorithm that can be used to compute a (local) minimum of \mathcal{F} in $[0, 1]^n$ is the iterative reweighting procedure (abbreviated as RW) introduced in Algorithm 1.

Algorithm 1 Reweighting (RW)

- 1: Initialize: $x(0) \in [0, 1]^n$; $\lambda > 0$
- 2: **for all** $t = 1, \dots, T_{stop}$ **do**
- 3: Update weights:

$$w_i(t) = 1 - x_i(t) \text{ for any } i \in \{1, \dots, n\}$$

- 4: Lasso:

$$x(t+1) = \operatorname{argmin}_{x \in [0, 1]^n} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \sum_{i=1}^n w_i(t) x_i$$

- 5: **end for**
-

$1 - x_i(t+1)$ is the derivative of $g(x_i(t+1)) = x_i(t+1) - \frac{1}{2} x_i(t+1)^2$. In the literature, such reweighting method has been studied for different concave penalties (e.g., $\log(|x| + \epsilon)$, $(|x| + \epsilon)^q$, $q \in (0, 1)$) and proved to reach a local minimum of concave-penalized functionals of kind $\frac{1}{2} \|Ax - y\|_2^2 + \lambda \sum_i g(|x_i|)$ [19, 22–24]. Even though limited to local minimization, practical experiments show that reweighting algorithms are generally very efficient and produces accurate estimates in many applications, e.g., magnetic resonance imaging, sensor selection and distributed CS in sensor networks [18, 19, 25–28].

For our experiments, we use reweighting, which turns out to be efficient. We however notice that other approaches might be tested, ranging from alternating minimization (subproblems of dimensions m are convex) and block-coordinate descent to the alternating direction method of multipliers (ADMM, [29]). We leave a comparison analysis for future work.

4.2 Results

We consider the following setting for our simulations¹. The desired signal is $\tilde{x} \in \{0, 1\}^n$ with $n = 100$ and sparsity level $k = 5$. m measurements are taken, $y = A\tilde{x}$, through a Gaussian sensing matrix $A \sim \mathcal{N}(0, \frac{1}{m})$, $m \in [10, 40]$. The Lasso step in Algorithm 1 is iteratively solved via ADMM, which is stopped when the sum of squared primal and dual residuals is below 10^{-6} [29]. We study two possibilities: the first one, indicated as RW, is the ℓ_1 reweighting procedure introduced in Algorithm 1 with initial condition $x(0) = 0$; in the second one, indicated by RWR, we re-run RW when the obtained solution is verified to be not the desired one via Theorem 3, restarting from $x(0) \in [0, 1]^n$ generated uniformly at random. This is an heuristic method to test different initializations over $[0, 1]^n$. We set $T_{stop} = 4$ (as noticed in [19], the most of benefit is generally obtained in the first reweighting iterations), while the maximum number of re-initializations in RWR is fixed to 20. We compare the proposed method to Lasso and Basis Pursuit (BP) [4], both solved via ADMM. BP is conceived for the free-noise case, while Lasso is known to have a bias. Lasso and BP do not envisage the prior information on the binary nature of the signal, but one can quantize the obtained solution over $\{0, 1\}$: we compute the performance metrics in both standard and quantized solutions. Finally, we notice that, for all the methods (RW, RWR, Lasso, BP), when k is known, we can add the equation $\sum_i x_i = k$ to the system $Ax = y$, which is expected to improve the performance. The following performance metrics are evaluated (we indicate by \hat{x} the estimate of \tilde{x}):

- relative square error $RSE = \|\hat{x} - \tilde{x}\|_2^2 / \|\tilde{x}\|_2^2$;
- false positive rate, that is, the normalized number of occurrences $\hat{x}_i \neq 0$ when $\tilde{x}_i = 0$;
- false negative rate, that is, the normalized number of occurrences $\hat{x}_i = 0$ when $\tilde{x}_i \neq 0$;
- exact recovery, which is defined as $RSE < 10^{-3}$ and no false positives/negatives;

¹Code available at <https://github.com/sophie27/Non-convex-approach-to-binary-compressed-sensing>

- total number of ADMM iterations.

The results, displayed in Figure 1, are averaged over 500 runs. When k is unknown (left column), RW and RWR always achieve better performance than classical methods when $m \geq 15$. In particular, the random re-initialization in RWR gives a substantial gain, and always gets the exact solution at $m = 25$, where classical methods do not overpass 90% of success. When k is known (thus equation $\sum_i x_i = k$ is added), as expected, the general recovery accuracy is improved. In this case, BP is as accurate as RW, while again RWR performs better, achieving 100% of success at $m = 20$.

The price of the improvement obtained by RW and RWR is the increased number of iterations. Future work will consider the development of faster strategies to get the desired local minimum. However, ADMM iterations are low-complex and the time spent for these experiments is acceptable (10^5 iterations require less than 20 seconds on a CPU @ 1.80GHz, RAM 16Gb).

5 Conclusion

In this paper, we have introduced a new efficient framework to tackle the recovery of sparse binary signals acquired according to the compressed sensing paradigm. We have formulated the problem as local minimization of a non-convex, polynomial cost functional, which has a (local) minimum that corresponds to the desired conditions under very mild conditions. The search of such local minimum can be efficiently performed via iterative algorithms, such as reweighting procedures. In future work, we will study the conditions under which the desired minimum is the global minimum of the functional, and we will extend the approach to larger, non-binary alphabets and to systems with noise.

References

- [1] U. Nakarmi and N. Rahnavard, “BCS: Compressive sensing for binary sparse signals,”

- in *IEEE Military Communications Conference (MILCOM)*, 2012, pp. 1–5.
- [2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] S. Keiper, G. Kutyniok, D. G. Lee, and G. E. Pfander, “Compressed sensing for finite-valued signals,” *Linear Algebra and its Applications*, vol. 532, no. Supplement C, pp. 570–613, 2017.
- [4] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York: Springer, 2013.
- [5] M. Stojnic, “Recovery thresholds for ℓ_1 optimization in binary compressed sensing,” in *IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 1593–1597.
- [6] Z. Tian, G. Leus, and V. Lottici, “Detection of sparse signals under finite-alphabet constraints,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 2349–2352.
- [7] M. Shirvanimoghaddam, Y. Li, B. Vucetic, J. Yuan, and P. Zhang, “Binary compressive sensing via analog fountain coding,” *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6540–6552, 2015.
- [8] J. H. Ahn, “Compressive sensing and recovery for binary images,” *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4796–4802, 2016.
- [9] N. Lee, “Map support detection for greedy sparse signal recovery algorithms in compressive sensing,” *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 4987–4999, 2016, support detection method, and a sufficient condition for perfect signal recovery is derived for the case when the sparse signal is binary.
- [10] T. Liu and D. G. Lee, “Fast binary compressive sensing via ℓ_0 gradient descent,” 2018, <https://arxiv.org/pdf/1801.09937.pdf>. [Online]. Available: <https://arxiv.org/pdf/1801.09937.pdf>
- [11] A. Bay, D. Carrera, S. M. Fosson, P. Fragneto, M. Grella, C. Ravazzi, and E. Magli, “Block-sparsity-based localization in wireless sensor networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 182, pp. 1–15, 2015.
- [12] S. M. Fosson, “Binary input reconstruction for linear systems: A performance analysis,” *Non-linear Analysis: Hybrid Systems*, vol. 7, no. 1, pp. 54 – 67, 2013.
- [13] F. Fagnani and S. M. Fosson, “Analysis of reduced-search bcjr algorithms for input estimation in a jump linear system,” *Signal Processing*, vol. 108, pp. 341 – 350, 2015.
- [14] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1996.
- [15] C.-H. Zhang, “Nearly unbiased variable selection,” *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.
- [16] C.-H. Zhang and T. Zhang, “A general theory of concave regularization for high-dimensional sparse estimation problems,” *Statist. Sci.*, vol. 27, no. 4, pp. 576 –593, 2012.
- [17] J. Woodworth and R. Chartrand, “Compressed sensing recovery via nonconvex shrinkage penalties,” *Inverse Problems*, vol. 32, no. 7, pp. 75 004–75 028, 2016.
- [18] S. M. Fosson, J. Matamoros, C. Antón-Haro, and E. Magli, “Distributed recovery of jointly sparse signals under communication constraints,” *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3470–3482, 2016.
- [19] E. J. Candès, M. B. Wakin, and S. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journ. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [20] J.-B. Lasserre, *An introduction to polynomial and semi-algebraic optimization*, ser. Cambridge texts in applied mathematics. Cambridge, UK: Cambridge University Press,, 2015.

- [21] R. J. Tibshirani, “The Lasso problem and uniqueness,” *Electronic Journal of Statistics*, vol. 7, pp. 1456–1490, 2013.
- [22] H. Zou and R. Li, “One-step sparse estimates in nonconcave penalized likelihood models,” *Annals of Statistics*, vol. 36, no. 4, p. 1509, 2008.
- [23] M. Fazel, H. Hindi, and S. Boyd, “Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices,” in *IEEE Proc. American Control Conference (ACC)*, vol. 3, 2003, pp. 2156–2162.
- [24] S. M. Fosson, “A biconvex analysis for lasso ℓ_1 reweighting,” *IEEE Signal Process. Lett.*, vol. early access, no. nn, pp. 1–1, 2018.
- [25] M. Asif and J. Romberg, “Fast and accurate algorithms for re-weighted ℓ_1 -norm minimization,” *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5905–5916, 2013.
- [26] R. Ahmad and P. Schniter, “Iteratively reweighted ℓ_1 approaches to sparse composite regularization,” *IEEE Trans. Computat. Imag.*, vol. 1, no. 4, pp. 220–235, 2015.
- [27] M. Calvo-Fullana, J. Matamoros, C. Antón-Haro, and S. M. Fosson, “Sparsity-promoting sensor selection with energy harvesting constraints,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3766–3770.
- [28] S. M. Fosson, J. Matamoros, C. Antón-Haro, and E. Magli, “Distributed support detection of jointly sparse signals,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 6434–6438.
- [29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1 – 122, 2010.

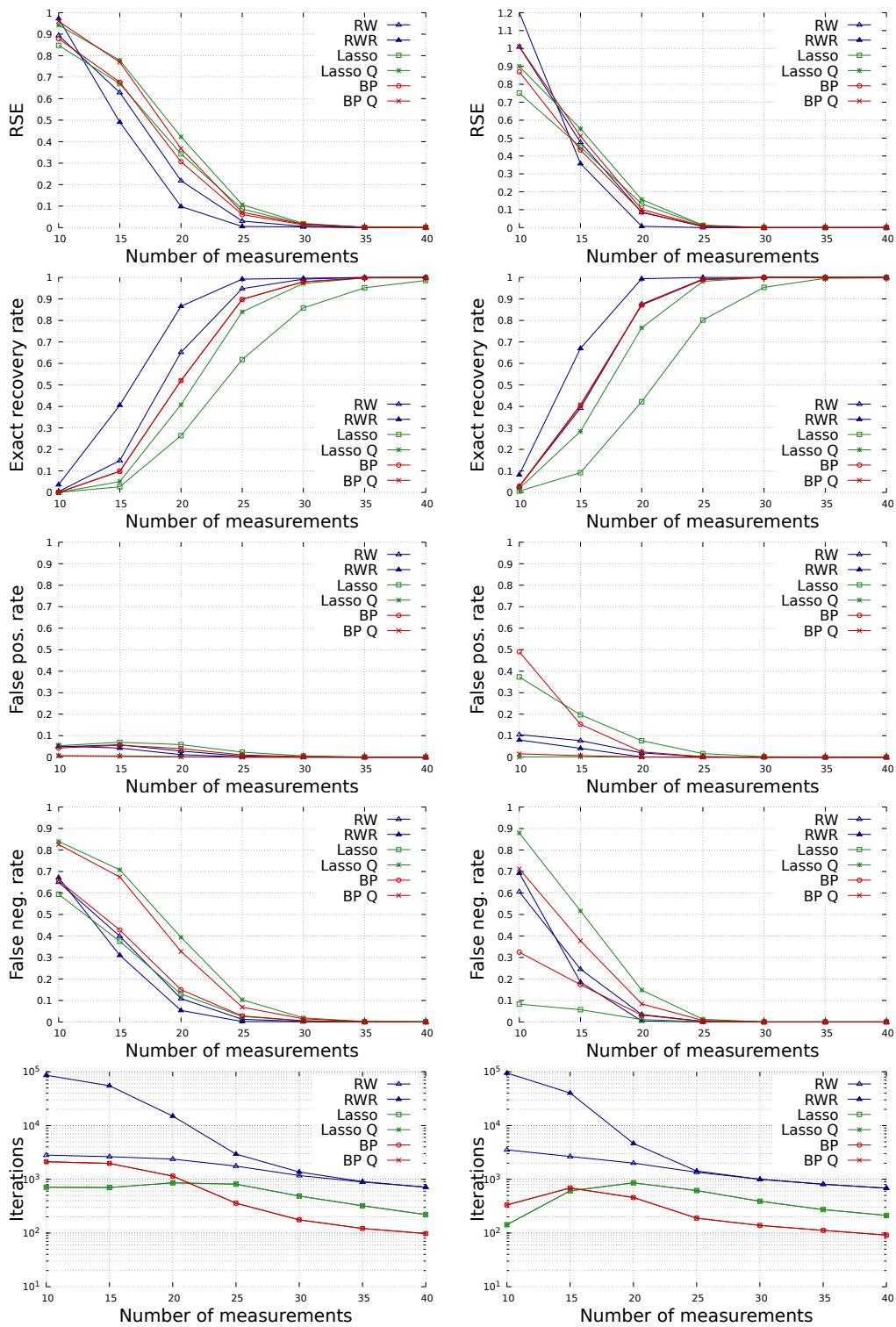


Figure 1: Noise-free measurements, $n = 100$, $k = 5$.⁸ Left column: unknown k , right column: known k . For Lasso, RW, RWR: $\lambda = 10^{-2}$. Lasso Q and BP Q refer to the solutions obtained by quantizing the Lasso and BP solutions over $\{0, 1\}^n$ (this is particularly useful for Lasso, which produces biased solutions in the absence of noise).