From L1 Minimization to Entropy Minimization: A Novel Approach for Sparse Signal Recovery in Compressive Sensing

Miguel Heredia Conde, Otmar Loffeld
Center for Sensorsystems (ZESS)
University of Siegen, Paul-Bonatz-Straße 9-11
57076 Siegen, Germany
E-mail: {heredia, loffeld}@zess.uni-siegen.de

***Abstract*—The groundbreaking theory of compressive sensing (CS) enables reconstructing many common classes or real-world signals from a number of samples that is well below that prescribed by the Shannon sampling theorem, which exclusively relates to the bandwidth of the signal. Differently, CS takes profit of the *sparsity* or *compressibility* of the signals in an appropriate basis to reconstruct them from few measurements. A large number of algorithms exist for solving the sparse recovery problem, which can be roughly classified in *greedy* pursuits and l_1 minimization algorithms. Chambolle and Pock's (C&P) primal-dual l_1 minimization algorithm has shown to deliver state-of-the-art results with optimal convergence rate. In this work we present an algorithm for l_1 minimization that operates in the null space of the measurement matrix and follows a Nesterov-accelerated gradient descent structure. Restriction to the null space allows the algorithm to operate in a minimal-dimension subspace. A further novelty lies on the fact that the cost function is no longer the l_1 norm of the temporal solution, but a weighted sum of its entropy and its l_1 norm. The inclusion of the entropy pushes the l_1 minimization towards a *de facto* quasi-l_0 minimization, while the l_1 norm term avoids divergence. Our algorithm globally outperforms C&P and other recent approaches for l_1 minimization in terms of l_2 reconstruction error, including a different entropy-based method.**

I. Introduction

The Shannon sampling theorem states that a function, defined in time domain over a certain interval and whose frequency spectrum is bounded, is fully determined by a set of equally-spaced temporal samples at a rate that is twice the maximum frequency contained in the signal, also known as *Nyquist* rate. Obviously, in acquisition systems where a large bandwidth is desired this translates into massive data streams, with the corresponding requirements in terms of storage and communications. The groundbreaking theory of *compressive* (or *compressed*) sensing (CS) goes a step further and states that signal reconstruction is feasible from a number of measurements that is no longer directly linked to the signal bandwidth, but to the signal *sparsity* or *compressibility* in an appropriate basis. This means that signals exhibiting some structure, which can be represented using only few vectors from an appropriate basis or dictionary, can be exactly reconstructed from a number of measurements that relates linearly to the sparsity and is independent from the bandwidth of the signal.

The CS pipeline can be divided in two steps: first sensing in a *compressed* fashion, that is, condensing as much information of the signal in as few measurements as possible. Second, reconstructing the original signal from this reduced set of measurements. If no restrictions on the acquisition hardware are given, the first step can be considered solved. Random matrices, e. g., with coefficients drawn from *i.i.d.* Gaussian or Bernoulli distributions, have been proven to be good CS sensing matrices [1]. The remaining challenge is then reconstructing the original signal in its sparse representation from the measurements obtained using such sensing matrices. This means, in fact, solving a linearly-constrained l_0 minimization, which is an NP-hard problem. Fortunately, it can be proven that under certain conditions a linearly-constrained l_1 minimization converges to the same minimizer as its l_0 counterpart [2]. For this reason most methods for CS sparse signal recovery look for the solution with minimal l_1 norm that satisfies the measurements.

Of special interest are those methods operating in the null space of the measurement matrix, such as the l_1-minimizing Kalman filter [3], [4]. In this paper we propose a novel method that, also working in the null space, no longer minimizes the l_1 norm, but a weighted sum of the l_1 norm and an *entropy*-like function of the signal components. The proposed algorithm outperforms state-of-the-art approaches solving pure l_1 minimization in terms of sparse recovery error.

II. The Compressive Sensing Scenario

Differently from the Shannon sampling theorem, which requires the signal to be bandlimited, CS theory [5], [6] imposes the more general requirement of the signal being *sparse* in some basis or tight frame. If this holds and some additional requirements regarding the sensing scheme are satisfied, then the signal can be exactly recovered from few non-adaptive measurements. Let $\vec{x} \in \mathbb{C}^n$ be the discrete signal we want to recover, in its sparse representation. The l_0 norm of $\vec{x}$ is defined as:

$$\|\vec{x}\|_0 \coloneqq \lim_{p \to 0} \|\vec{x}\|_p^p = |\operatorname{supp}(\vec{x})| \quad (1)$$

ISBN 978-90-827970-1-5 © EURASIP 2018

that is, the cardinality of the support of $\vec{x}$, and $\vec{x}$ is called an s-sparse signal if $\|\vec{x}\|_0 \leq s$. Provided that the sparsity requirement is satisfied ($s \ll n$), the challenge is to reconstruct $\vec{x}$ from $m \ll n$ linear measurements. Thus, the measurement model is described by the underdetermined linear system:

$$\vec{y} = \boldsymbol{A}\vec{x} \tag{2}$$

where $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is the *measurement matrix*, which explains how the vector of measurements $\vec{y} \in \mathbb{C}^m$ relates to $\vec{x}$ and may be the composition of the actual *sensing matrix*, and a dictionary. As mentioned before, finding the sparsest $\vec{x}$ satisfying Eq. 2 is known to be NP-hard, and a common workaround is convexifying the problem turning the l_0 minimization into l_1, yielding:

$$\hat{\vec{x}} = \underset{\vec{x} \in \mathbb{C}^n}{\arg\min} \|\vec{x}\|_1 \text{ subject to } \vec{y} = \boldsymbol{A}\vec{x}, \tag{3}$$

The equality constraint in Eq. 3 can be implicitly enforced by conducting the minimization process within the *null space* or *kernel* of the matrix, defined as the set of vectors that are projected to zero by it, that is,

$$\mathcal{N}(\boldsymbol{A}) = \left\{ \vec{x} \mid \boldsymbol{A}\vec{x} = \vec{0} \right\} \tag{4}$$

Provided that Eq. 2 only imposes m constraints, the null space is an $n-m$-dimensional subspace of $\mathbb{C}^n$. Given any particular solution to Eq. 2, $\vec{x}_0$, Eq. 3 can be reformulated as:

$$\begin{aligned} \hat{\vec{x}} &= \underset{\vec{x} \in \{\mathcal{N}(\boldsymbol{A}) + \vec{x}_0\}}{\arg\min} \|\vec{x}\|_1 \\ &= \vec{x}_0 + \boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})}\hat{\vec{n}}, \text{ with:} \\ \hat{\vec{n}} &= \underset{\vec{n} \in \mathbb{C}^{n-m}}{\arg\min} \left\| \vec{x}_0 + \boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})}\vec{n} \right\|_1 \end{aligned} \tag{5}$$

where $\boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})} \in \mathbb{C}^{n \times (n-m)}$ denotes a basis of $\mathcal{N}(\boldsymbol{A})$.

III. Related Work

Primal-dual interior point methods have shown to be efficient tools for solving Eq. 3. Provided that the problem in Eq. 3 can be directly expressed as a linear program, the classical Newton method can be used to approach a solution, as in the l_1-magic library [7]. A prominent alternative is the Chambolle and Pock's (C&P) primal-dual algorithm [8], a first-order primal-dual method for convex optimization problems with saddle-point structure and convergence rate $\mathcal{O}(1/n)$. A novel adaptive approach for solving the constrained l_1 minimization problem based on the scale space method was proposed in [9].

Using a Kalman filter for estimating sequences of sparse signals from a reduced set of *compressed* measurements was initially proposed in [11]. Differently from [11], [12], where the sparse recovery algorithm is uncoupled from the probabilistic filter, in [13] the norm to minimize, typically the l_1 norm, is included as an additional measurement. Independently from the aforementioned works, a Kalman filter for solving Eq. 3 was proposed in [3] and further studied in [4], [14]. The l_1 norm of the temporal solution is the only measurement of the filter, which has the peculiarity of operating in $\mathcal{N}(\boldsymbol{A})$. At each iteration the algorithm tries to push down the l_1 norm of the temporal solution by incorporating a measurement that is slightly lower than the actual l_1 norm.

Apart from the previous approaches for l_1 minimization, a number of *greedy* algorithms exist that try first to estimate the signal support and then solve the resulting overdetermined problem. Known examples are Matching Pursuit (MP) [15], OMP [10], Order Recursive Matching Pursuit (ORMP) [16], Regularized Orthogonal Matching Pursuit (ROMP) [17], and Compressive Sampling Matching Pursuit (CoSaMP) [18], among many others. Greedy algorithms are specially appealing when the sparsity s is very low, since the number of iterations is directly given by s.

Reviewers pointed out some references where entropy-like priors are combined with sparsity-promoting penalties, e. g., in Nuclear Magnetic Resonance (NMR) spectrum reconstruction, but these works focus on entropy *maximization*, while we use entropy as a sparsity-promoting term itself, to be minimized. More related is the entropy-based algorithm in [19], which we include in our experimental evaluation.

IV. Sparse Signal Recovery via Entropy Minimization

Let $\mathcal{Z}$ be a discrete random variable with n possible values $z_1, z_2, \ldots, z_n$ and let $P(z_i)$ denote the probability of occurrence of z_i. Then the classical Shannon entropy is given by:

$$H(\mathcal{Z}) = -\sum_{i=1}^{n} P(z_i) \log P(z_i) \tag{6}$$

At this point we propose considering the sparse vector to recover $\vec{x}$ as a representation of some discrete probability density function, whose entropy is to be minimized. Despite this may seem to be an exotic interpretation, note that minimizing the entropy of a discrete probability distribution will concentrate the probability in only few discrete events. In fact, minimal (zero) entropy is attained when all probability is concentrated in a single event, since this is a sure event and entropy measures *unpredictability*. In our case this means looking for the signal support set of minimal cardinality, that is, *de facto* approaching a solution to Eq. 1. Formally, we establish the following correspondence:

$$P(z_i) = \frac{|x_i|}{\|\vec{x}\|_1}, \ \forall 1 \leq i \leq n \tag{7}$$

Using the change of variables in Eq. 7 we can write our entropy-like cost function acting directly over $\vec{x}$ as:

$$\begin{aligned} H_{\text{vec}}(\vec{x}) &= -\sum_{i=1}^{n} \frac{|x_i|}{\|\vec{x}\|_1} \log\left(\frac{|x_i|}{\|\vec{x}\|_1}\right) \\ &= \frac{1}{\|\vec{x}\|_1} \left(\|\vec{x}\|_1 \log \|\vec{x}\|_1 - \sum_{i=1}^{n} |x_i| \log |x_i| \right) \end{aligned} \tag{8}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

From both expressions in Eq. 8 it clearly follows that the minimum value $H_{\mathrm{vec}}(\vec{x}) = 0$ is attained when there is a single nonzero coefficient, say at position i_{nnz}, and we have that $\|\vec{x}\|_1 = |x_{i_{\mathrm{nnz}}}|$.

In this work we aim to solve the signal recovery problem in $\mathcal{N}(\boldsymbol{A})$ (Eq. 4), which ensures satisfaction of Eq. 2 and operation in a minimum-dimensional subspace. One might be tempted of substituting the l_1 norm by Eq. 8 in Eq. 5 directly. Note that this will not work, since $H_{\mathrm{vec}}(\cdot)$ is insensitive to the actual values of the coefficients x_i, whose amplitude can grow without bounds, due to the l_1 normalization in Eq. 7. To solve this issue we include also the l_1 norm in the cost function with a weighting parameter $0 < \alpha \leq 1$, yielding:

$$H_{l_1}(\vec{x}) = \frac{\alpha}{\|\vec{x}\|_1}\left(\|\vec{x}\|_1 \log \|\vec{x}\|_1 - \sum_{i=1}^{n} |x_i| \log |x_i|\right) + (1-\alpha)\|\vec{x}\|_1 \quad (9)$$

Eq. 5 can now be rewritten using $H_{l_1}(\cdot)$ instead of $\|\cdot\|_1$, yielding:

$$\begin{aligned} \hat{\vec{x}} &= \underset{\vec{x}\in\{\mathcal{N}(\boldsymbol{A})+\vec{x}_0\}}{\arg\min} \; H_{l_1}(\vec{x}) \\ &= \vec{x}_0 + \boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})}\hat{\vec{n}}, \text{ with:} \\ \hat{\vec{n}} &= \underset{\vec{n}\in\mathbb{C}^{n-m}}{\arg\min} H_{l_1}\left(\vec{x}_0 + \boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})}\vec{n}\right) \end{aligned} \quad (10)$$

The minimization in Eq. 10 is not a convex problem. First, the l_1 norm is not differentiable over the full domain, and second, it can be easily shown that our entropy-like function is not convex. Nevertheless, as it is often the case in practice, some simple methods for convex optimization offer very good results solving some types of non-convex problems. In this work we propose using a Nesterov-accelerated *gradient descent* (GD) scheme for solving the optimization in Eq. 10. We adopt the *general scheme* for convergence acceleration proposed in [20] with backtracking line search. The algorithm requires knowledge of the gradient of $H_{l_1}(\vec{x})$ w.r.t. the optimization variable $\vec{n} \in \mathbb{C}^{n-m}$, which was found to be:

$$\frac{\delta H_{l_1}(\vec{x}_0 + \boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})}\vec{n})}{\delta \vec{n}} = \left\{\left[\left(1 - \alpha\left(1 + \frac{\log|x_i|}{\|\vec{x}\|_1}\right)\right)\operatorname{sign}(x_i)\right]_{i=1}^{n}\right\}^* \boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})} \quad (11)$$

where the term that premultiplies $\boldsymbol{E}_{\mathcal{N}(\boldsymbol{A})}$ is an n-dimensional row vector. It is worth noting that equivalence to the pure l_1 minimization problem holds if:

$$-\frac{\log|x_i|}{\|\vec{x}\|_1} = 1 \iff |x_i| = e^{-\|\vec{x}\|_1} \; \forall i \quad (12)$$

We found out that our accelerated GD scheme exhibits good performance in practice, quite independently from α.

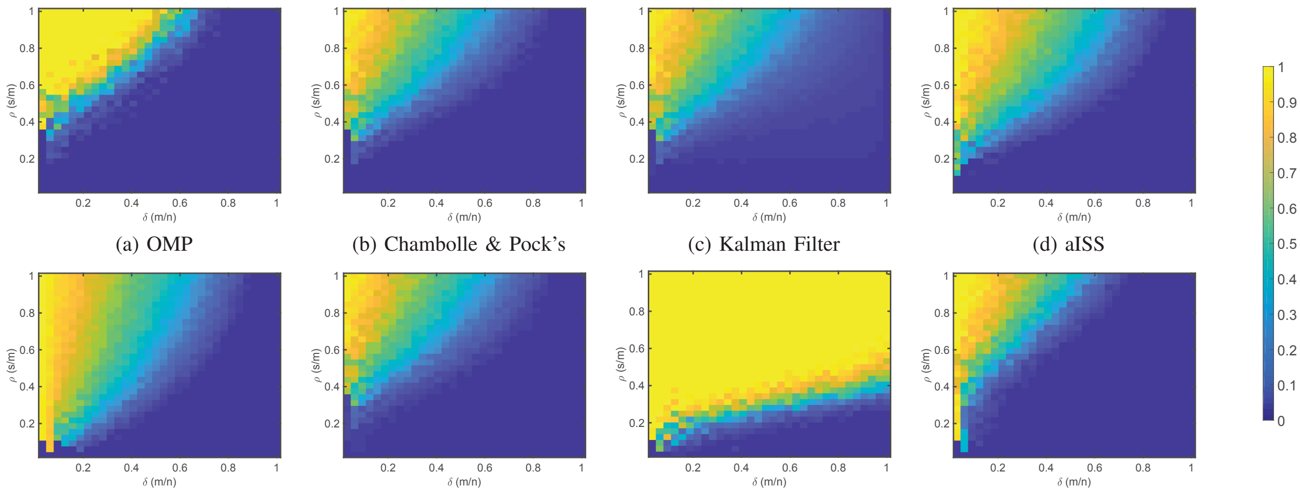
V. Experiments and Results

The proposed null-space-based hybrid l_1 and entropy-minimizing algorithm is compared to the following well-known alternatives: OMP [10], the Chambolle and Pock's (C&P) primal-dual algorithm [8], Loffeld's l_1-minimizing Kalman filter [3], the Adaptive Inverse Scale Space (aISS) algorithm [9], a baseline minimizing the l_1 norm via GD on null-space domain (with and without Nesterov acceleration), and the entropy-based algorithm in [19]. Only Loffeld's l_1-minimizing Kalman filter operates in $\mathcal{N}(\boldsymbol{A})$. We use an improved version of the l_1-minimizing Kalman filter including process noise covariance. Our implementation of the entropy-based algorithm in [19] handles complex signals. The weighting parameter of our algorithm is set to $\alpha = 0.9$.

A series of experiments has been carried out to evaluate the different approaches in terms of sparse recovery performance in the CS sparse recovery problem. We use *best complex antipodal spherical codes* (BCASCs) as close-to-optimal measurement matrices $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ in Eq. 2. Our own fast implementation of the method in [21] was used to construct each $\boldsymbol{A}$. We generate full Donoho-Tanner graphs of normalized l_2 recovery error. Each pixel of the graph means a different combination of the parameters $\delta = m/n$ and $\rho = s/m$. Our graphs are generated with 32×32 experimental cases, using equally-spaced discrete steps per parameter. For each pixel 32 independent experiments are conducted and averaged results are shown. For each experiment a different s-sparse signal $\vec{x} \in \mathbb{C}^n$ is randomly generated. Both the real and imaginary parts of the nonzero complex coefficients are drawn from *i.i.d.* normal distributions of zero mean and unit variance, and the resulting $\vec{x}$ is then l_2-normalized. For all experiments the signal length is set to $n = 128$.

Figs. 1 and 2 show the Donoho-Tanner graphs of normalized l_2 recovery error obtained for each algorithm after 250 (Fig. 1) and 2000 (Fig. 2) iterations. From the results in Figs. 1 and 2 it seems that C&P offers the best performance among algorithms solving an l_1 minimization, while OMP delivers the smallest failure region. In comparison to OMP our method offers superior performance in the failure region (top-left) of the greedy pursuit, but is inferior elsewhere. Both the l_1-minimizing Kalman filter and our Nesterov-accelerated GD baseline converge to exactly the same results as C&P, offering no further improvement. Operation in $\mathcal{N}(\boldsymbol{A})$ and the absence of a thresholding step may be considered advantages over C&P. The aISS method offers a performance that also approaches quite tightly that of C&P, but is inferior for the cases of lowest ratio $\delta = m/n$. The proposed method is the only one that, not being a greedy pursuit, is able to offer a sparse signal recovery performance that is sensibly superior to that of C&P. The failure region appears further confined in the top-left corner and the transition border between success and failure has been pushed towards the left, especially in the upper half of the graph. Rather surprising is the bad performance of the method in [19], which uses a modified entropy functional. Also, the method is between one and two orders

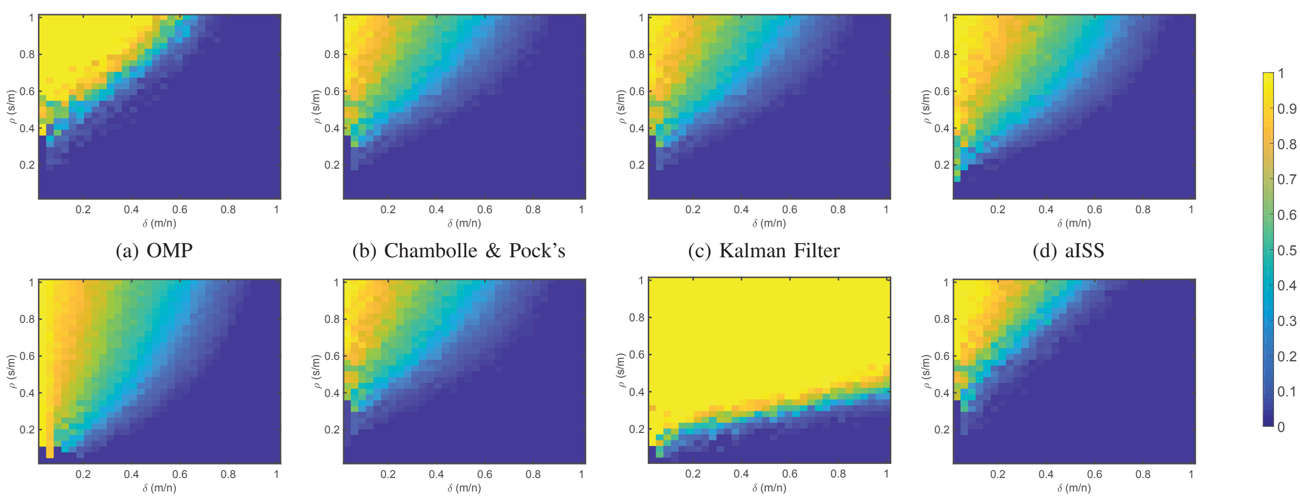
ISBN 978-90-827970-1-5 © EURASIP 2018



(a) OMP (b) Chambolle & Pock's (c) Kalman Filter (d) aISS

(e) GD in null space (f) Nesterov acc. GD (g) Modified Entropy (h) Ours

Fig. 1: Donoho-Tanner graphs of the recovery errors obtained after 250 iterations using OMP (a), the Chambolle and Pock's algorithm (b), the l_1-minimizing Kalman filter (c), aISS (d), l_1-minimization via gradient descent on null-space domain without (e) and with (f) Nesterov acceleration, a modified-entropy minimization (g) and our approach (h).



(a) OMP (b) Chambolle & Pock's (c) Kalman Filter (d) aISS

(e) GD in null space (f) Nesterov acc. GD (g) Modified Entropy (h) Ours

Fig. 2: Donoho-Tanner graphs of the recovery errors obtained after 2000 iterations using the same methods as in Fig. 1.

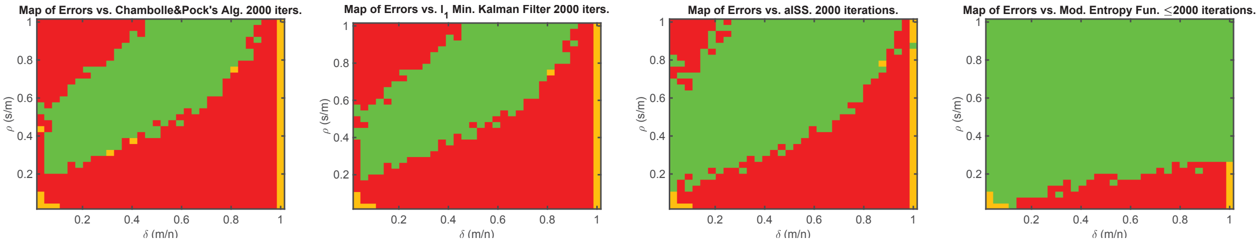
of magnitude slower than the rest. A deeper analysis of this algorithm showed that it is not suited for sparse signals with small nonzero coefficients, like the l_2-normalized signals we use. This is due to the combined effect of the absence of, e. g., an l_1 normalization of the amplitude within the logarithm of the entropy functional and the addition of a constant factor to prevent the argument of the logarithm from reaching zero.

For a better visualization of the improvement we use *semaphoric graphs*, where a triplet of colors is used for the cases of better (green), equivalent (amber), and worse performance (red). The semaphoric graphs obtained against C&P, the Kalman filter, aISS, and the modified entropy method are shown in Fig. 3. Clearly our method brings a massive improvement exactly in the area of the Donoho-Tanner graphs where it is most desired, namely the transition region, further confining the failure area in the top-left corner. Note that the red area in the top-left corner of Fig. 3a-c is not relevant, since it correspond to failure cases in the competitor methods.

VI. Conclusion

In this paper we have proposed solving the CS sparse reconstruction problem by means of minimizing a linear combination of the l_1 norm and an entropy-like function. The minimization problem is formulated in null space domain, so that our algorithm effectively works in a subspace of minimal dimension. Despite the non-convexity of our cost function, we have observed that classical methods for solving convex problems can offer, in practice, a very good approximation to the proposed minimization problem. We use a Nesterov-accelerated gradient descent scheme for our iterative solver.

ISBN 978-90-827970-1-5 © EURASIP 2018



(a) Chambolle & Pock's (b) Kalman Filter (c) aiSS (d) Modified Entropy

Fig. 3: Semaphoric graphs obtained from the Donoho-Tanner graphs in Fig. 2. The graphs show in which cases the proposed method performs better (green), equivalently (amber), or worse (red) than comparable alternatives.

We carried out extensive simulations to evaluate the performance of the proposed approach as sparse recovery algorithm. In the experiments many different synthetically-generated s-sparse n-dimensional complex signals were recovered from $m < n$ measurements. Full Donoho-Tanner graphs of the (average) normalized reconstruction error were generated for the entire range of the parameters $0 < \delta = m/n \leq 1$ and $0 < \rho = s/m \leq 1$. The proposed approach was compared to other seven reference algorithms, namely: OMP, the C&P's primal-dual algorithm, Loffeld's l_1-minimizing Kalman filter, the aISS method, two GD-based baselines (our approach with $\alpha = 0$, with and without Nesterov acceleration), and a method using a modified entropy function. The results show that the proposed method offers a sparse recovery performance that is globally superior to that of all other non-greedy methods and also locally superior to OMP. While the l_1-minimizing Kalman filter and our Nesterov-accelerated GD-based baseline converge to the same results as C&P, our method sensibly shifts the transition in the Donoho-Tanner graphs to the left.

Acknowledgment

The authors would like to thank Prof. M. Möller, Dr. K. Kose, and D. A. Hage for providing the codes of the complex extension of the aISS algorithm, the l_1 minimization method using a modified entropy functional, and an extension of Loffeld's l_1-minimizing Kalman filter accounting for process noise and using optimized parameters, respectively.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008. [Online]. Available: http://dx.doi.org/10.1007/s00365-007-9003-x

[2] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006. [Online]. Available: http://dx.doi.org/10.1002/cpa.20132

[3] O. Loffeld, T. Espeter, and M. Heredia Conde, "From weighted least squares estimation to sparse cs reconstruction: l_1-minimization in the framework of recursive kalman filtering," in *3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, Jun. 2015, pp. 149–153.

[4] O. Loffeld, A. Seel, M. Heredia Conde, and L. Wang, "A nullspace based L1 minimizing kalman filter approach to sparse CS reconstruction," in *11th European Conference on Synthetic Aperture Radar (EUSAR 2016)*, Hamburg, Germany, Jun. 2016.

[5] E. J. Candès, "Compressive sampling," in *Proceedings oh the International Congress of Mathematicians*, Aug. 2006, pp. 1433–1452.

[6] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] E. J. Candès and J. Romberg, "$l1$-magic: Recovery of sparse signals via convex programming," 2005.

[8] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, May 2011. [Online]. Available: http://dx.doi.org/10.1007/s10851-010-0251-1

[9] M. Burger, M. Möller, M. Benning, and S. Osher, "An adaptive inverse scale space method for compressed sensing," *Math. Comput.*, vol. 82, no. 281, pp. 269–299, 2013. [Online]. Available: https://doi.org/10.1090/S0025-5718-2012-02599-3

[10] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, Twenty-Seventh Asilomar Conference on*, Nov. 1993, pp. 40–44 vol.1.

[11] N. Vaswani, "Kalman filtered compressed sensing," in *Proceedings of the International Conference on Image Processing, ICIP 2008*, San Diego, California, USA, Oct. 2008, pp. 893–896. [Online]. Available: https://doi.org/10.1109/ICIP.2008.4711899

[12] ——, "LS-CS-residual (LS-CS): Compressive sensing on least squares residual," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4108–4120, Aug. 2010.

[13] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2405–2409, Apr. 2010.

[14] O. Loffeld, A. Seel, M. Heredia Conde, and L. Wang, "Sparse cs reconstruction by nullspace-based l_1 minimizing kalman filtering," in *11th International Conference on Communications (COMM 2016)*, Bucharest, Romania, Jun. 2016.

[15] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[16] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995. [Online]. Available: http://dx.doi.org/10.1137/S0097539792240406

[17] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[18] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520308000638

[19] K. Kose, O. Gunay, and A. E. Cetin, "Compressive sensing using the modified entropy functional," *Digital Signal Processing*, vol. 24, pp. 63 – 70, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200413002157

[20] Y. Nesterov, "Introductory lectures on convex programming," *Lecture Notes*, 1998.

[21] H. Zörlein and M. Bossert, "Coherence optimization and best complex antipodal spherical codes," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6606–6615, Dec. 2015.