A Comparative Study of Sparse Recovery and Compressed Sensing Algorithms with Application to AoA Estimation

Ahmad Bazzi*†, Dirk T.M. Slock*, Lisa Meilhac†, and Swarnalathaa Panneerselvan*
*EURECOM Mobile Communications Department, 450 route des Chappes, 06410 Biot Sophia Antipolis, France
Email: {bazzi,slock,panneerselvan}@eurecom.fr
†CEVA-RivieraWaves, 400, avenue Roumanille Les Bureaux, Bt 6, 06410 Biot Sophia Antipolis, France
Email:{ahmad.bazzi, lisa.meilhac}@ceva-dsp.com

***Abstract*—We investigate the performance of some sparse recovery and compressed sensing algorithms when applied to the Angle-of-Arrival (AoA) estimation problem. In particular, we review three different approaches in compressed sensing, namely *Pursuit-type*, *Thresholding-type*, and *Bayesian-based* algorithms. The compressed sensing algorithms reviewed herein are of vast interest when applied to AoA estimation problems because of their ability to resolve sources with a single snapshot and without prior knowledge of the number of sources. We compare the performance of these algorithms in terms of Mean-Square Error (MSE) through simulations.**

***Index Terms*—Angle-of-Arrival Estimation, Compressed sensing, Sparse recovery, Pursuit, Thresholding, Bayesian**

I. Introduction

The estimation of the angles of arrival, or AoAs, of multiple sources is a well known problem in the context of array signal processing [1]. In fact, this problem emanates in many engineering applications such as navigation, tracking of objects, radar, sonar, and wireless communications (see [2], [3]). Therefore, many algorithms were used to solve this issue, such as the optimal Maximum Likelihood (ML) technique [4]. Since the ML cost function is highly nonlinear in signal parameters, its direct optimisation would require cumbersome optimisation techniques. Techniques that deal with optimising the ML cost function in a computationally acceptable way are: Iterative Quadratic ML (IQML) [5], Alternating Projections [6], Expectation-Maximisation (EM) [7], Space-Alternating Generalised EM (SAGE) [8], and so forth. However, these algorithms also demand multiple 1D searches, or convergence is in some cases slow or not guaranteed (such as EM). As a result, suboptimal techniques such as MUSIC [9], Root-MUSIC [10], and ESPRIT [11] received more attention. However, all those algorithms require the knowledge of the number of incoming sources. Furthermore, the suboptimal algorithms can not operate with a single snapshot.

Recently, sparse recovery optimisation and compressed sensing algorithms have become popular and found many applications in diverse areas, such as signal processing. Compressed sensing was initiated in 2006 by two ground breaking papers, namely [12] by Donoho and [13] by Candes, Romberg, and Tao. In particular, consider the following linear model, which will be oriented towards the AoA estimation problem in the next section:

$$\mathbf{x} = \mathcal{A}\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathcal{A} \in \mathbb{C}^{\mathrm{N}\times\mathrm{K}}$ is a known overcomplete dictionary. Each column of $\mathcal{A}$ is referred to as an atom. The vector $\mathbf{s} \in \mathbb{C}^{\mathrm{K}\times 1}$ is composed of unknown coefficients that we would like to retrieve using the observed vector $\mathbf{x} \in \mathbb{C}^{\mathrm{N}\times 1}$. In general, this problem is underdetermined and therefore ill-posed. However, a typical remedy for this indeterminacy is to pose a sparse constraint on $\mathbf{s}$, which leads to the following sparse optimisation problem:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{x} - \mathcal{A}\mathbf{s}\|^2 + \lambda\|\mathbf{s}\|_p \tag{2}$$

where $\|\mathbf{s}\|_p$ is the l_p norm of $\mathbf{s}$ and $0 < p \leq 2$. Note that we have excluded $p = 0$ since l_0 is a pseudo-norm (the triangular inequality of norms is not satisfied), which counts the number of non-zero elements. Also note that $\|\mathbf{x}\|_2 = \|\mathbf{x}\|$.

Sparsity is most favored when $p = 0$. However, the above optimisation problem will become NP-hard [14]. As a response to this issue, *Greedy* algorithms have been implemented to solve the above optimisation problem under the l_0 constraint, such as Matching Pursuit (MP) [15] and Orthogonal MP (OMP) [16]. An alternative to the NP-hard problem is to relax the constraint so as the problem is convex, i.e. this happens when $p \geq 1$ [17]. Popular algorithms that are used for the l_1 optimisation problem are the Iterative Shrinkage Thresholding Algorithm (ISTA) [18] and the Basis Pursuit Denoising (BPDN) [19].

Another category of methods are based on the *Bayesian* approach, where the non-zero elements of the sparse signal $\mathbf{s}$ are assumed to have *a priori* statistical information. For instance, the method in [20] imposes a *Laplacian* distribution on the non-zero elements of the sparse signal $\mathbf{s}$. Whereas, the methods found in [21], [41] model the non-zero elements of $\mathbf{s}$ as *Gaussian*. The method in [41] is the Fast Bayesian Matching Pursuit (or FBMP), which was shown to outperform the follwing methods: Stagewise OMP (StOMP) [42], Sparse-Bayes [43], Gradient Projection for Sparse Reconstruction (GPSR) [44], and OMP [16]. More recently, the method in [40], which is called Sparse reconstruction using distribution Agnostic Bayesian Matching Pursuit (or SABMP for short), was shown to outperform the methods in [20], [21], [41]. As argued in [40], the assumption that the non-zero elements of $\mathbf{s}$ follow a certain distribution (whether *Gaussian* or *Laplacian*) is not realistic, because, in most real-world scenarios, it is not *Gaussian*, or it is unknown. Moreover, the SABMP approach is very interesting, since it is naturally agnostic to the distribution of the non-zero elements of the sparse signal. Even more recently, we have introduced an iterative Variational Bayes (VB) algorithm in [45] with the help of latent variables. Indeed, the paper was inspired by the work in [46]–[48]. The papers [46]–[48] focus on introducing latent variables[1] and imposing prior distributions on these variables that favor sparsity. The performance of this VB algorithm demonstrated its potential especially in the case of closely spaced sources.

The rest of this paper is organised as follows: Section II presents the system model used throughout the paper. An overview of different sparse recovery algorithms is given in Section III. In Section IV, we present some simulation results to demonstrate the potential of the algorithms revised in Section III when applied to the Angle-of-Arrival (AoA) estimation problem. We conclude the paper in Section V.

Notations: Upper-case and lower-case boldface letters denote matrices and vectors, respectively. $(.)^{\mathrm{T}}$ and $(.)^{\mathrm{H}}$ represent the transpose and the transpose-conjugate operators. $\odot$ represents the pointwise *Hadamard* product. The k^{th} entry of a vector $\mathbf{x}$ is denoted as $\mathbf{x}_k$. For any $z \in \mathbb{C}$, the magnitude of z is represented as $|z|$. For any vector $\mathbf{x}$, the representation $\mathbf{x}^{(\mathbf{n})}$ denotes the estimated value of $\mathbf{x}$ at an iteration n. Also, for any vector $\mathbf{x} \in \mathbb{C}^{N\times 1}$, the l_p norm ($p > 0$) of $\mathbf{x}$ is denoted as $\|\mathbf{x}\|_p \triangleq \left(\sum_{k=1}^{N} |\mathbf{x}_k|^p\right)^{\frac{1}{p}}$. For convenience, the l_2 norm of a vector $\mathbf{x}$ will be expressed as $\|\mathbf{x}\|$. For any two real numbers x_1 and x_2, the operator $\max(x_1, x_2)$ returns the maximum of the two values. Finally, the indicator function $\mathbb{1}_\Omega$ is equal to 1 if the statement Ω is true, otherwise it returns 0.

II. System Model

Assume a planar arbitrary array of N antennas. Furthermore, consider $q < N$ narrowband sources attacking the array from different angles, i.e. $\theta_1 \dots \theta_q$. A single observed snapshot could be written as [22]

$$\mathbf{x} = \mathbf{A}\mathbf{t} + \mathbf{n} \tag{3}$$

where $\mathbf{x} \in \mathbb{C}^{\mathrm{N}\times 1}$ is a single observed snapshot. The vector $\mathbf{t} \in \mathbb{C}^{\mathrm{q}\times 1}$ is the transmitted signal from q sources. The steering matrix $\mathbf{A} \in \mathbb{C}^{\mathrm{N}\times \mathrm{q}}$ is composed of q steering vectors, i.e. $\mathbf{A} = [\mathbf{a}(\theta_1) \dots \mathbf{a}(\theta_q)]$. Each vector $\mathbf{a}(\theta_i)$ is the response of the array to a source impinging the array from direction θ_i. The form of $\mathbf{a}(\theta_i)$ depends on the array geometry. In general, the vector $\mathbf{a}(\theta)$ has the following functional form

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{-j\frac{w_c}{c}(\bar{x}_1 \sin(\theta) + \bar{y}_1 \cos(\theta))} \\ \vdots \\ e^{-j\frac{w_c}{c}(\bar{x}_N \sin(\theta) + \bar{y}_N \cos(\theta))} \end{bmatrix} \tag{4}$$

where $(\bar{x}_k, \bar{y}_k)$ is the position of the k^{th} antenna. The term $w_c = 2\pi f_c$ is the angular frequency, and c is the speed of light in vacuum. The vector $\mathbf{n} \in \mathbb{C}^{\mathrm{N}\times 1}$ is modelled as a white circular complex Gaussian process of zero mean and covariance $\sigma^2\mathbf{I}_{\mathrm{N}}$ and independent from $\mathbf{s}$.

Now, we recast the problem statement in (3) to the model in equation (1), where matrix $\mathcal{A} \in \mathbb{C}^{\mathrm{N}\times \mathrm{K}}$ is an overcomplete dictionary given as

$$\mathcal{A} = [\mathbf{a}(\theta^1) \dots \mathbf{a}(\theta^{\mathrm{K}})] \tag{5}$$

and $\mathbf{s} \in \mathbb{C}^{\mathrm{K}\times 1}$ is a $q-$sparse (only q elements of $\mathbf{s}$ are not set to zero) vector. Note that the non-zero elements of $\mathbf{s}$ are equal to the corresponding elements of $\mathbf{t}$.

III. Compressive Sensing and Sparse Recovery AoA Estimation

Consider the optimisation problem in penalized form given in equation (2). This problem is referred to as l_p-optimisation. When $p = 0$, note that $\|\mathbf{s}\|_0$ counts the number of non-zero elements of $\mathbf{s}$. Also note that $\|\mathbf{s}\|_0$ is a quasinorm, since the triangular inequality of norms is not satisfied in this case. Solving the problem in (2), when $p = 0$, is known to favor sparse solutions the most. However, this comes with a price of having an NP-hard problem in hand to solve. In this paper, we aim to study the performance of three broad categories of compressed sensing algorithms, namely:

- Pursuit-type algorithms.
- Thresholding-type algorithms.
- *Bayesian*-based algorithms.

A. Pursuit-type algorithms

Pursuit-type algorithms are popular algorithms in the field of compressed sensing. More specifically, matching pursuit algorithms deal with an approximate solution of the l_0-optimisation problem. For uniqueness of the l_0 problem, we refer the reader to [23]. However, basis pursuit relax the l_0-optimisation problem to an l_1-optimisation one. The l_1-optimisation problem is also known as LASSO [25]. For uniqueness of the l_1 problem, we refer the reader to [24]. An advantage of this relaxation is that the problem is now convex. It remains to see when the unique solution provided by the l_1-optimisation problem coincides with that of the l_0 one. The papers in [23], [26] give sufficient conditions for $\hat{\mathbf{s}}$ to be a unique solution of the l_0 and l_1-optimisation problems. Moreover, the necessary conditions for that to happen are found in [27], [28].

[1]Latent variables are often referred to as hyperparameters.

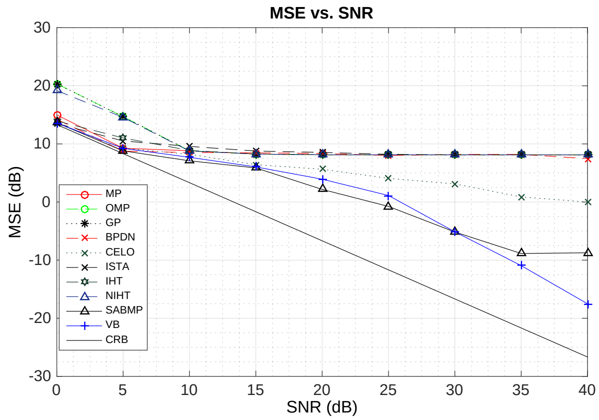

Fig. 1: Two sources impinging the array from directions $\theta_1 = 0°$ and $\theta_2 = 5°$. The number of antennas is 10.

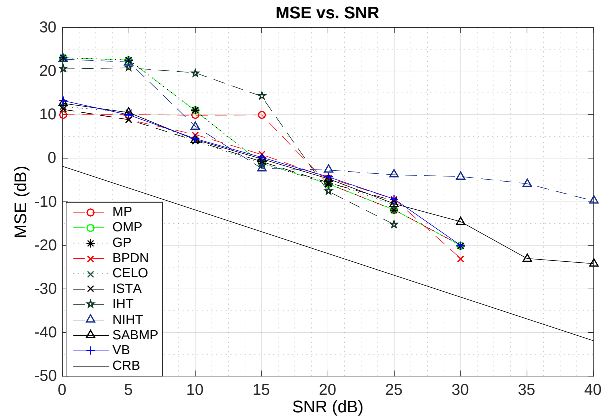

Fig. 2: Two sources impinging the array from directions $\theta_1 = 0°$ and $\theta_2 = 30°$. The number of antennas is 10.

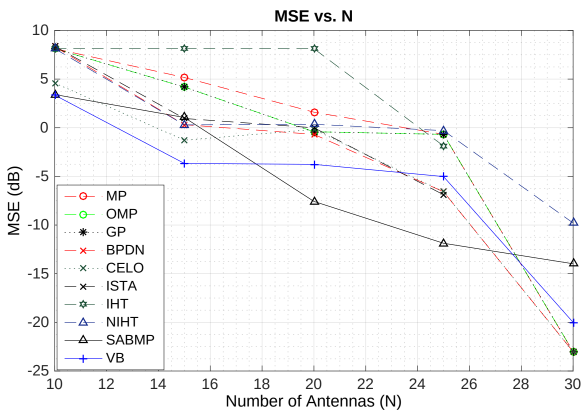

Fig. 3: Two sources impinging the array from directions $\theta_1 = 0°$ and $\theta_2 = 5°$. The SNR is 20dB.

The pursuit algorithms that are evaluated in the context of AoA estimation in this paper are the following:

- Matching Pursuit (MP) [15]
- Orthogonal MP (OMP) [16]
- Gradient, or directional, Pursuit (GP) [29]
- Basis Pursuit De-Noising (BPDN) [19]

The first three algorithms: MP, OMP, and GP are also referred to as *Greedy* algorithms. These algorithms start by initialising $\hat{\mathbf{s}}$ to a zero vector, then estimate a set of non-zero components of $\hat{\mathbf{s}}$ by adding new components to those non-zero terms, in an iterative manner [17]. A brief summary of *Greedy* algorithms is given in **Table 1**. Indeed, the algorithms: MP, OMP, and GP differ in how the "Element Selection" and "Coefficient Updates" are done. For example, MP updates one element at each iteration (this entry corresponds to the maximum magnitude of $\mathbf{g}^{(\mathbf{n})}$). However, OMP updates multiple entries at the same iteration using Least-Square fit. For more information regarding *Greedy* methods, we encourage the reader to refer to [17] and [30]. Furthermore, many work has been done on figuring out a good "Stopping Criterion" for *Greedy* algorithms. For example, in [31], [32], a necessary condition was given in order to recover $\mathbf{s}$ with error threshold

Table 1: General framework of *Greedy* algorithms

INPUT:
Given the data $\mathbf{x}$ and the dictionary $\mathcal{A}$.
INITIALISATION:
$\mathbf{r}^{(\mathbf{0})} = \mathbf{x}$, $\hat{\mathbf{s}}^{(\mathbf{0})} = \mathbf{0}$, and $n = 0$.
MAIN LOOP:
while *Stopping Criterion is not met* **do**

- Element Selection: Select the columns of $\mathcal{A}$ based on the largest magnitude of entries of $\mathbf{g}^{(\mathbf{n})} = \mathcal{A}^{\mathrm{H}}\mathbf{r}^{(\mathbf{n})}$
- Coefficient Update: Obtain a new estimate $\hat{\mathbf{s}}^{(\mathbf{n})}$ that minimises $\|\mathbf{x} - \mathcal{A}\mathbf{s}\|^2$ then increment n.

$\delta = 0$, i.e. when $\|\hat{\mathbf{s}}^{(\mathbf{n})}\| \leq \delta = 0$.

On the other hand, BPDN aims at an l_1-optimisation problem, or equivalently the following

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{subject to} \quad \|\mathbf{x} - \mathcal{A}\mathbf{s}\|^2 \leq \epsilon \tag{6}$$

The regularization parameter ϵ has to be chosen appropriately depending on the noise, which is a major disadvantage of this algorithm.

B. Thresholding-type algorithms

The *Greedy* algorithms are easy and computationally efficient. However, they do not promise recovery of $\mathbf{s}$ as strong as the l_1-optimisation problem. In this sub-section, we are interested in the following:

- Iterative Hard Thresholding (IHT) [33], [34]
- Normalised IHT (NIHT) [35]
- Iterative Shrinkage-Thresholding Algorithm (ISTA) [18]

It was shown in [36] that solutions of (2) are given as follows

$$\mathbf{s} = \mathrm{prox}_{\|\cdot\|_p}\Big(\mathbf{s} - \gamma\mathcal{A}^{\mathrm{H}}(\mathcal{A}\mathbf{s} - \mathbf{x})\Big) \tag{7}$$

where $\gamma > 0$ and the *proximity* function is given by

$$\mathrm{prox}_{\|\cdot\|_p}(\mathbf{z}) = \arg\min_{\mathbf{s}} \left(\|\mathbf{s}\|_p + \frac{1}{2}\|\mathbf{s} - \mathbf{z}\|_2^2\right) \tag{8}$$

which has a unique solution $\mathbf{s}$ for every $\mathbf{z} \in \mathbb{C}^{\mathrm{K}\times 1}$ [37]. Now, equation (7) could be solved using fixed-point in an iterative fashion, viz.

$$\mathbf{s}^{(n+1)} = \mathrm{prox}_{\|\cdot\|_p}\Big(\mathbf{s}^{(n)} - \gamma\mathcal{A}^{\mathrm{H}}(\mathcal{A}\mathbf{s}^{(n)} - \mathbf{x})\Big) \qquad (9)$$

When $p = 0$, the *proximity* in (9) gives the hard threshold, and therefore the IHT algorithm

$$\mathrm{prox}_{\lambda\gamma\|\cdot\|_0}(\mathbf{z}) = [\ldots, \mathbf{z}_i \mathbb{1}_{|\mathbf{z}_i| > \sqrt{2\lambda\gamma}}, \ldots]^{\mathrm{T}} \qquad (10)$$

However, when $p = 1$, the *proximity* in (9) gives the soft threshold. Hence, we obtain the ISTA algorithm

$$\mathrm{prox}_{\lambda\gamma\|\cdot\|_1}(\mathbf{z}) = [\ldots, \frac{\mathbf{z}_i}{|\mathbf{z}_i|}\max(|\mathbf{z}_i| - \lambda\gamma, 0), \ldots]^{\mathrm{T}} \qquad (11)$$

Convergence and recovery properties of IHT are found in [33], [38], [39]. To further enhance IHT, the normalised IHT (NIHT) was obtained by a simple modification [35]. This modification yields a faster algorithm, whilst keeping theoretical performance similar to IHT, in some scenarios.

C. Bayesian-based algorithms

In this sub-section, the sparse signal $\mathbf{s}$ is no longer treated as deterministic, but rather as probabilistic, or random. In other words, a *Bayesian* approach is adopted. Here, we briefly discuss the ideas of:

- Sparse reconstruction using distribution Agnostic Bayesian Matching Pursuit (SABMP) [40]
- Iterative Variational Bayes (VB) with latent variables.

SABMP [40] performs *Bayesian* estimates of the sparse signal $\mathbf{s}$ even when it is modelled as non-Gaussian, thus the term ”Agnostic”. Even more, this method makes use of *a priori* statistics of the noise and the sparsity rate of the signal. More specifically, the signal $\mathbf{s}$ is modelled as $\mathbf{s} = \mathbf{s}_A \odot \mathbf{s}_B$, where $\mathbf{s}_A$ consists of elements that are drawn from some unknown distribution (Agnostic), whereas $\mathbf{s}_B$ are drawn i.i.d. from a *Bernoulli* distribution with success probability p. Note that p controls the sparsity of $\mathbf{s}$, and thus it plays a major role in activating elements[2] of $\mathbf{s}$. The SABMP method was shown, through simulations, to outperform BPDN [19] and Fast *Bayesian* Matching Pursuit (FBMP) [41].

On the other hand, we have recently introduced an iterative Variational Bayes (VB) algorithm in [45] with the help of latent variables. Indeed, the paper was inspired by the work in [46]–[48]. The papers [46]–[48] focus on introducing latent, or hidden, variables and imposing prior distributions on these variables that favor sparsity. In [45], we also introduce the latent variables discussed in [46]–[48], which leads to a novel iterative Variational Bayes [49] algorithm that allows recovering $\mathbf{s}$ from a single observation $\mathbf{x}$ with the help of the latent variables that were introduced.

[2]By activating elements of $\mathbf{s}$, we mean to set these elements to non-zero. Actually, this term was taken from [40].

IV. Simulations

This section presents some simulation results regarding Mean-Square Error (MSE) of the compressed sensing algorithms revised in the previous section. We have simulated three different scenarios. Furthermore, we fix the following simulation parameters: Consider a Uniform Linear Antenna array composed of N antennas spaced at half a wavelength. Furthermore, assume $q = 2$ sources attacking the array from directions $\theta_1 = 0°$ and θ_2. The dictionary $\mathcal{A}$ is composed of $K = 91$ atoms discretized from $-45°$ till $+45°$ with a grid step of $1°$. All our experiments are done using $M = 100$ Monte Carlo trials.

In Scenario 1 (Figure 1), we fix $N = 10$ antennas and $\theta_2 = 5°$. Moreover, we plot the MSE vs. SNR and we notice that all algorithms except for CELO, SABMP, and VB were not able to resolve the closely spaced sources. This phenomenon is explained in [50] and is known as the Restricted Isometry Property (RIP). In short, the RIP condition (in the context of AoA estimation) relates the number of resolvable sources[3] with the number of antennas N that should be used to resolve these sources. Furthermore, we observe that the MSE of SABMP and VB are close to the Cramer-Rao Bound (CRB), whereas CELO has inferior performance when compared to VB or SABMP. In order to validate the RIP condition, we have simulated Scenarios 2 and 3.

In Scenario 2 (Figure 2), we fix $N = 10$ antennas and $\theta_2 = 30°$. One could verify that the RIP condition is now validated for 2 sources when seperated at $30°$. As one can now see, all the algorithms now recover the sparse signal, and thus properly estimate the AoAs at a sufficiently high SNR. For example, IHT presents no error when SNR ≥ 25 dB. Furthermore, MP, OMP, GP, and BPDN present no error when SNR ≥ 30 dB.

In Scenario 3 (Figure 3), we fix the SNR to be 20 dB and $\theta_2 = 5°$. Furthermore, we plot the MSE vs. the number of antennas (N). We notice that all pursuit and thresholding algorithms promise exact recovery of the closely spaced sources when the number of antennas N exceeds a certain level. For instance, ISTA and IHT promise exact recovery at 20 dB of two sources spaced at $5°$ when $N > 25$. As for MP, OMP, GP, and BPDN, the required number of antennas should exceed 30 to guarantee exact recovery.

V. Conclusion

In short, we have revised some methods that fall under three types of compressed sensing algorithms, namely *Pursuit-type*, *Thresholding-type*, and *Bayesian-based* algorithms. Furthermore, we have provided some simulation results demonstrating the performance of these algorithms in the context of angle-of-arrival estimation. Indeed, the algorithms that were revised here, amongst others as well, are appealing due to their ability of resolving sources with a single snapshot and without the knowledge of the number of sources.

[3]By number of resolvable sources, we mean the number of sources that could be resolved, given that the angular seperation between these sources exceed a certain threshold.

ACKNOWLEDGMENT

This work was supported by RivieraWaves, a CEVA company, and a Cifre scholarship. EURECOM's research is partially supported by its industrial members: ORANGE, BMW Group, SFR, ST Microelectronics, Symantec, SAP, Monaco Telecom, iABG, and by the EU projects ADEL (FP7) and HIGHTS (H2020).

REFERENCES

[1] Tuncer, E. and B. Friedlander, "Classical and Modern Direction of Arrival Estimation," Elsevier, Burlington, MA, 2009.
[2] A. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 2440, Jul. 2005.
[3] G. Sun, J. Chen, W. Guo, and K. J. R. Liu, "Signal processing techniques in network-aided positioning," *IEEE Signal Process. Mag.*, vol. 22, pp. 12-23, Jul. 2005.
[4] W. S. Ligget, "Passive sonar: Fitting models to multiple time series," in *NATO ASI on Signal Processing*, J. W. R. Griffiths et al., Eds. New York: Academic, 1973, pp. 327-345.
[5] Bresler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 10811089, Oct. 1986.
[6] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp.15531560, 1988.
[7] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Transactions on Acous., Speech, Signal Processing*, vol. 36, no. 4, pp. 477-489, Apr. 1988.
[8] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 26642677, Oct. 1994.
[9] R. 0 . Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. AP-34, pp. 276-280, 1986.
[10] A. J. Barabell, "Improving the resolution performance of eigenstructure-based direction finding algorithms," in *Proc. IEEE ICASSP*, 1983, pp. 336-339.
[11] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol.37, no. 7, pp. 984-995, July 1989.
[12] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, 52:12891306, 2006.
[13] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal recon- struction from highly incomplete Fourier information," *IEEE Transactions on Information Theory*, 52:489-509, 2006.
[14] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, 24, pp. 227-234,1995.
[15] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, 41:33973415, 1993.
[16] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," In *Proc. of the 27th Asilomar Conference on Signals, Systems and Computers*, 1:40-44, 1993.
[17] M. Mishali and Y. C. Eldar, "Compressed Sensing: Theory and Applications," *Cambridge*, 2012.
[18] I. Daubechies, M. Defrise, and C. de Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, 57 (2004), pp. 1413-1457.
[19] S. Chen, D. Donoho, and M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, 20,pp. 33-61, 1998.
[20] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 5363, Jan. 2010.
[21] E. G. Larsson and Y.Selen, "Linear regression with a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 55, pp. 451460, Feb. 2007.
[22] H. L. Van Trees, "Detection, Estimation, and Modulation Theory," Part IV, Optimum Array Processing. *New York: Wiley*, 2002.
[23] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization," *Proc. Natl. Acad. Sci.* USA 100 (2003), 21972202.
[24] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Am. Math. Soc.*, 22:211231, 2009.
[25] W. J. Fu, "Penalized regressions: the bridge versus the LASSO," *Journal of Computational and Graphical Statistics*, vol. 7, no. 3, Sept. 1998.
[26] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, 48:25582567, 2002.
[27] D. L. Donoho and J. Tanner, "Neighborliness of Randomly-Projected Simplices in High Dimensions," *Proc. Natl. Acad. Sci.* USA, 102:94529457, 2005.
[28] D. L. Donoho and J. Tanner, "Sparse Nonnegative Solutions of Underdetermined Linear Equations by Linear Programming," *Proc. Natl. Acad. Sci.* USA, 102:94469451, 2005.
[29] T. Blumensath and M. Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, 56(6):23702382, 2008.
[30] R. Baraniuk, Compressive sensing, *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118121, Jul. 2007.
[31] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, 50:22312242, 2004.
[32] D.L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, 52:618, 2006.
[33] T. Blumensath and M. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal Appl*, 14(5):629654, 2008.
[34] J. Portilla, "Image restoration through l0 analysis-based sparse optimization in tight frames," *Proc IEEE Int Conf Image*, 2009.
[35] T. Blumensath and M. E. Davies, "Normalized iterated hard thresholding: guaranteed stability and performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298 309, 2010.
[36] P. L. Combettes and V.R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul. 4* , pp. 1168-1200, 2005.
[37] J. J. Moreau, "Proximite et dualite dans un espace hilbertien," *Bull. Soc. Math.*, France 93, 273299, 1965.
[38] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Appl Comput Harmonic Anal*, 27(3):265274, 2009.
[39] R. Garg and R. Khandekar, "Gradient Descent with Sparsification: an iterative algorithm for sparse recovery with restricted isometry property," *Proc Int Conf Machine Learning*, Montreal, Canada; 2009.
[40] M. Masood and T. Y. Al-Naffouri, "Sparse reconstruction using distribution agnostic Bayesian matching pursuit," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 52985309, Nov. 2013.
[41] P. Schniter, L. C. Potter, and J. Ziniel, "Fast bayesian matching pursuit," in *Proc. Workshop Inform. Theory Appl. (ITA)*, 2008.
[42] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, pp. 10941121, Feb. 2012.
[43] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learning Res.*, vol. 1, no. 3, pp. 211244, 2001.
[44] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, pp. 586597, Dec. 2007.
[45] A. Bazzi, D. T.M. Slock, L. Meilhac, "Sparse Recovery using an Iterative Variational Bayes Algorithm and Application to AoA Estimation," submitted, 2016.
[46] D. Wipf, "Bayesian methods for finding sparse representations," Ph. D., Univ. Calif., San Diego, 2006.
[47] D. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Transactions on Information Theory*, 2010.
[48] D. Wipf, J. Palmer, B. Rao, and K. Kreutz-Delgado, "Performance analysis of latent variable models with sparse priors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2007.
[49] M. J. Beal, "Variational algorithms for approximate Bayesian inference," PhD Thesis, University College London, 2003.
[50] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, 51, pp. 4203-4215, 2005.