GRIDLESS COMPRESSIVE SENSING

Ashkan Panahi, Mats Viberg

Depertement of Signals and Systems
Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

The effect of off-grid atoms has become the prominent problem in application of the Compressed Sensing (CS) techniques to the cases where there is an underlying continuous parametrization. In this work, we develop a generalizing CS framework which shows that sampling to a finite grid is not necessary toward compressive estimation. We propose an alternative procedure over infinite dictionaries, which we show to be theoretically consistent in many cases of interest and then propose a robust implementation. We illustrate the general properties of our technique in some difficult practical instances of frequency estimation.

1. INTRODUCTION

Compressed Sensing (CS) has appeared as a general solution to many problems of seemingly different nature, essentially reflecting a common linear structure as well as sparsity [1]. For that, the key target of many pioneering CS studies has been to develop a unified numerically feasible technique, which at the same time enjoys good theoretical properties[2, 3]. Although the advent of the convex optimization techniques, e.g. the interior point [4] and the theoretical machinery of the Restricted Isometry Property (RIP) [5, 2] remarkably enhanced the development [6], recent practical difficulties such as the off-grid atom problem call for further improvement. On the other hand, relying on the RIP condition has led to misunderstand the off-grid atom problem, which also stands for some skeptical views. In this study, we aim to reach the above mentioned original goal of CS by theoretically demonstrating the possibility of nearly optimal atomic decomposition over a continuum without discretization, and further providing a numerically stable implementation.

The general CS approach deals with the so called atomic decomposition problem; Given a set of data and a set of candidate atom vectors, what is the unique smallest set of the atoms to which the data can be linearly decomposed? The different translations of atomic decomposition include numerous practical problems, most of which essentially concern a continuum of atoms [7, 8, 9]. However, the main body of the CS research deals with a finite number of candidates and the continuous cases of interest seem to stay out of the scope of CS. This is traditionally responded by finitely sampling the set of atoms and creating a so called atom grid, which brings up ambiguity when the true atomic decomposition (data representation) includes off-grid atoms [10]. This is recently addressed in [11, 12], with a solution that generally suppresses but does not eliminate the effect.

Regarding the off-grid element problem, one natural solution is to improve the sampling to provide a sufficiently "dense" grid to meet the practical specifications. However, relative to the gained accuracy, the additional computational burden is high and the classical convex methods are frequently reported to run into numerical problems. On the other hand, the RIP-based theory also restricts the grid size, so that it seems straightforward to draw the conclusion that the oversampling strategy fundamentally fails by reaching the RIP limit. However, it is seen that the practical oversampling failure bound is far above the best known resolvable RIP level. This suggests that the RIP may be improved. As we later show, it is in fact possible to include an arbitrary grid size in an alternative theoretical framework, so that the numerical problems may not be related to the theoretical properties of the method.

Finally, we show the possibility of overcoming the numerical instability by proposing an alternative optimization procedure considering an equivalent dual form of the Basis Pursuit method [1], called the noiseless global matched filter. This form surprisingly concerns a finite dimensional convex optimization, while giving continuous estimates. We provide some further both theoretical and empirical results suggesting the robustness and accuracy of the proposed grid-less CS approach.

2. MATHEMATICAL MODELING

In this part, we introduce the problem of atomic decomposition and its solution through BP. Then, we introduce an exact optimality condition for BP which will be used to develop our theory. Consider a set $\mathcal{A} \subset \mathbb{C}^m$ of atoms, which we may also refer to as the dictionary and a vector $\mathbf{x} \in \mathbb{C}^m$. Then it is said that $\mathbf{x}$ can be decomposed to $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathcal{A}$ when there exist amplitudes $s_1, s_2, \ldots, s_n \in \mathbb{C}$ such that

$$\mathbf{x} = \sum_{k=1}^{n} s_k \mathbf{a}_n, \tag{1}$$

where the order n can be arbitrary. Such an atomic decomposition is called ideal if there is no smaller number of atoms

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

n in $\mathcal{A}$ to which $\mathbf{x}$ can be decomposed. Finally, the ideal decomposition to n or fewer atoms is always unique if every subset of $2n$ atoms is linearly independent [13]. This does not necessarily mean that $\mathcal{A}$ is finite. For example, it is well known that the Fourier atoms given by

$$\mathcal{A} = \{[1\ e^{j\phi}\ e^{2j\phi} \dots e^{j(m-1)\phi}]^T \mid 0 \le \phi < 2\pi\} \quad (2)$$

give unique ideal decomposition when $2n \le m$. In this case the ideal decomposition is equivalent to frequency estimation and the condition $2n \le m$ leads to the Nyquist Theorem and its generalization [14]. As seen, the abstract formalism has a great potential in providing general results.

Although in certain special cases the problem of ideal decomposition may be solved by algebraic techniques, e.g. [15], the general case remains difficult, especially when m and n grow. When $\mathcal{A}$ is finite, i.e. $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N\}$, this can be solved by the proximal convex optimization, known as Basis Pursuit (BP) and given by

$$\begin{aligned} &\min_{\mathbf{s}\in\mathbb{C}^N} \sum_{k=1}^{N} |s_k| \\ &\text{s.t.} \\ &\mathbf{x} = \mathbf{A}\mathbf{s} \end{aligned} \quad (3)$$

where $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2 \dots \mathbf{a}_N]$ and $\mathbf{s} = [s_1\ s_2 \dots s_N]^T$. As the solution of BP is shown to include few non-zero elements, it automatically selects the atoms corresponding to non zero elements as the desired decomposition.

Under which conditions does the solution of BP coincide with the ideal atomic decomposition? We are going to explain this in details in the sequel. However, it should first be noted that the question has been also the center of CS studies for a long time, and is partly answered by the following so called RIP condition [16]:

If the finite dictionary $\mathcal{A}$ is such that every choice of $2n$ atoms remains almost isometric (i.e. almost orthogonal), in the case of existing an ideal atomic decomposition to n or less number of atoms, the solution of BP is the ideal decomposition.

To avoid confusion, we omit more mathematical details. Still, it is simple to see that the BP defined in (3) and the above RIP condition are exclusive to the finite dictionary case as otherwise the summation in (3) is not always well defined and $\mathcal{A}$ also necessarily includes arbitrarily coherent atoms violating the RIP. Yet, as we show later, BP can be generalized to include infinite dictionaries with a guaranteed recovery result. For that, we first need to express some fundamental facts about BP.

Let us write down the so called Karush-Kuhn-Tucker (KKT) condition, which characterizes the solution of (3). As it contains equality constraints a vector of dual parameters $\mathbf{z} \in \mathbb{C}^m$ should be considered. Then, the KKT Theorem [17] leads to

$$\mathbf{a}_k^H \mathbf{z} + \xi_k = 0 \quad k = 1, 2, \dots, N \quad (4)$$

where $\xi_k \in \partial|s_k|$ is a subgradient. Note that the subgradients of the absolute function $|s_k|$ are characterized by $\xi_k = s_k/|s_k|$ when $s_k \neq 0$ and $|\xi_k| \le 1$ when $s_k = 0$. Thus, after some simple manipulations the KKT condition in (4) implies that a vector $\mathbf{s} \in \mathbb{C}^N$ is an optimal point of (3) if and only if $\mathbf{x} = \mathbf{A}\mathbf{s}$ and there exists a vector $\mathbf{z} \in \mathbb{C}^m$ such that

$$\forall k \in \{1, 2, \dots, N\}; \quad \begin{cases} |\mathbf{a}_k^H \mathbf{z}| \le 1 & s_k = 0 \\ \mathbf{a}_k^H \mathbf{z} = \gamma_k = \frac{s_k}{|s_k|} & s_k \neq 0 \end{cases} \quad (5)$$

For example, suppose that the vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ belonging to a finite dictionary $\mathcal{A}$ ideally decompose a vector $\mathbf{x}$ as in (1). Remembering that the unemployed atoms are accompanied by zero amplitudes $s = 0$ in BP, from (5) we conclude the following fact.

For BP to recover an ideal decomposition as in (1), there must exist a vector $\mathbf{z}$ such that $\mathbf{a}_k^H \mathbf{z} = s_k/|s_k|$ for $k = 1, 2, \dots, n$ and $|\mathbf{a}^H \mathbf{z}| \le 1$ for any other unemployed atom $\mathbf{a} \in \mathcal{A}$. We will refer to this as the Dual Null Space Property (DNSP).

Note that as $\mathbf{a}_k^H \mathbf{z} = s_k/|s_k|$ implies $|\mathbf{a}_k^H \mathbf{z}| \le 1$, the condition $|\mathbf{a}^H \mathbf{z}| \le 1$ can be equivalently assumed for all atoms. How does DNSP result in RIP? As seen, the possibility of recovering an ideal decomposition depends on finding a proper dual vector $\mathbf{z}$ satisfying the DNSP. Then, restricting $\mathbf{z}$ to certain families results in a tighter recovery bound. In [5] for example, it is shown that if RIP holds DNSP is satisfied in every possible sparse case by a simple selection procedure of $\mathbf{z}$. Thus, RIP implies perfect recovery. However, in certain individual cases, it is possible to select $\mathbf{z}$ differently, resulting in perfect recovery beyond the prediction of RIP.

3. NOISELESS GLOBAL MATCHED FILTER

The results in the previous section are valid only for the finite dictionary case, but it is seen that the DNSP is also well defined and can be verified in the case of infinite dictionaries. However, the connection to BP is established only when $\mathcal{A}$ is finite. On the other hand, the interesting properties of BP are equivalently dedicated to the DNSP. Thus, providing a procedure over infinite dictionaries that terminates only at the points satisfying DNSP will retrieve the desired properties in the cases of interest. There is an optimization expression over infinite dictionaries which results in DNSP as the optimality condition as follows:

$$\begin{aligned} &\max_{\mathbf{z}\in\mathbb{C}^m} \Re(\mathbf{z}^H \mathbf{x}) \\ &\text{s.t.} \\ &\forall \mathbf{a} \in \mathcal{A}; \quad |\mathbf{a}^H \mathbf{z}| \le 1 \end{aligned} \quad (6)$$

To respect and remind the important and well connected work in [18], we will refer to (6) as the noiseless Global Matched Filter (NL-GMF).

It is easy to conclude DNSP for the above NL-GMF scheme. Note that there exists an infinite number of inequality constraints $|\mathbf{a}^H\mathbf{z}| \leq 1$, each identified by an atom $\mathbf{a} \in \mathcal{A}$. Assume that for the optimal $\mathbf{z}$ vector the bound constraints, i.e. the ones holding equality instead of inequality, are given by $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. Then, the KKT condition implies that there exist corresponding positive real numbers $r_1, r_2, \ldots, r_n$ such that

$$\mathbf{x} = \sum_{k=1}^{n} r_k \frac{\mathbf{a}_k^H \mathbf{z}}{|\mathbf{a}_k^H \mathbf{z}|} \mathbf{a}_k \tag{7}$$

Defining $s_k = r_k \mathbf{a}_k^H \mathbf{z}/|\mathbf{a}_k^H \mathbf{z}|$ it is seen that the resulting vector $\mathbf{z}$ satisfies DNSP as $s_k/|s_k| = \mathbf{a}_k^H \mathbf{z}/|\mathbf{a}_k^H \mathbf{z}| = \mathbf{a}_k^H \mathbf{z}$. Also, note that if DNSP is satisfied for a given case with an infinite dictionary $\mathcal{A}$, the same parameters satisfy it for a sub-dictionary $\mathcal{B} \subset \mathcal{A}$ if the former resulting decomposition atoms are included in $\mathcal{B}$. In other words, we have the following:

For any vector $\mathbf{x}$*, the solution of NL-GMF coincides with the solution of BP over a finite grid if the former NL-GMF atom estimates are on-grid.*

How can (6) be implemented without discretization? To answer this question we propose the following local search algorithm which we show to converge to the NL-GMF global optimal point characterized by DNSP. For simplicity, we assume that the atoms are indexed by a continuous real parameter ω, i.e. $\mathcal{A} = \{\mathbf{a}(\omega) \mid 0 \leq \omega < 1\}$, where $\mathbf{a}(\omega)$ is a differentiable function. We note in passing that the constraint (6) resembles an equi-ripple filter design, a fact that can be used to inspire a search algorithm.

3.1. Implementing GMF

In simple words, each $\mathbf{z}$ updating iteration consists of finding the atoms violating the constraint in (6) and finding an infinitesimal update which increases the cost $\Re(\mathbf{z}^H\mathbf{x})$, while suppressing the spectrum $p(\omega) = |\mathbf{a}^H(\omega)\mathbf{z}|$ at the violation points. Note that the set of violation points, recognized by $|\mathbf{a}^H(\omega)\mathbf{z}| > 1$, generally consists of a number of intervals. However, as each interval contains a finite number of peak points, it suffices to suppress the spectrum at the violating peak points. The following is an algorithm sketch:

- Initialize by an arbitrary $\mathbf{z}$ vector and iteratively follow the following procedure:
- Recognize the peak points (local maxima) in the spectrum $p(\omega) = |\mathbf{a}^H(\omega)\mathbf{z}|$.
- Select the peak points $\omega_1, \omega_2, \ldots$ at which the spectrum $p(\omega)$ is greater than 1.
- Find a feasible ascend direction $\boldsymbol{\delta}$, such that updating $\mathbf{z}$ to $\mathbf{z} + \epsilon\boldsymbol{\delta}$ for a sufficiently small value of ϵ decreases $p(\omega_k)$ and increases $\Re(\mathbf{z}^H\mathbf{x})$.

we subsequently explain each step in details.

3.1.1. Finding all peaks

As the spectrum $p(\omega)$ contains multiple local maxima, we may start by searching over a sufficiently dense grid $\Omega = \{\omega_1^g, \omega_2^g, \ldots, \omega_N^g\}$ and later update the estimates by local search. We employ a simple but numerically stable iterative bisection scheme with the following iteration:

- Compute $p(\omega_1^g), p(\omega_2^g), \ldots, p(\omega_N^g)$ and select peak points $\omega_{i_1}^g, \omega_{i_2}^g, \ldots, \omega_{i_r}^g$ obeying $p(\omega_{i_k}^g) \geq p(\omega_{i_k+1}^g)$ and $p(\omega_{i_k}^g) \geq p(\omega_{i_k-1}^g)$. Set the estimate upper and lower bounds $\omega_k^h = \omega_{i_k+1}^g$ and $\omega_k^l = \omega_{i_k-1}^g$ respectively.
- Set the estimates $\omega_k^c = (\omega_k^h + \omega_k^l)/2$.
- Compute the sign of the gradient at every ω_k^c given by the sign of $\Re(\mathbf{a}^H(\omega_k^c)\mathbf{z}\mathbf{z}^H d\mathbf{a}(\omega_k^c)/d\omega)$.
- For the points ω_k with a positive gradient set $\omega_k^l = \omega_k^c$. For the others, take $\omega_k^h = \omega_k^c$.

3.1.2. Finding a feasible ascend direction

Once the peak points $\omega_1^c, \omega_2^c, \ldots$ are identified by the above procedure, its subset $\omega_1, \omega_2, \ldots, \omega_n$ of all the violating peaks, given by $p(\omega_k) > 1$ is recognized. Then, we look for a direction $\boldsymbol{\delta}$ such that updating to $\mathbf{z} + \epsilon\boldsymbol{\delta}$ for a sufficiently small value of ϵ decreases $p(\omega_k)$ and increases $\Re(\mathbf{z}^H\mathbf{x})$. This leads to

$$\forall 1 \leq k \leq n \quad \Delta p(\omega_k) < 0 \rightarrow \Re(\mathbf{z}^H \mathbf{a}(\omega_k)\mathbf{a}^H(\omega_k)\boldsymbol{\delta}) < 0 \tag{8}$$

Now we bring the above linear constraints into a standard real-valued form. Defining $\mathbf{b}_k = \left(\mathbf{a}^H(\omega_k)\mathbf{z}\right)\mathbf{a}(\omega_k)$ and decomposing to real and imaginary parts as $\mathbf{b}_k = \mathbf{b}_k^r + j\mathbf{b}_k^m$ and $\boldsymbol{\delta}_k = \boldsymbol{\delta}^r + j\boldsymbol{\delta}^m$, we may re-write (8) as

$$\forall 1 \leq k \leq n \quad (\mathbf{b}_k^r)^T\boldsymbol{\delta}^r + (\mathbf{b}_k^m)^T\boldsymbol{\delta}^m < 0 \tag{9}$$

which may be denoted by $\mathbf{W}^T\mathbf{d} \prec 0$ where $\prec$ denotes elementwise inequality, $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_n]$ is the collection of $\mathbf{w}_k = \left[(\mathbf{b}_k^r)^T \quad (\mathbf{b}_k^m)^T\right]^T$ and $\mathbf{d} = \left[(\boldsymbol{\delta}^r)^T \quad (\boldsymbol{\delta}^m)^T\right]^T$. In other words, there exist an all-negative vector $\mathbf{r}$ such that

$$\mathbf{W}^T\mathbf{d} = \mathbf{r} \prec \mathbf{0} \tag{10}$$

This is called the feasibility cone. In a similar manner, the condition that the NL-GMF cost should be increased can also be written in a standard form as

$$\Delta\Re(\mathbf{z}^H\mathbf{x}) > 0 \rightarrow \Re(\boldsymbol{\delta}^H\mathbf{x}) > 0 \rightarrow \mathbf{y}^T\mathbf{d} > 0 \tag{11}$$

where denoting $\mathbf{x} = \mathbf{x}^r + j\mathbf{x}^m$, we define $\mathbf{y} = [(\mathbf{x}^r)^T\ (\mathbf{x}^m)^T]^T$.

Finding a vector $\mathbf{d}$ satisfying (10) and (11) is a well-studied problem which can be solved by classical iterative

techniques. However, to reduce complexity, we avoid extra iterations by introducing the following technique. The general solution to (10) is obtained as

$$\mathbf{d} = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{r} + \mathbf{n} \tag{12}$$

where $\mathbf{n}$ is arbitrary vector belonging to the null space of $\mathbf{W}^T$, i.e. $\mathbf{W}^T\mathbf{n} = 0$. From, (12), the condition of (11) simplifies to

$$\mathbf{y}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{r} + \mathbf{y}^T\mathbf{n} > 0 \tag{13}$$

To ensure the above, we select $\mathbf{n}$ and $\mathbf{r}$ so that each term remains positive. In order to increase coherence $\mathbf{y}^T\mathbf{n}$, The vector $\mathbf{n}$ is simply chosen to be the orthogonal projection of $\mathbf{y}$ into the orthogonal complement of the range space of $\mathbf{W}$,

$$\mathbf{n} = \mathbf{y} - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{y} \tag{14}$$

To select $\mathbf{r}$, note that the term $\mathbf{y}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{r}$ can be expanded as $\sum_i r_i\eta_i$ where η_i and r_i are the i^{th} element of the vectors $(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{y}$ and $\mathbf{r}$ respectively. Thus, the condition is guaranteed if

$$r_i = \begin{cases} 0 & \eta_i \geq 0 \\ \eta_i & \eta_i < 0 \end{cases} \tag{15}$$

In summary once the violating peaks are found, the ascend direction $\boldsymbol{\delta}$ is computed by the following method

- Compute matrices $\mathbf{W}$ and $\mathbf{y}$ from the atoms $\mathbf{a}(\omega_k)$ and the data $\mathbf{x}$ respectively.
- Compute $\mathbf{n}$ and $\mathbf{r}$ from (14) and (15) respectively.
- Compute $\mathbf{d}$ and $\boldsymbol{\delta}$ from (12).

Each iteration is completed by updating $\mathbf{z}$ to $\mathbf{z} + \epsilon\boldsymbol{\delta}$ for a small step size ϵ. Thus, the algorithm stops only when $\mathbf{d} = 0$, or more practically when it is small enough. Then, from the construction, $\mathbf{y}$ lies in the range space of $\mathbf{W}$ (as $\mathbf{n} = 0$) and the spanning components η_i are all negative. Thus, (13) may not have any solution and there does not exist any feasible ascend direction. From Farka's lemma [17], we conclude that the DNSP holds. This means that even though suboptimal in convergence speed, the proposed algorithm can only converge to the solution of the NL-GMF.

4. RESULTS

Due to lack of space, we only include some preliminary results of implementing NL-GMF for an example case of frequency estimation, given in (2). In all cases the step size $\epsilon = .001$ and 1000 iterations are considered. In the first scenario, we took $m = 4$, which gives a highly correlated atom dictionary, and chose two components at $\omega = 0, 1/8$ in terms of the normalized frequency. As seen in Figure 1, the BP implementation by the CVX toolbox can not totally converge

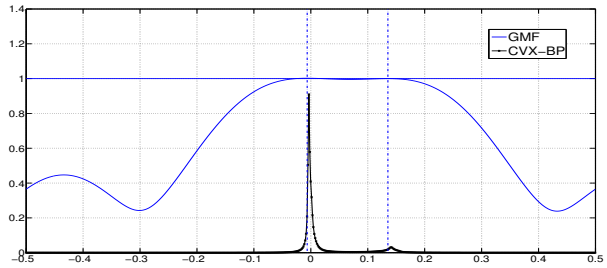

Fig. 1. The frequency spectrum for the NL-GMF compared to BP.

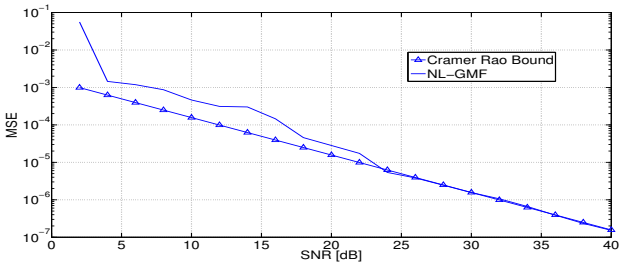

Fig. 2. The NL-GMF mean squared error in different SNRs.

and stops at a close non-sparse point. However, NL-GMF gives peaks exactly at the true points despite high correlation. In a different scenario we considered a noisy observation of two deterministic components $s_1 = s_2 = 1$ with $m = 15$ samples at the same frequencies as the previous experiment. Figure 2 shows the Mean Square Error (MSE) for 100 trials at each SNR level. As seen, when SNR decreases to 25 dB, NL-GMF is not statistically efficient anymore, as some extra peaks appear in the spectrum. We leave this discussion to future work.

5. CONCLUSION

In this work, we proposed a CS recovery technique over infinite and continuous dictionaries, which we refer to as the noiseless global matched filter. We reviewed its theoretical properties and its deep relation to BP, which suggests that NL-GMF may perfectly recover the ideal atomic decomposition in many cases of interest, modifying RIP to the DNSP. We then proposed a NL-GMF numerical implementation which we showed to stably converge. We finally demonstrated the properties of NL-GMF through simulation. The results showed that NL-GMF may converge and provide satisfactory results even when the original BP technique with a convex optimization is numerically deficient. Furthermore, the results showed that NL-GMF may be reliably used for noisy recovery at a high enough SNR. However, it fails when SNR decreases as some false peaks appear in the spectrum. This can be corrected by modifying the NL-GMF method, which is postponed to future studies.

6. REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Dec. 1998.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289 –1306, Apr. 2006.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., Series B (Methodological)*, vol. 58, pp. 267–288, Jan. 1996.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[5] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Mag.*, vol. 24, pp. 118–121, July 2007.

[7] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *IEEE J. Select. Topics Signal Processing*, vol. 4, pp. 375–391, Apr. 2010.

[8] H. Yao, P. Gerstoft, P. M. Shearer, and C. Mecklenbräuker, "Compressive sensing of the tohoku-oki mw 9.0 earthquake: Frequency-dependent rupture modes," *Geophys. Res. Lett.*, vol. 38, Oct. 2011.

[9] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Resonance Med. Mag.*, vol. 58, pp. 1182–1195, Dec. 2007.

[10] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2182–2195, 2011.

[11] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatch," *arXiv preprint arXiv:1309.0858*, 2013.

[12] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4735–4744, 2011.

[13] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[14] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *Signal Processing, IEEE Transactions on*, vol. 57, no. 3, pp. 993–1009, 2009.

[15] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 5, pp. 814–824, 1990.

[16] E. J. Candes, 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.

[17] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. Wiley, 2006.

[18] J. J. Fuchs, "On the application of the global matched filter to doa estimation with uniform circular arrays," *IEEE Trans. Signal Processing*, vol. 49, no. 4, pp. 702–709, Apr. 2001.