CHOOSING ANALYSIS OR SYNTHESIS RECOVERY FOR SPARSE RECONSTRUCTION

Nicolae Cleju * *Maria G. Jafari*† *Mark D. Plumbley*†

* Technical University "Gheorghe Asachi" of Iasi, Romania
Faculty of Electronics, Telecommunications and Information Technology
† Queen Mary University of London, Centre for Digital Music, London, UK

ABSTRACT

The analysis sparsity model is a recently introduced alternative to the standard synthesis sparsity model frequently used in signal processing. However, the exact conditions when analysis-based recovery is better than synthesis recovery are still not known. This paper constitutes an initial investigation into determining when one model is better than the other, under similar conditions. We perform separate analysis and synthesis recovery on a large number of randomly generated signals that are simultaneously sparse in both models and we compare the average reconstruction errors with both recovery methods. The results show that analysis-based recovery is the better option for a large number of signals, but it is less robust with signals that are only approximately sparse or when fewer measurements are available.

Index Terms— Analysis sparsity, synthesis sparsity, comparison, sparse reconstruction, signal recovery

1. INTRODUCTION

In recent years, the study of sparsity and its benefits has been an active research field in signal processing. The *de facto* sparsity model used in literature is a generative model defined as

$$x = D\gamma_S, \text{ with } \|\gamma_S\|_0 = k. \tag{1}$$

It requites that the signal $x \in \mathbb{R}^d$ be expressed as a weighted sum of at most k of the N atoms from a dictionary $D \in \mathbb{R}^{d\times N}$. The signal x is said to be k-*sparse*.

More recently, a new sparsity model known as *analysis sparsity* was introduced [1], describing instead what the signal is orthogonal to:

$$\gamma_A = \Omega x, \text{ with } \|\gamma_A\|_0 = N - l \tag{2}$$

with $\Omega \in \mathbb{R}^{N\times d}$ being an analysis operator such that x is orthogonal to at least l of its rows. Following [2], we call the quantity l the *cosparsity* of x, and x is said to be l-*cosparse*.

Both the synthesis and the analysis sparsity models can be used as regularizing terms in various ill-posed problems. A widely studied problem of this kind is that of recovering a signal that is acquired only through a set of m linear measurements arranged as the rows of a matrix $M \in \mathbb{R}^{m\times d}$, contaminated with noise z of energy $\epsilon = \|z\|_2^2$:

$$y = Mx + z. \tag{3}$$

It has long been established [3] that a sufficiently synthesis-sparse signal x can be recovered as long as the projections are incoherent with the sparsity dictionary by solving the NP-complete optimization problem

$$\hat{x} = D \arg\min_{\gamma_S} \|\gamma_S\|_0 \text{ with } \|y - MD\gamma_S\|_2^2 < \epsilon. \tag{4}$$

Recovering a signal with (4) is known as the compressed sensing problem [4], and it has been applied in a variety of applications in the recent years (e.g. [5, 6]).

Similarly to (4), it has recently been proved [2] that a sufficiently analysis cosparse signal x can also be recovered by solving

$$\hat{x} = \arg\min_{x} \|\Omega x\|_0 \text{ with } \|y - Mx\|_2^2 < \epsilon. \tag{5}$$

The optimization problem (5) is believed to be as difficult as (4) [7]. The use of ℓ_0 norm in (4) and (5) leads to instability in the presence of noise, or with signals which are only approximately sparse. Increased robustness can be achieved by relaxing the ℓ_0 norm to $\ell_p, 0 < p \leq 1$, at the expense of an increased number of measurements. A popular choice is the ℓ_1 norm [8, 9], which has the added benefit of turning (4) and (5) into convex optimization problems.

Although some key similarities and differences between the two sparsity models and their associated recovery problems are known, it is still unclear under what conditions one performs better than the other. In this paper we conduct an experimental investigation into the practical task of deciding which sparsity model and recovery problem is better suited in a given scenario.

This work was supported by: EURODOC "Doctoral Scholarships for research performance at European level" project, financed by the European Social Found and Romanian Government (NC), EU Framework 7 FET-Open project FP7-ICT-225913-SMALL: Sparse Models, Algorithms and Learning for Large-Scale data (MGJ) and a Leadership Fellowship (EP/G007144/1) from the UK Engineering and Physical Sciences Research Council (EPSRC) (MDP)

The rest of the paper is organized as follows. In section 2 we present the analysis model in more detail and we rewrite it into a more convenient form, revealing the similarity with the synthesis model as well as the key differences. Section 3 introduces signals that are jointly sparse and cosparse (i.e. *bisparse*), establishing the conditions for their existence. In section 4 we present the results of the comparison of synthesis and analysis recovery for a large number of bisparse signals . Final comments and concluding remarks are in section 5.

2. ANALYSIS AS LEAST-SQUARES SYNTHESIS SPARSITY

The relation between the synthesis and the analysis sparsity models has previously been analysed in detail in [1, 2]. In [1] it is proved than the synthesis and analysis models (1) and (2) and their associated recovery problems (4) and (5) are equivalent whenever the dictionary D is a basis, if $\Omega = D^\dagger$ and $l = d - k$, where $D^\dagger$ designates the pseudoinverse of D. The solutions of the synthesis and analysis recovery diverge, however, as D becomes an overcomplete dictionary. This is caused by the geometry of the sets of synthesis- and analysis-sparse signals being different for the overcomplete case, with the differences increasing as the overcompleteness factor increase [2].

The operator Ω is usually assumed to be in *general position* [2], i.e. any set of up to d rows is linear independent, which is an assumption that we follow in this paper as well. In this case $l \leq d-1$ since a non-zero d-dimensional signal can be orthogonal to at most $d-1$ linear independent signals.

Following our recent work [10], we reformulate the analysis sparsity model (2) into an equivalent augmented synthesis model. One observes that γ_A in (2) is the least-squares solution of (1) if $D = \Omega^\dagger$:

$$x = \Omega^\dagger \gamma_A = D\gamma_A \tag{6}$$

which can be written in the form

$$\begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} D \\ P_D \end{bmatrix} \gamma_A. \tag{7}$$

where P_D is any basis for the nullspace of D. The extra orthogonality constraint ensures that γ_A is the least-squares solution of x in D. Here D is the pseudoinverse of the analysis operator Ω, and it is regarded as an overcomplete dictionary. As the upper part of (7) is similar to the synthesis model (1), (6) can be considered an augmented synthesis sparsity model, with the addition of an extra constraint.

The equation system (7) reveals the analysis sparsity model to be a *least-squares constrained synthesis* sparsity model. In [10] we rigorously prove that for the noiseless case ($\epsilon = 0$) the standard analysis recovery problem (5) is identical to the augmented synthesis problem

$$\hat{x} = D \arg\min_{\gamma} \|\gamma\|_0 \text{ with } \begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} MD \\ P_D \end{bmatrix} \gamma \tag{8}$$

with $D = \Omega^\dagger$. The additional orthogonality constraint is the defining feature of the least-squares solution. It can easily be enforced with many existing synthesis recovery algorithms, enabling them to also be used for analysis recovery with (8). In [10] we have shown the practical viability of this approach through simulations with various synthesis-based solvers. Throughout the rest of this paper, when referring to "analysis recovery" we will have this approach in mind.

In this paper we compare synthesis recovery (4) with $\epsilon = 0$ with the reformulated analysis recovery (8), aiming to establish when one is better than the other, for a general dictionary D. One observes that the two problems now have a very similar form. The difference is that synthesis recovery (4) seeks the *sparsest* decomposition γ_S of x in D, whereas analysis recovery (8) requires finding the *least-squares* decomposition γ_A of x in D. As such, analysis recovery (8) benefits from having available an additional orthogonality constraint, at the expense of seeking a solution which is less sparse, since the sparsity of γ_A is upper bounded by $l \leq d-1$ while no such restriction exists for γ_S.

Choosing between synthesis and analysis-based recovery implies, therefore, evaluating the benefits of having an extra orthogonality constraint versus higher sparsity of the solution. In this paper we evaluate this trade-off from a practical point of view, using numerical simulations to establish when one recovery problem is more successful than the other, depending on the sparsity and cosparsity of a signal. This kind of empirical numerical experiments have been presented in compressed sensing literature before [11], providing insight into the theory as well as significant help for practical applications.

A meaningful comparison requires the synthesis dictionary and the analysis operator to be pseudoinverses of each other, i.e. using the same D in (4) and (8). This also follows naturally from the equivalence of the two models in the complete case. In a practical application, however, if different dictionary and analysis operator are available, one must keep in mind that the difference of their inherent quality will correspondingly make one recovery method preferable.

3. BISPARSE SIGNALS

In this paper we consider signals that are jointly synthesis and analysis sparse for some random dictionary, and we perform reconstructions with both synthesis and analysis algorithms, separately, in order to establish which recovery is better, depending on the sparsity and cosparsity of the signals. We focus mainly on exact-sparse signals, but we optionally add small non-sparse random components to the signals to also investigate robustness against non-exact sparsity.

Let us consider a signal $x \in \mathbb{R}^d$ and a dictionary $D \in \mathbb{R}^{d\times N}$. Denote with γ_S the sparsest decomposition of x in D, and with γ_A the least-squares decomposition of x in D, with $k = \|\gamma_S\|_0$ (the number of non-zero coefficients in γ_S) and $l = N - \|\gamma_A\|_0$ (the number of zero coefficients in γ_A). We refer to such a signal as a (k, l)-*bisparse* signal, or simply a *bisparse* signal, throughout the rest of this paper.

We point out that we are interested in bisparse signals not for their practical relevance in applications, but as a way to compare the usefulness of cosparsity against sparsity for the same signal. We seek to investigate the following question: given k and l, is it better to recover the signal using synthesis-based recovery (4) or using analysis-based recovery (5) reformulated as (8)? As $k \in 1, 2, ...d$ and $l \in 0, 1, ...(d-1)$, there are d^2 possible pairs (k, l) in all. We test every combination and determine in which region of the (k, l) space is synthesis recovery performing better than its analysis counterpart, and vice-versa.

To generate (k, l)-bisparse signals for some dictionary D, we must first see when is it possible for such signals to exist. We rely on Theorem 3.1 that formulates the existence conditions.

Theorem 3.1. *Consider an overcomplete dictionary $D \in \mathbb{R}^{d \times N}$ and denote $M = D^{\dagger} D$. Given any subsets $I, J \subset \{1, 2, ...N\}$ with $card(I) = l$ and $card(J) = k$, there exists a non-zero vector $x \in \mathbb{R}^d$ having simultaneously a decomposition γ_S with the non-zero coefficients on locations in J and a least-squares decomposition vector γ_A with zero elements on locations I if and only if the rank of the $l \times k$ minor matrix M_{IJ} obtained by keeping only the rows with indices I and columns J from M is strictly smaller than k.*

Proof. The signal x satisfies both (1) and (2) with $\Omega = D^{\dagger}$. Replacing x from (2) with (1) yields:

$$\gamma_A = \underbrace{D^{\dagger} D}_{M} \gamma_S, \text{ with } \|\gamma_A\|_0 = N - l \text{ and } \|\gamma_S\|_0 = k. \quad (9)$$

If there exist γ_S and γ_A obeying (9), then, if we keep only the rows in I and the columns in J from M in (9), we have

$$0 = M_{IJ} \gamma_{S_J} \quad (10)$$

where M_{IJ} is a minor matrix obtained by keeping only the rows with indices I and the column with indices J from M, and γ_{S_J} is the restriction of γ_S to the indices J. Since γ_{S_J} is nonzero, this means that the k columns of the M_{IJ} are linear dependent, i.e. $rank(M_{IJ}) < k$.

Conversely, if an $l \times k$ minor matrix M_{IJ} of M has rank smaller than k, it means that its k columns are linear dependent, and therefore there exists a set of non-zero coefficients γ_{S_J} such as (10) is true. One has only to find such a solution of (10) and then place the coefficients on the locations J of γ_S. The signal x can be then generated as $x = D\gamma_S$. The least-squares solution $\gamma_A = D^{\dagger} x$ will have zeros at the locations in I.

When generating bisparse signals this way, it is possible for γ_A to accidentally have additional zero coefficients besides the locations in I. Thus, an $l \times k$ matrix M_{IJ} with rank smaller than k only guarantees that cosparsity l' of γ_A is at least l, but not necessarily equal to it, $l' \geq l$. □

Theorem 3.1 shows that it is always possible to find k-sparse signals that are also l-cosparse up to $l \leq k - 1$, irrespective of where the non-zero coefficients of γ_S and the zeros of γ_A are located, since $l < k$ implies that $rank(M_{IJ}) < k, \forall I, J$.

For $l \geq k$, however, there exist (k, l)-bisparse signals only if the sparsity and co-sparsity patterns I and J happen to correspond to a rank deficient minor matrix M_{IJ} of $D^{\dagger} D$. Thus, such signals are not guaranteed to exist. Their existence is strictly determined by the distribution of linear dependent minors in the $D^{\dagger} D$ matrix. Our simulations show that the second case is negligible for a reasonable high overcompleteness factor (about $N/d > 1.5$), i.e. the vast majority of bisparse signals have $l < k$. An in-depth characterization of this distribution for a general dictionary D would be interesting, but is outside the scope of the current paper.

4. RESULTS: COMPARING SYNTHESIS AND ANALYSIS RECOVERY

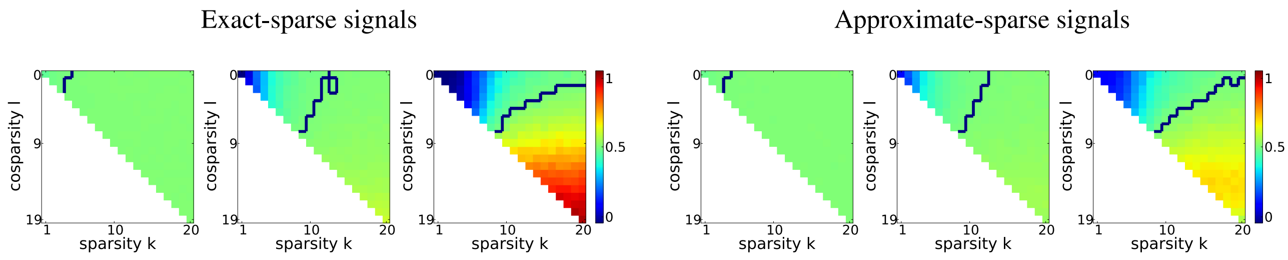
For each of the d^2 pairs (k, l) with $k = 1, 2, ...d$ and $l = 0, 1, 2, ...d - 1$ we attempt to generate 1000 (k, l)-bisparse test signals with a dictionary D. We generate D as a random tight frame of size 20×50. The signals are generated according to the following procedure: (i) choose random subsets $I, J \subset \{1, 2, ...N\}$ with $card(I) = l$ and $card(J) = k$; (ii) check whether the minor matrix M_{IJ} has rank smaller than k (Theorem 3.1); (iii) if yes, find a random solution to (10) and place the coefficients on the locations J of γ_S; (iv) compute the actual signal $x = D\gamma_S$, and (v) compute $\gamma_A = D^{\dagger} x$, count the number l' of zeros and assign x to the set of (k, l')-bisparse signals.

To investigate robustness against approximate sparsity, when generating signals we optionally add a small non-sparse random component to the exact-sparse decomposition γ_S, with energy equal to 1% of γ_S. Thus the resulting signal x will be only approximately sparse and cosparse.

We take m zero-mean, unit-norm random linear measurements of each signal and reconstruct with synthesis-based recovery (4) and analysis-based recovery (8), respectively. For synthesis recovery we use the Smooth L0 algorithm (SL0) [12] for ℓ_0 recovery (the choice of the algorithm is not critical). For analysis recovery we also use a version of SL0, adapted to analysis recovery by incorporating the extra orthogonality constraint of (7) (see [10] for details as well as examples of using this approach for other synthesis recovery algorithms). We compute the percentage RMS (Root-Mean-Square) error of the reconstructed signal $\hat{x}$ defines as

$$R(x) = \sqrt{\frac{\sum (x_i - \hat{x}_i)^2}{\sum x_i^2}}. \quad (11)$$

A smaller value of R indicates a better reconstruction, with $R = 0$ meaning perfect reconstruction. For every pair (k, l)



(a) $m = 3$ (b) $m = 11$ (c) $m = 17$ (d) $m = 3$ (e) $m = 11$ (f) $m = 17$

Fig. 1: Ratio of average reconstruction errors obtained with synthesis and analysis recovery, respectively, for signals simultaneously k-sparse and l-cosparse. On the left side, the signals are exact-sparse, whereas on the right side a small non-sparse component of 1% energy is added. A small value indicates smaller errors with synthesis recovery, large values indicate smaller errors for analysis recovery. The dark separation line indicates the 0.5 frontier (similar performance)

we define the averaged error $R_{kl}^{S,A}$ as the average $R(x)$ for all the signals of the (k, l) pair, with indices S and A indicating synthesis or analysis recovery, respectively. The ratio R_{kl}^{S}/R_{kl}^{A} indicates which kind of recovery is better: a value smaller than 1 indicates that synthesis reconstruction achieves lower average errors, otherwise analysis recovery is the better option.

In Fig.1, we plot the quantity

$$R_{kl} = \left(1 + \frac{R_{kl}^{A}}{R_{kl}^{S}}\right)^{-1} \tag{12}$$

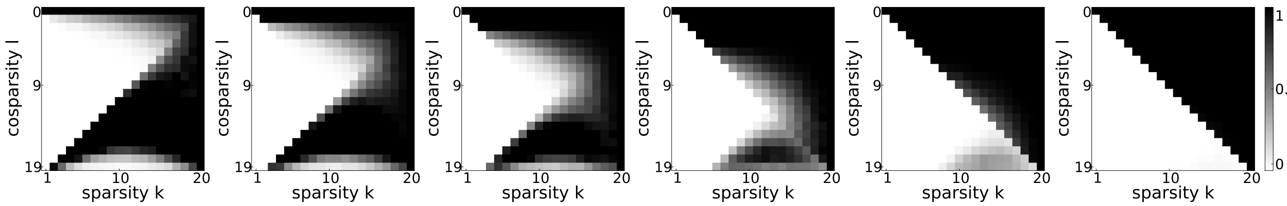
for exact-sparse signals as well as approximate sparse signals. A value of $R_{kl} = 0.5$ indicates similar performance of the two reconstruction algorithms. If synthesis performance is better, the value of R_{kl} tends to 0, whereas if analysis is better, it approaches 1. For each (k, l) pair we attempt to generate 1000 signals. However, for $l \geq k$ we cannot guarantee to find as many, as explained in Section 3. We use the intensity of the color to indicate the number of (k, l)-bisparse signals actually found in 1000 attempts, with pure white indicating that no such signal could be found. We show the results for three different values of the m (3, 11, and 17) to illustrate the influence of the number of measurements.

The first notable fact is that we could not find any (k, l)-bisparse signals for $l \geq k$, and thus we have no data for the lower triangular part of every plot. Further simulations confirm this finding for all other random dictionaries, as long as the overcompleteness factor N/d remains reasonably large (e.g. N/d larger than about 1.5). Under these assumptions, the results suggest that the $(k, k-1)$ main diagonal is in general a maximum limit for joint sparsity and cosparsity, i.e. a k-sparse signal cannot be in general more than $(k-1)$-cosparse at the same time. In other words, a very synthesis-sparse signal x cannot simultaneously be very analysis-cosparse for the same D, and vice-versa.

This result is an interesting argument on the difference between synthesis and analysis reconstruction in the overcomplete case, suggesting that the two sparsity models cannot be in general simultaneously adequate for a signal. However, it does not hold for certain dictionaries which enforce particular relations between the atoms. For example, if D is an equiangular tight frame, all the off-diagonal elements of $D^{\dagger}D$ have the same absolute value, and it is therefore much easier to find linear dependent minor matrices of $D^{\dagger}D$ of larger size than usual, and thus (k, l)-bisparse signals that are more sparse than usual. This suggests than the $(k, k-1)$ joint sparsity limit is valid only in a probabilistic sense. Even though, it can still be useful when working with learned dictionaries, which generally do not have any particular relation enforced between the atoms.

When approaching the complete case (N/d approaching 1), the pattern evolves dramatically, as depicted in Fig.2. Theorem 3.1 shows that finding a (k, l)-bisparse signal depends on the rank of the corresponding M_{IJ} minor matrix being smaller than k. The patterns in Fig.2 are, therefore, an illustration of the probability of finding linear dependent minor matrices of size $l \times k$ inside the matrix $D^{\dagger}D$, for random dictionaries D of various sizes. When D is a basis, $N/d = 1$, the plot concentrates strictly on the first diagonal. This is because in this case the two solutions γ_S and γ_A are identical, and therefore a k-sparse signal is automatically $(d-k)$-cosparse. Further analysis of these distribution patterns, although interesting, is however outside the scope of this paper.

A second interesting observation is that analysis recovery performs well with sufficiently cosparse signals when the number of measurements is large, Fig.1(c), but is less reliable than its synthesis counterpart for fewer measurements and for approximately-sparse signals. In Fig.1(b), analysis recovery is not accurate even for the most cosparse signals (lower-right corner), while synthesis recovery works better for very sparse signals (upper-left corner). In Fig.1(d),(e),(f), a small non-sparse component of 1% energy is added to the test signals. One observes that analysis recovery is significantly more affected by this non-sparse component than synthesis recovery. We conclude therefore that analysis recovery is less robust to approximate sparsity and insufficient measurements than synthesis recovery.



(a) $d \times N = 20 \times 21$ (b) $d \times N = 20 \times 22$ (c) $d \times N = 20 \times 23$ (d) $d \times N = 20 \times 25$ (e) $d \times N = 20 \times 27$ (f) $d \times N = 20 \times 30$

Fig. 2: Percentage of successfully generated signals that are jointly k-sparse and l-cosparse, for a random dictionary of size $d \times N$ with small overcompleteness factor. White indicates that no (k, l)-bisparse signals could be generated in 1000 attempts, and black indicates the all 1000 (k, l)-bisparse signals have been generated.

5. CONCLUSIONS

In this paper we conducted an experimental investigation into determining when is one of the analysis and synthesis sparsity models better than the other in terms of recovering a signal from a few random measurements. Our approach is based on reformulating analysis sparsity as least-squares constrained synthesis sparsity. We consider (k, l)-*bisparse* signals, i.e. signals that are simultaneously k-sparse in a random dictionary and l-cosparse with the pseudoinverse of that dictionary. We generate bisparse test signals for every possible (k, l) pair, reconstruct them separately using both synthesis and analysis recovery, and compare the average recovery errors obtained with the two methods.

The results indicate than the two recovery options perform similarly when recovering signals that are sparse according to the corresponding sparsity model, when the number of measurements is sufficient. However, analysis recovery is significantly more affected than synthesis recovery by a reduction in the number of measurements, and is also less robust with signals that are only approximately sparse. In addition, we find that for random dictionaries with reasonably large overcompleteness factor there is a limit of how much joint sparse and cosparse a signal can be, with a k-sparse signal being in general at most $(k-1)$-cosparse.

As future work, it will be interesting to conduct similar simulations with ℓ_1-sparse signals instead of exact-sparse, increasing the practical usefulness of the results. We also aim to fully investigate the influence of the dictionary size for small overcompleteness factors.

6. REFERENCES

[1] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, pp. 947–968, 2007.

[2] S. Nam, M. Davies, M. Elad, and R. Gribonval, "Cosparse analysis modeling - uniqueness and algorithms," in *Proc. ICASSP 2011*, 2011, pp. 5804–5807.

[3] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006.

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

[5] C. M. Fira, L. Goras, C. Barabasa, and N. Cleju, "ECG compressed sensing based on classification in compressed space and specified dictionaries," in *Proc. EUSIPCO 2011*, 2011, pp. 1573–1577.

[6] N. Cleju, C. M. Fira, C. Barabasa, and L. Goras, "Robust reconstruction of compressively sensed ECG signals," in *Proc. International Symposium on Signals, Circuits and Systems*, 2011, pp. 507–510.

[7] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," Research Report 00602205, 2011.

[8] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, 2005.

[9] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal*, vol. 31, pp. 59–73, 2011.

[10] N. Cleju, M. Jafari, and M. D. Plumbley, "Analysis-based sparse reconstruction with synthesis-based solvers," in *Proc. ICASSP 2012*, 2012, pp. 5401–5404.

[11] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proc. IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

[12] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed l0 norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289 –301, Jan. 2009.