On Sparsity Averaging

Rafael E. Carrillo*, Jason D. McEwen†, and Yves Wiaux*‡§
* Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.
† Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.
‡ Department of Radiology and Medical Informatics, University of Geneva (UniGE), CH-1211 Geneva, Switzerland.
§ Department of Radiology, Lausanne University Hospital (CHUV), CH-1011 Lausanne, Switzerland.

***Abstract*—Recent developments in [1] and [2] introduced a novel regularization method for compressive imaging in the context of compressed sensing with coherent redundant dictionaries. The approach relies on the observation that natural images exhibit strong *average sparsity* over multiple coherent frames. The associated reconstruction algorithm, based on an *analysis* prior and a *reweighted* ℓ_1 scheme, is dubbed Sparsity Averaging Reweighted Analysis (SARA). We review these advances and extend associated simulations establishing the superiority of SARA to regularization methods based on sparsity in a single frame, for a generic spread spectrum acquisition and for a Fourier acquisition of particular interest in radio astronomy.**

I. Introduction

Consider a complex-valued signal $\boldsymbol{x} \in \mathbb{C}^N$, assumed to be sparse in some orthonormal basis $\Psi \in \mathbb{C}^{N\times N}$, and also consider the measurement model $\boldsymbol{y} = \Phi\boldsymbol{x}+\boldsymbol{n}$, where $\boldsymbol{y} \in \mathbb{C}^M$ denotes the measurement vector, $\Phi \in \mathbb{C}^{M\times N}$ with $M < N$ is the sensing matrix and $\boldsymbol{n} \in \mathbb{C}^M$ represents the observation noise. The most common approach in compressed sensing (CS) is to recover $\boldsymbol{x}$ from $\boldsymbol{y}$ solving the following convex problem [3]:

$$\min_{\bar{\boldsymbol{\alpha}}\in\mathbb{C}^N} \|\bar{\boldsymbol{\alpha}}\|_1 \text{ subject to } \|\boldsymbol{y} - \Phi\Psi\bar{\boldsymbol{\alpha}}\|_2 \leq \epsilon, \tag{1}$$

where ϵ is an upper bound on the ℓ_2 norm of the noise and $\|\cdot\|_1$ denotes the ℓ_1 norm of a complex-valued vector. The signal is recovered as $\hat{\boldsymbol{x}} = \Psi\hat{\boldsymbol{\alpha}}$, where $\hat{\boldsymbol{\alpha}}$ denotes the solution to (1). Such problems that solve for the representation of the signal in a sparsity basis are known as synthesis-based problems. The standard CS theory provides results for the recovery of $\boldsymbol{x}$ from $\boldsymbol{y}$ if Φ obeys a Restricted Isometry Property (RIP) and Ψ is orthonormal [3]. However, signals often exhibit better sparsity in an overcomplete dictionary [4]–[6].

Recent works have begun to address the case of CS with redundant dictionaries. In this setting the signal $\boldsymbol{x}$ is expressed in terms of a dictionary $\Psi \in \mathbb{C}^{N\times D}$, $N < D$, as $\boldsymbol{x} = \Psi\boldsymbol{\alpha}$, $\boldsymbol{\alpha} \in \mathbb{C}^D$. Rauhut et al. [7] find conditions on the dictionary Ψ such that the compound matrix $\Phi\Psi$ obeys the RIP to accurately recover $\boldsymbol{\alpha}$ by solving a synthesis-based problem. Candès et al. [8] provide a theoretical analysis of the ℓ_1 analysis-based problem. As opposed to synthesis-based problems, analysis-based problems recover the signal itself solving:

$$\min_{\bar{\boldsymbol{x}}\in\mathbb{C}^N} \|\Psi^\dagger\bar{\boldsymbol{x}}\|_1 \text{ subject to } \|\boldsymbol{y} - \Phi\bar{\boldsymbol{x}}\|_2 \leq \epsilon, \tag{2}$$

where $\Psi^\dagger$ denotes the adjoint operator of Ψ. The aforementioned work [8] extends the standard CS theory to coherent and redundant dictionaries, providing theoretical stability guarantees based on a general condition of the sensing matrix Φ, coined the Dictionary Restricted Isometry Property (D-RIP).

In [1] and [2], we proposed a novel sparsity analysis prior for compressive imaging in the context of CS with coherent and redundant dictionaries, relying on the observation that natural images are simultaneously sparse in various frames, in particular wavelet frames, or in their gradient. Promoting *average sparsity* over multiple frames, as opposed to single frame sparsity, is an extremely powerful prior. The associated reconstruction algorithm, based on an *analysis* prior and a *reweighted* ℓ_1 scheme, is dubbed Sparsity Averaging Reweighted Analysis (SARA)[1].

In this work, we review and further discuss these recent advances. The superiority of SARA to regularization methods based on sparsity in a single frame, as established through simulations for a generic spread spectrum acquisition, is described with an additional extensive visual support. Moreover, we bring a novel illustration for a realistic continuous Fourier sampling strategy of particular interest for radio interferometry in astronomy. We finally discuss possible avenues to establish explicit theoretical stability results for the algorithm.

II. Sparsity Averaging Reweighted Analysis

Natural images are often complicated and encompass several types of structures admitting sparse representations in different frames. For example, piecewise smooth structures exhibit gradient sparsity, while extended structures are better encapsulated in wavelet frames. Observing that natural images actually exhibit sparsity in multiple frames, we hypothesise in [1] and [2] that average sparsity over multiple coherent frames represents a strong prior. We thus proposed the use of a dictionary composed of a concatenation of q frames, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \ldots, \Psi_q], \tag{3}$$

with $\Psi \in \mathbb{C}^{N\times D}$, $N < D$, and an analysis ℓ_0 prior,

$$\|\Psi^\dagger\bar{\boldsymbol{x}}\|_0 \sim \frac{1}{q}\sum_{i=1}^{q} \|\Psi_i^\dagger\bar{\boldsymbol{x}}\|_0, \tag{4}$$

to promote this average sparsity. Note that in this setting each frame contains all the signal information as opposed

[1] In [9], similar ideas were applied to the reverberant audio source separation problem exploiting sparsity in a redundant short time Fourier transform.

to component separation approaches such as [4] and [5]. Also note on a theoretical level that a single signal cannot be arbitrarily sparse simultaneously in a set of incoherent frames. For example, a signal extremely sparse in the Dirac basis is completely spread in the Fourier basis. As discussed in [2], each frame, Ψ_i, should be highly coherent with the other frames in order for the signal to have a sparse representation in the dictionary Ψ. Concatenation of the first eight orthonormal Daubechies wavelet bases (Db1-Db8) is an example of interest. The first Daubechies wavelet basis, Db1, is the Haar wavelet basis. It can be used as an alternative to gradient sparsity, usually imposed by a total variation (TV) prior, to promote piecewise smooth signals. The Db2-Db8 bases provide smoother decompositions. Coherence between the bases is ensured by the compact support of the Daubechies wavelets.

A reweighted ℓ_1 minimization scheme [10] promotes average sparsity through the prior (4). The algorithm replaces the ℓ_0 norm by a weighted ℓ_1 norm and solves a sequence of weighted ℓ_1 problems with weights essentially the inverse of the values of the solution of the previous problem:

$$\min_{\bar{\boldsymbol{x}} \in \mathbb{C}^N} \|\mathsf{W}\Psi^\dagger \bar{\boldsymbol{x}}\|_1 \text{ subject to } \|\boldsymbol{y} - \Phi\bar{\boldsymbol{x}}\|_2 \leq \epsilon, \qquad (5)$$

where $\mathsf{W} \in \mathbb{R}^{D\times D}$ is a diagonal matrix with positive weights. The solution to (5) is denoted as $\Delta(\boldsymbol{y}, \Phi, \mathsf{W}, \epsilon)$. We update the weights at each iteration, i.e. after solving a complete weighted ℓ_1 problem, by the function $f(\gamma, a) \propto (\gamma + |a|)^{-1}$, where a denotes the coefficient value estimated at the previous iteration and γ plays the role of a stabilization parameter, avoiding undefined weights when the signal value is zero. Note that as $\gamma \to 0$ the solution of the weighted ℓ_1 problem approaches the solution of the ℓ_0 problem. We use a homotopy strategy and solve a sequence of weighted ℓ_1 problems with a decreasing sequence $\{\gamma^{(t)}\}$, with t denoting the iteration time variable.

The sparsity averaging reweighted analysis (SARA) algorithm is defined in Algorithm 1. A rate parameter $\beta \in (0, 1)$ controls the decrease of the sequence through $\gamma^{(t)} = \beta\gamma^{(t-1)}$. However, the noise standard deviation σ_α in the representation domain, rough estimate for a baseline above which significant signal components could be identified, serves as a lower bound: $\gamma^{(t)} \geq \sigma_\alpha = \sqrt{M/D}\sigma_n$, with σ_n the noise standard deviation in measurement space. As a starting point we set $\hat{\boldsymbol{x}}^{(0)}$ as the solution of the ℓ_1 problem and $\gamma^{(0)} = \sigma_s\left(\Psi^\dagger\hat{\boldsymbol{x}}^{(0)}\right)$, where $\sigma_s(\cdot)$ takes the empirical standard deviation of a signal. The re-weighting process ideally stops when the relative variation between successive solutions is smaller than some bound $\eta \in (0, 1)$, or after the maximum number of iterations allowed, $N_{\max}$, is reached. We fix $\eta = 10^{-3}$ and $\beta = 10^{-1}$.

Algorithm 1 SARA algorithm

Input: $\boldsymbol{y}$, Φ, ϵ, σ_α, β, η and $N_{\max}$.
Output: Reconstructed image $\hat{\boldsymbol{x}}$.

```
Initialize t = 1, W^(0) = I and ρ = 1.
Compute
  x̂^(0) = Δ(y, Φ, W^(0), ε), γ^(0) = σ_s(Ψ† x̂^(0)).
while ρ > η and t < N_max do
    Update W_ij^(t) = f(γ^(t−1), α̂_i^(t−1)) δ_ij,
    for i, j = 1, ..., D with α̂^(t−1) = Ψ† x̂^(t−1).
    Compute a solution x̂^(t) = Δ(y, Φ, W^(t), ε).
    Update γ^(t) = max{βγ^(t−1), σ_α}.
    Update ρ = ‖x̂^(t) − x̂^(t−1)‖_2 / ‖x̂^(t−1)‖_2
    t ← t + 1
end while
```

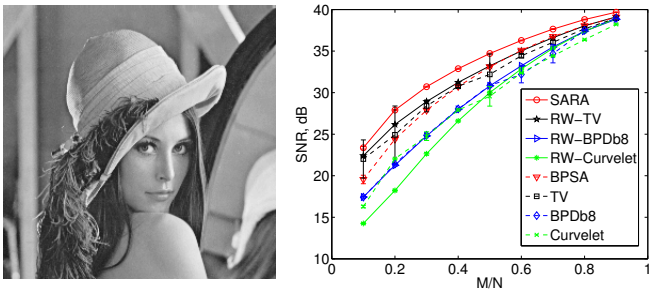

Figure 1. Reconstruction quality results for Lena in the context of a spread spectrum acquisition. Left: original image. Right: SNR results against the undersampling ratio for an input SNR of 30 dB (average values over 100 simulations are shown with corresponding standard deviations).

III. Simulations

In this section, the superiority of SARA to regularization methods based on sparsity in a single frame, as established through simulations in the context of a generic spread spectrum acquisition, is described with a new extensive visual support. Moreover, we bring a novel illustration for a realistic continuous Fourier sampling strategy of particular interest for radio interferometry.

For the first experiment we recover a 256×256 version of Lena from compressive measurements. The spread spectrum technique described in [11] is used as measurement operator. We compare SARA to analogous analysis algorithms, and their reweighted versions, changing the sparsity dictionary Ψ in (2) and (5) respectively. Three different dictionaries are considered: the Daubechies 8 wavelet basis, the redundant curvelet frame [6] and the concatenation of the first eight Daubechies bases described above for SARA. The associated algorithms are respectively denoted BPDb8, Curvelet and BPSA for the non reweighted case. The reweighted versions are respectively denoted RW-BPDb8, RW-Curvelet and SARA. We also compare to the TV prior, where the TV minimization problem is formulated as a constrained problem like (2), but replacing the ℓ_1 norm by the image TV norm. The reweighted version of TV is denoted as RW-TV. Since the image of interest is positive, we impose the additional constraint that $\bar{\boldsymbol{x}} \in \mathbb{R}^N_+$ for all problems. The reconstruction quality of SARA is evaluated as a function of the undersampling ratio M/N, for M/N in the range $[0.1, 0.9]$. The input SNR is set to 30 dB. The SNR results comparing SARA against all the other benchmark methods are shown in the right panel of Figure 1. The results demonstrate that SARA outperforms state-of-the-art methods

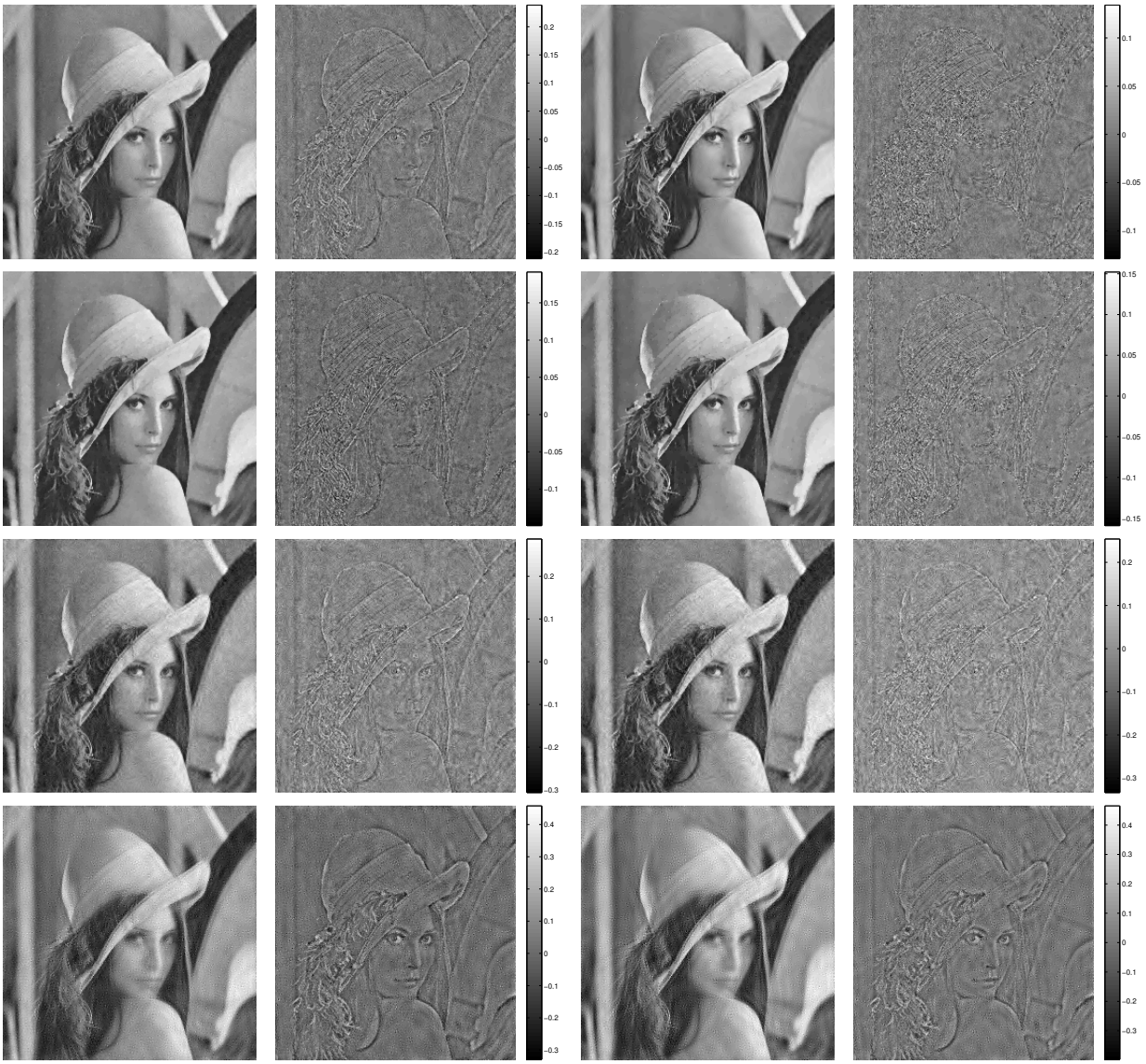

Figure 2. Reconstruction example of Lena for spread spectrum acquisition, with $M = 0.2N$ and input SNR set to 30 dB. First and third columns show the reconstructed images and the second and fourth columns show the error images. First row: BPSA(24.4 dB) and SARA (27.9 dB). Second row: TV(26.3 dB) and RW-TV (26.6 dB). Third row: BPDb8 (21.4 dB) and RW-BPDb8 (21.2 dB). Fourth row: Curvelet (18.7 dB) and RW-Curvelet (18.3 dB).

for all undersampling ratios. RW-TV provides the second best results. BPSA achieves better SNRs than BPDb8, curvelet and their reweighted versions for all undersampling ratios. It also achieves similar SNRs to TV in the range 0.4-0.9. Figure 2 presents a visual assessment for $M = 0.2N$, showing both reconstructed and error images. SARA provides an impressive reduction of visual artifacts relative to the other methods in this high undersampling regime. In particular RW-TV exhibits expected cartoon-like artifacts. Other methods do not yield results of comparable quality, either in SNR or visually, with associated reconstructions full of visual artifacts.

The second experiment illustrates the performance of SARA in the context of radio interferometric imaging by recovering a 256×256 version of the well known M31 galaxy from simulated continuous Fourier samples associated with a realistic radio telescope sampling pattern (superposition of arcs of ellipses). The number of measurements is $M = 9413$, affected by 30 dB of input noise. The dictionary for SARA is the concatenation of the first eight Daubechies bases *and* the Dirac basis. The Dirac basis is added given the sparsity in image space due to the large field of view. For comparison, we use two different methods: BP, constrained ℓ_1-minimization in the Dirac basis (used as benchmark in the field), and BPDb8, constrained analysis-based ℓ_1-minimization in the Db8 basis. Figure 3 shows the original test image, the sampling pattern and the corresponding dirty image, i.e. the inverse Fourier transform of the measurements, with non-measured points set to zero. The reconstructed images for BP, BPDb8 and SARA are also reported. Once more, SARA provides not only a drastic SNR increase but also a significant reduction of visual

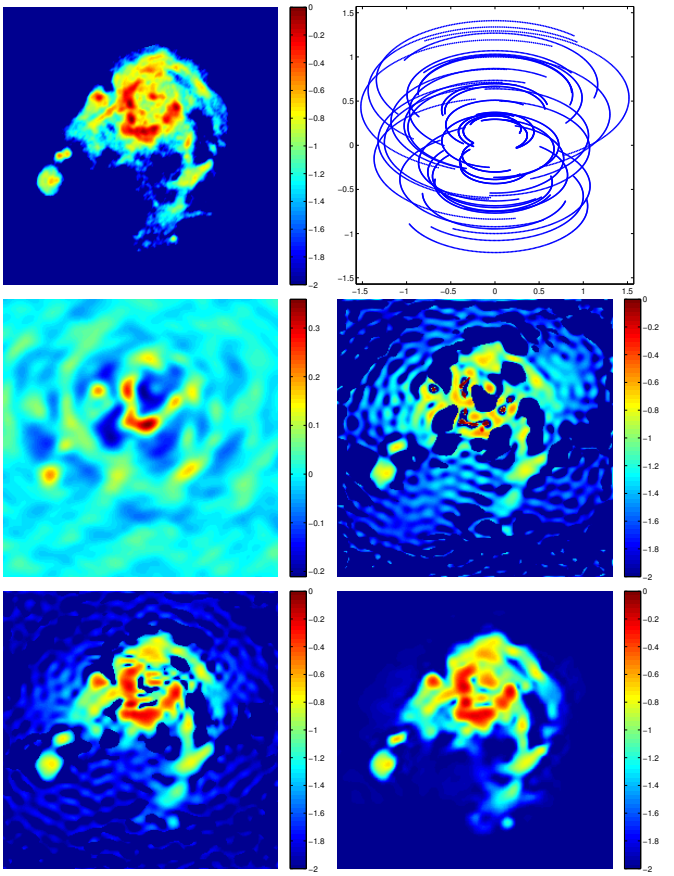

Figure 3. Radio Astronomy example. From left to right. Top row: original test image in $\log_{10}$ scale and Fourier sampling pattern. Middle row: corresponding dirty image in linear scale and reconstruction results for BP (3.9 dB) in $\log_{10}$ scale. Bottom row: reconstruction results for BPDb8 (10.3 dB) and SARA (14.1 dB) in $\log_{10}$ scale.

artifacts relative to the other methods.

IV. Conclusion and Discussion

In this paper we have reviewed recent advances in the average sparsity model and the associated algorithm SARA. Extended simulations demonstrating the superiority of SARA for compressive imaging reconstruction were described. Novel results on the application of SARA to a realistic radio interferometric imaging scenario were also described.

Future work will concentrate on finding a theoretical framework for the average sparsity model. In [2] we have put average sparsity in the context of theory developed in [8]. However, specialized results for the particular case of concatenation of frames (or orthogonal bases) are needed. The co-sparsity analysis model [12] proposes a general framework for general analysis operators. Similar properties to the D-RIP coined Ω-RIP are introduced in [13] to analyze greedy algorithms in the context of the co-sparsity analysis model. It would be interesting to explore the connections between average sparsity and the co-sparsity model to have an estimate on the number of measurements needed for reconstruction compared to single frame representations.

The proposed approach relies on the observation that natural images exhibit strong average sparsity, i.e. the signals of interest have so-called simultaneous structured models. Recently, it was shown in [14] that combinations of convex relaxations of the individual structured models do not yield better results than an algorithm that exploits only one of the structured models, while *non-convex* approaches that approximate the simultaneous model can exploit the multiple structured models. Those results suggest that the *re-weighting* approach in SARA to approximate the ℓ_0 norm is fundamental to exploit average sparsity, as observed in the simulation results (see the gap between SARA and BPSA in Fig. 1 and Fig. 2).

Acknowledgment

REC is supported by the Swiss National Science Foundation (SNSF) under grant 200021-130359. JDM is supported by a Newton International Fellowship from the Royal Society and the British Academy. YW is supported by the Center for Biomedical Imaging (CIBM) of the Geneva and Lausanne Universities and EPFL.

References

[1] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.

[2] R. E. Carrillo, J. D. McEwen, D. V. D. Ville, J.-P. Thiran, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Processing Letters*, 2013, accepted for publication. Preprint available at http://infoscience.epfl.ch/record/180382.

[3] M. Fornasier and H. Rauhut, *Handbook of Mathematical Methods in Imaging*. Springer, 2011, ch. Compressed sensing.

[4] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[5] J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D. Donoho, "Morphological component analysis: an adaptative thresholding strategy," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.

[6] J. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, Cambridge, GB, 2010.

[7] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.

[8] E. J. Candès, Y. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2010.

[9] S. Arberet, P. Vandergheynst, R. E. Carrillo, J.-P. Thiran, and Y. Wiaux, "Sparse reverberant audio source separation via reweighted analysis," *IEEE Transactions on Audio Speech and Language Processing*, 2013, accepted for publication. Preprint available at http://infoscience.epfl.ch/record/180378.

[10] E. J. Candès, M. Wakin, and S. Boyd, "Enhacing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[11] G. Puy, P. Vandergheynst, R. Gribonval, and Y. Wiaux, "Universal and efficient compressed sensing by spread spectrum and application to realistic fourier imaging techniques," *EURASIP Journal on Applied signal Processing*, vol. 2012, no. 3, 2012.

[12] S. Nam, M. Davies, R.Gribonval, and M. Elad, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[13] R. Giryes, S. Nam, M. Elad, R.Gribonval, and M. Davies, "Greedy-like algorithms for the cosparse analysis model," 2013, preprint, arXiv:1207.2456v2.

[14] S. Oymak, A. Jalali, M. Fazel, Y. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," 2013, preprint, arXiv:1212.3753v2.