

Sparsity Averaging for Compressive Imaging

Rafael E. Carrillo, Jason D. McEwen, Dimitri Van De Ville, Jean-Philippe Thiran, and Yves Wiaux

Abstract—We discuss a novel sparsity prior for compressive imaging in the context of the theory of compressed sensing with coherent redundant dictionaries, based on the observation that natural images exhibit strong average sparsity over multiple coherent frames. We test our prior and the associated algorithm, based on an analysis reweighted ℓ_1 formulation, through extensive numerical simulations on natural images for spread spectrum and random Gaussian acquisition schemes. Our results show that average sparsity outperforms state-of-the-art priors that promote sparsity in a single orthonormal basis or redundant frame, or that promote gradient sparsity. Code and test data are available at <https://github.com/basp-group/sopt>.

Index Terms—Compressed sensing, sparse approximation.

I. INTRODUCTION

Compressed sensing (CS) introduces a signal acquisition framework that goes beyond the traditional Nyquist sampling paradigm [1]. Consider a complex-valued signal $\mathbf{x} \in \mathbb{C}^N$, assumed to be sparse in some orthonormal basis $\Psi \in \mathbb{C}^{N \times N}$, i.e. $\mathbf{x} = \Psi\alpha$ for $\alpha \in \mathbb{C}^N$ sparse. Also consider the measurement model $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$, where $\mathbf{y} \in \mathbb{C}^M$ denotes the measurement vector, $\Phi \in \mathbb{C}^{M \times N}$ with $M < N$ is the sensing matrix, and $\mathbf{n} \in \mathbb{C}^M$ represents noise. The most common approach to recover \mathbf{x} from \mathbf{y} is to solve the following convex problem [1]: $\min_{\tilde{\alpha} \in \mathbb{C}^N} \|\tilde{\alpha}\|_1$ subject to $\|\mathbf{y} - \Phi\Psi\tilde{\alpha}\|_2 \leq \epsilon$, where ϵ is an upper bound on the ℓ_2 norm of the noise and $\|\cdot\|_1$ denotes the ℓ_1 norm. The signal is recovered as $\hat{\mathbf{x}} = \Psi\hat{\alpha}$, where $\hat{\alpha}$ denotes the solution to the above problem. Such problems, solving for the signal representation in a sparsity basis, are known as synthesis-based problems. Standard CS provides results if Φ obeys a Restricted Isometry Property (RIP) and Ψ is orthonormal [1]. However, signals often exhibit better sparsity in a redundant dictionary [2]–[4].

Recent works have begun to address CS with redundant dictionaries, i.e. where $\Psi \in \mathbb{C}^{N \times D}$, with $N < D$, so that $\mathbf{x} = \Psi\alpha$ with $\alpha \in \mathbb{C}^D$. Rauhut et al. [5] find conditions on Ψ such that $\Phi\Psi$ obeys the RIP to recover α in a synthesis formulation. Candès et al. [6] provide a theoretical analysis of the ℓ_1 analysis-based problem. As opposed to synthesis, the analysis formulation solves for the signal itself:

$$\min_{\tilde{\mathbf{x}} \in \mathbb{C}^N} \|\Psi^\dagger \tilde{\mathbf{x}}\|_1 \text{ subject to } \|\mathbf{y} - \Phi\tilde{\mathbf{x}}\|_2 \leq \epsilon, \quad (1)$$

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, and by the SNSF under grant PP00P2-123438.

REC, DVDV, JPT and YW are with the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. DVDV and YW are also with the Department of Radiology and Medical Informatics, University of Geneva (UniGE), CH-1211 Geneva, Switzerland. JDM is with the Department of Physics and Astronomy, University College London, London WC1E 6BT, U.K.

E-mail: rafael.carrillo@epfl.ch (REC).

where Ψ^\dagger denotes the adjoint operator of Ψ . The aforementioned work [6] 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). The D-RIP is a natural extension of the standard RIP. In fact many random matrices that obey the standard RIP also obey the D-RIP, like Gaussian or Bernoulli ensembles. Also, the subsampled Fourier matrix multiplied by a random sign matrix satisfies the D-RIP [7], which provides a fast sensing operator. Interestingly, this approach falls within the spread spectrum framework proposed in [8]. If Φ satisfies the D-RIP and Ψ is a general frame, Candès et al. prove in [6] that the solution to (1), denoted $\hat{\mathbf{x}}$, satisfies the following error bound:

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0\epsilon + C_1K^{-1/2} \|\Psi^\dagger \mathbf{x} - (\Psi^\dagger \mathbf{x})_K\|_1, \quad (2)$$

where $(\Psi^\dagger \mathbf{x})_K$ denotes the best K -term approximation of $\Psi^\dagger \mathbf{x}$ and C_0 and C_1 are numerical constants. Similar properties to the D-RIP coined Ω -RIP are introduced in [9] in the context of the co-sparsity analysis model.

In [10] some of the authors of this paper proposed a novel sparsity analysis prior in the context of Fourier imaging in radio astronomy. Our approach relies on the observation that natural images are simultaneously sparse in various frames, in particular wavelet frames, or in their gradient, so that promoting average signal sparsity over multiple frames should be a powerful prior. In the present work, the average sparsity prior is put in the generic context of compressive imaging within the theory of CS with coherent redundant dictionaries. The associated reconstruction algorithm, based on an analysis reweighted ℓ_1 formulation, is dubbed Sparsity Averaging Reweighted Analysis (SARA). We evaluate SARA through extensive numerical simulations for spread spectrum and Gaussian acquisition schemes. Our results show that the average sparsity prior outperforms state-of-the-art priors.

II. SPARSITY AVERAGING REWEIGHTED ANALYSIS

Natural images are often complicated and include 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. Therefore, in [10] we observed that promoting average sparsity over multiple bases rather than a single basis is an extremely powerful prior. Here, we propose using a dictionary composed of a concatenation of q frames Ψ_i with $1 \leq i \leq q$. We focus on the particular case of concatenation of Parseval frames, creating the Parseval frame $\Psi \in \mathbb{C}^{N \times D}$, with $N < D$, as:

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \dots, \Psi_q]. \quad (3)$$

The analysis-based framework is a suitable approach to promote average sparsity and thus we propose the following prior, proportional to the average sparsity:

$$\|\Psi^\dagger \bar{x}\|_0 = \sum_{i=1}^q \|\Psi_i^\dagger \bar{x}\|_0. \quad (4)$$

Note that in this setting each frame contains all the signal information. Such a prior cannot be formulated in a synthesis-based perspective. Previous works considering multiple frames, e.g. [2], [3], consider a component separation approach, decomposing the signal as $x = \sum_{i=1}^q x_i$, where each component x_i is sparse in the i -th frame. This is a completely different problem, where each component bears only part of the signal information, which can be addressed either in an analysis or in a synthesis framework.

Also note on a theoretical level that a single signal cannot be arbitrarily sparse simultaneously in a set of incoherent frames [11]. For example, a signal extremely sparse in the Dirac basis is completely spread in the Fourier basis and thus (2) does not provide a good error bound. As discussed by Candès et al. in [6], what is important is that the columns of the Gram matrix $\Psi^\dagger \Psi$ are reasonably sparse such that $\Psi^\dagger x$ is sparse when x admits a sparse representation α with $x = \Psi \alpha$. This requirement is nothing else than a coherence condition on Ψ . In our case of concatenations of frames, this leads to the condition that each Ψ_i is highly coherent with itself and mutually coherent with the other frames. The component separation approaches in [2], [3] use incoherent frames for the decomposition, while our average sparsity prior takes the opposite direction. The concatenation of the first eight orthonormal Daubechies wavelet bases (Db1-Db8, $q = 8$) represents a good and simple candidate for a dictionary in imaging applications. The first Daubechies wavelet basis, Db1, is the Haar wavelet basis, which can be used as an alternative to gradient sparsity (usually imposed by a total variation (TV) prior [12]) to promote piecewise smooth signals. The Db2-Db8 bases provide smoother sparse decompositions. All Daubechies bases are mutually coherent thanks to their compact support and identical sampling positions.

In order to promote average sparsity through the prior (4) we adopt a reweighted ℓ_1 minimization scheme [13]. 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:

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

where $\mathbf{W} \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights. Assuming i.i.d. complex Gaussian noise with variance σ_n , the ℓ_2 norm term in (5) is identical to a bound on the χ^2 with $2M$ degrees of freedom governing the noise level estimator. Therefore, we set this bound as $\epsilon^2 = (2M + 4\sqrt{M})\sigma_n^2/2$, where $\sigma_n^2/2$ is the variance of both the real and imaginary parts of the noise. This choice provides a likely bound for $\|\mathbf{n}\|_2$ [10]. To solve (5), we use the Douglas-Rachford splitting algorithm [14]. The solution is denoted as $\Delta(\mathbf{y}, \Phi, \mathbf{W}, \epsilon)$. The weights are updated at each iteration, i.e. after solving a complete

weighted ℓ_1 problem, by the function $f(\gamma, a) = \gamma(\gamma + |a|)^{-1} \in (0, 1]$, where a denotes the coefficient value estimated at the previous iteration and $\gamma \neq 0$ plays the role of a stabilization parameter, avoiding undefined weights when the signal value is zero. Note that as $\gamma \rightarrow 0$ 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 using a decreasing sequence $\{\gamma^{(t)}\}$, with t denoting the iteration time variable. The resulting algorithm, dubbed sparsity averaging reweighted analysis (SARA), is defined in Algorithm 1¹. See [10] for more details.

Algorithm 1 SARA algorithm

Input: $\mathbf{y}, \Phi, \epsilon, \sigma_\alpha, \beta, \eta$ and N_{\max} .

Output: Reconstructed image \hat{x} .

- 1: Initialize $t = 1$, $\mathbf{W}^{(0)} = \mathbf{I}$ and $\rho = 1$.
 - 2: Compute $\hat{x}^{(0)} = \Delta(\mathbf{y}, \Phi, \mathbf{W}^{(0)}, \epsilon)$, $\gamma^{(0)} = \sigma_s(\Psi^\dagger \hat{x}^{(0)})$.
 - 3: **while** $\rho > \eta$ and $t < N_{\max}$ **do**
 - 4: Update $\mathbf{W}_{ij}^{(t)} = f(\gamma^{(t-1)}, \hat{\alpha}_i^{(t-1)}) \delta_{ij}$,
for $i, j = 1, \dots, D$ with $\hat{\alpha}^{(t-1)} = \Psi^\dagger \hat{x}^{(t-1)}$.
 - 5: Compute a solution $\hat{x}^{(t)} = \Delta(\mathbf{y}, \Phi, \mathbf{W}^{(t)}, \epsilon)$.
 - 6: Update $\gamma^{(t)} = \max\{\beta \gamma^{(t-1)}, \sigma_\alpha\}$.
 - 7: Update $\rho = \|\hat{x}^{(t)} - \hat{x}^{(t-1)}\|_2 / \|\hat{x}^{(t-1)}\|_2$
 - 8: $t \leftarrow t + 1$
 - 9: **end while**
-

III. EXPERIMENTAL RESULTS

In this section we evaluate the reconstruction performance of SARA by recovering a 256×256 pixel version of the Lena test image from compressive measurements following the measurement model presented in Section I. We use the suggested Db1-Db8 concatenation as the dictionary for SARA. In order to have a fast measurement operator that obeys the D-RIP, we use for a first experiment the spread spectrum technique described in [8]. Spread spectrum incorporates a modulating sequence on top of Fourier sampling, defining the measurement operator as $\Phi = \mathbf{MFC}$, where $\mathbf{C} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with elements with unit norm and randomized sign, $\mathbf{F} \in \mathbb{C}^{N \times N}$ is the discrete Fourier operator and $\mathbf{M} \in \mathbb{R}^{M \times N}$ is a binary mask defining the random selection operator. For a second experiment we consider Gaussian random measurement matrices.

We compare SARA to analogous analysis algorithms, and their reweighted versions, changing the sparsity dictionary Ψ in (1) and (5) respectively. The three different dictionaries

¹A rate parameter $\beta \in (0, 1)$ controls the decrease of the sequence $\gamma^{(t)} = \beta \gamma^{(t-1)}$. In practice $\gamma^{(t)}$ should however not reach zero. The noise standard deviation in the sparsity domain $\sigma_\alpha = \sqrt{M/D} \sigma_n$, with σ_n the noise standard deviation in measurement space, is a rough estimate for a baseline above which significant signal components could be identified. Hence we set $\gamma^{(t)} = \max\{\beta \gamma^{(t-1)}, \sigma_\alpha\}$ so that $\gamma^{(t)}$ is lower-bounded by σ_α . As a starting point we set $\hat{x}^{(0)}$ as the solution of the ℓ_1 problem and $\gamma^{(0)} = \sigma_s(\Psi^\dagger \hat{x}^{(0)})$, where $\sigma_s(\cdot)$ takes the empirical standard deviation of a signal. The re-weighting process stops when the relative variation between successive solutions is smaller than some bound $\eta \in (0, 1)$, or after a maximum number of iterations N_{\max} . We fix $\eta = 10^{-3}$ and $\beta = 10^{-1}$.

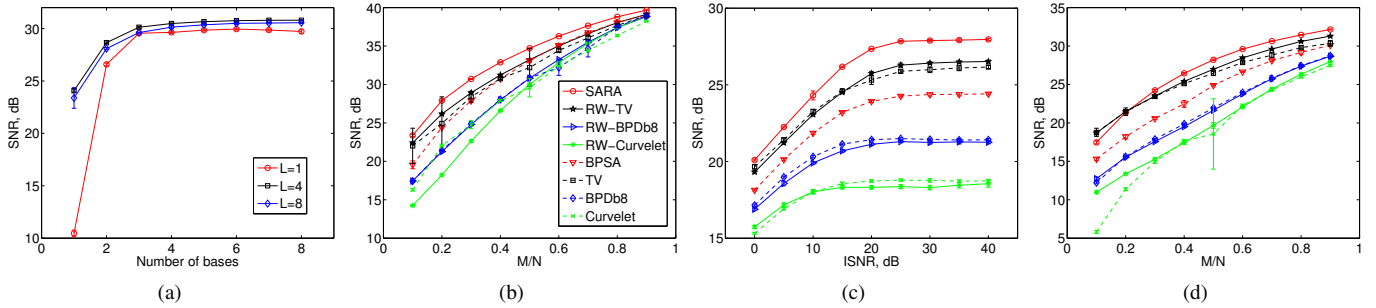


Figure 1. Reconstruction quality results for Lena and spread spectrum measurements. (a) SNR as a function of the number of bases in the dictionary for decomposition depths $L = 1, 4, 8$ ($M = 0.3N$, $\text{ISNR} = 30$ dB). (b) SNR results against the undersampling ratio ($\text{ISNR} = 30$ dB). (c) SNR as a function of ISNR ($M = 0.2N$). (d) Results for random random Gaussian measurements. SNR against the undersampling ratio for cropped Lena image ($\text{ISNR} = 30$ dB).

are: the Daubechies 8 wavelet basis, the redundant curvelet frame [4] and the Db1-Db8 concatenation. The associated algorithms are respectively denoted BPD8, Curvelet and BPSA for the non reweighted case. The reweighted versions are respectively denoted RW-BPD8, RW-Curvelet and SARA. We also compare to the TV prior [12], where the TV minimization problem is formulated as a constrained problem like (1), 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 $\bar{x} \in \mathbb{R}_+^N$ for all problems.

We use as reconstruction quality metric the standard signal-to-noise ratio (SNR), defined as $\text{SNR} = 20 \log_{10} (\|\mathbf{x}\|_2 / \|\mathbf{x} - \hat{\mathbf{x}}\|_2)$, where \mathbf{x} and $\hat{\mathbf{x}}$ denote the original and the estimated image respectively. Average values over 30 simulations and associated 1σ error bars are reported for all experiments. The measurements are corrupted by complex Gaussian noise. The associated input SNR is defined as $\text{ISNR} = 20 \log_{10} (\|\mathbf{y}_0\|_2 / \|\mathbf{n}\|_2)$, where \mathbf{y}_0 identifies the clean measurement vector.

We start by evaluating SARA for spread spectrum acquisition. Prior to our main analysis, we study the reconstruction performance of SARA as a function of the number of wavelet bases in the dictionary. We test depths $L = 1, 4, 8$ in the Daubechies decomposition for all dictionaries, fixing $M = 0.3N$ and $\text{ISNR} = 30$ dB. We add bases in parametric order, i.e., one basis means Db1 alone, two bases Db1 and Db2 and so on until we reach the eight bases from Db1-Db8. The results for Lena are summarized in Figure 1(a). We can observe that the best performance is obtained when $L = 4$ and the worst when $L = 1$. We can also observe that the reconstruction quality improves as the number of bases increases until it saturates between 4 to 8 bases. These results corroborate our choice for 8 bases, and $L = 4$.

Having validated the dictionary choice, we now proceed to evaluate the reconstruction quality of SARA as a function of the undersampling ratio M/N . We fix $\text{ISNR} = 30$ dB and vary the undersampling ratio from 0.1 to 0.9. The SNR results comparing SARA against all the other benchmark methods are shown in Figure 1(b). The results demonstrate that SARA outperforms state-of-the-art methods for all undersamplings. SARA achieves gains between 0.9 and 1.9 dB with the largest gains observed for undersampling ratios in the range 0.2-0.5.

Notably, BPSA achieves better SNR than BPD8, curvelet and their reweighted versions for all undersampling ratios. It also achieves similar SNR to TV in the range 0.4-0.9.

The following experiment studies the robustness of SARA against measurement noise in the spread spectrum acquisition setting. We fix $M = 0.2N$ and vary the ISNR in the range 0 to 40 dB. The results are summarized in Figure 1(c). As expected from the bound in (2), the relationship between SNR and ISNR is linear with slope 1 for low ISNR until it is high enough and the reconstruction quality is dominated by the undersampling effect. Notably, SARA outperforms the benchmark methods for all ISNR , achieving an SNR of 20 dB for an ISNR of 0 dB. Again, BPSA yields a better performance than BPD8, Curvelet and their reweighted versions.

Next we present a visual assessment of the reconstruction quality of SARA compared to the benchmark methods, still in the spread spectrum acquisition setting. Figure 2 shows the reconstructions for $M = 0.2N$ and $\text{ISNR} = 30$ dB for the three best algorithms in SNR: SARA (28.1 dB), RW-TV (26.3 dB) and BPD8 (21.4 dB). 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. BPD8 does not yield results of comparable visual quality.

We now study the performance of SARA with Gaussian random matrices as measurements operators. Due to computational limitations for the use of a dense sensing matrix, for this experiment we use a cropped version of Lena, around the head, of dimension 128×128 as test image. We compare SARA against all the benchmark methods for this sensing modality. We fix $\text{ISNR} = 30$ dB and vary the undersampling ratio in the range 0.1 to 0.9. The SNR results are reported in Figure 1(d). These results confirm the performance of SARA for compressive imaging with a different sensing matrix, outperforming the benchmark methods for $M \geq 0.3N$. For $M = 0.1N$ SARA is 1 dB below TV and RW-TV and for $M = 0.2N$ it achieves the same SNR.

As final experiment, we present a magnetic resonance (MR) imaging illustration. We reconstruct a 224×168 positive brain image from standard variable density Fourier measurements, for an adverse undersampling ratio of $M = 0.05N$, well beyond current state of the art in the field. The ISNR is set to 30 dB. In this case, the sparsity dictionary for SARA is

augmented with the Dirac basis as the brain is quite localized in the field of view. Figure 3 shows a zoom of the original brain image and reconstructed images for SARA and TV, which yield the two best reconstructions in SNR. In addition to an SNR gain of 1.5 dB, SARA achieves an impressively better reconstruction from the visual standpoint.



Figure 2. Reconstruction example for Lena in spread spectrum acquisition setting ($M = 0.2N$, ISNR = 30 dB). From left to right and top to bottom: original image, reconstructed images for SARA (28.1 dB), RW-TV (26.3 dB) and BPDb8 (21.4 dB).

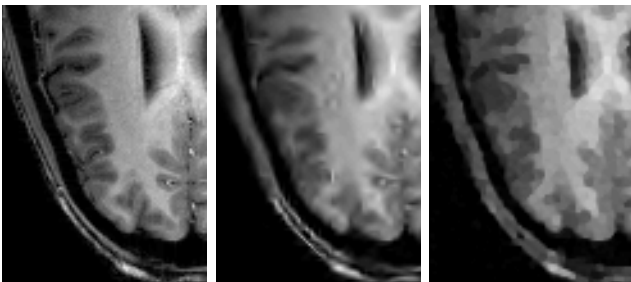


Figure 3. MR illustration: reconstruction of a brain image from Fourier acquisition ($M = 0.05N$, ISNR = 30 dB). From left to right: original image, SARA (18.8 dB) and TV (17.3 dB) reconstructions.

IV. CONCLUSION

In this letter we have discussed the novel SARA regularization method and algorithm for compressive imaging in the theoretical context of CS with coherent redundant dictionaries. The approach relies on the observation that natural images exhibit strong average sparsity. We have evaluated SARA under two different acquisition schemes: spread spectrum and random Gaussian measurements. Experimental results demonstrate that the sparsity averaging prior embedded in the analysis reweighted ℓ_1 formulation of SARA outperforms

state-of-the-art priors, based on single frame or gradient sparsity, both in terms of SNR and visual quality. An MR imaging illustration also corroborates these conclusions for Fourier imaging. Code and test data are available at <https://github.com/basp-group/sopt>.

Future work will concentrate on finding a theoretical framework for the average sparsity model. Specialized results are indeed needed in the particular case of concatenation of frames for an estimate of the number of measurements required for accurate image reconstruction. It would be interesting to explore the connections between average sparsity and the co-sparsity model, which proposes a general framework for general analysis operators (see [9] and references therein). Also, it was recently shown in [15] that combinations of convex relaxation priors do not yield better results than exploiting only one of those priors, while non-convex approaches can exploit multiple models. Those results suggest that the re-weighting approach in SARA to approximate the non-convex ℓ_0 norm is fundamental to exploit average sparsity, as observed in the simulation results.

REFERENCES

- [1] M. Fornasier and H. Rauhut, *Handbook of Mathematical Methods in Imaging*. Springer, 2011, ch. Compressed sensing.
- [2] 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.
- [3] J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D. Donoho, "Morphological component analysis: an adaptive thresholding strategy," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.
- [4] J. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, Cambridge, GB, 2010.
- [5] 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.
- [6] 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.
- [7] F. Krahermer and R. Ward, "New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property," *SIAM J. Math. Anal.*, vol. 43, no. 3, pp. 1269–1281, 2011.
- [8] 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.
- [9] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. Davies, "Greedy-like algorithms for the cosparsity analysis model," 2013, preprint, arXiv:1207.2456v2.
- [10] 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.
- [11] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *Information Theory, IEEE Transactions on*, vol. 48, no. 9, pp. 2558 – 2567, sep 2002.
- [12] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 89–97, 2004.
- [13] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.
- [14] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.
- [15] 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.