On Sparsity Averaging

Rafael E. Carrillo*, Jason D. McEwen†, and Yves Wiaux*‡§

† 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 \( \ell_1 \) 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 \( x \in \mathbb{C}^{N} \), assumed to be sparse in some orthonormal basis \( \Psi \in \mathbb{C}^{N \times N} \), and also consider the measurement model \( y = \Phi x + n \), where \( y \in \mathbb{C}^{M} \) denotes the measurement vector, \( \Phi \in \mathbb{C}^{M \times N} \) with \( M < N \) is the sensing matrix and \( n \in \mathbb{C}^{M} \) represents the observation noise. The most common approach in compressed sensing (CS) is to recover \( x \) from \( y \) solving the following convex problem [3]:

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

where \( \epsilon \) is an upper bound on the \( \ell_2 \) norm of the noise and \( \| \cdot \|_1 \) denotes the \( \ell_1 \) norm of a complex-valued vector. The signal is recovered as \( \hat{x} = \Psi \hat{\alpha} \), where \( \hat{\alpha} \) 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 \( x \) from \( y \) if \( \Phi \) obeys a Restricted Isometry Property (RIP) and \( \Psi \) 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 \( x \) is expressed in terms of a dictionary \( \Psi \in \mathbb{C}^{N \times D} \), \( N < D \), as \( x = \Psi \alpha \), \( \alpha \in \mathbb{C}^{D} \). Rauhut et al. [7] find conditions on the dictionary \( \Psi \) such that the compound matrix \( \Phi \Psi \) obeys the RIP to accurately recover \( \alpha \) by solving a synthesis-based problem. Candès et al. [8] provide a theoretical analysis of the \( \ell_1 \) analysis-based problem. As opposed to synthesis-based problems, analysis-based problems recover the signal itself solving:

\[
\min_{\alpha \in \mathbb{C}^{N}} \| \Psi^\dagger \alpha \|_1 \text{ subject to } \| y - \Phi \alpha \|_2 \leq \epsilon, \tag{2}
\]

where \( \Psi^\dagger \) denotes the adjoint operator of \( \Psi \). 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 \( \Phi \), 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 \( \ell_1 \) 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 \( \ell_0 \) prior,

\[
\| \psi^\dagger \alpha \|_0 \sim \frac{1}{q} \sum_{i=1}^{q} \| \psi_i^\dagger x \|_0, \tag{4}
\]

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

\(^1\)In [9], similar ideas were applied to the reverberant audio source separation problem exploiting sparsity in a redundant short time Fourier transform.
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, $\Psi$, should be highly coherent with the other frames in order for the signal to have a sparse representation in $\Psi$. Concatenation of the first eight orthonormal Daubechies wavelet bases (Db1-D8) 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-D8 bases provide smoother decompositions. Coherence between the bases is ensured by the compact support of the Daubechies wavelets.

A reweighted $\ell_1$ minimization scheme [10] promotes average sparsity through the prior (4). The algorithm replaces the $\ell_0$ norm by a weighted $\ell_1$ norm and solves a sequence of weighted $\ell_1$ problems with weights essentially the inverse of the values of the solution of the previous problem:

$$\min_{\hat{x} \in \mathbb{C}^N} ||W\hat{\Psi}\hat{x}||_1 \text{ subject to } ||y - \Phi \hat{x}||_2 \leq \epsilon,$$  \hspace{1cm} (5)$$

where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights. The solution to (5) is denoted as $\Delta(y, \Phi, W, \epsilon)$. We update the weights at each iteration, i.e., after solving a complete weighted $\ell_1$ problem, by the function $f(\gamma, \alpha) \propto (\gamma + |\alpha|)^{-1}$, where $\alpha$ denotes the coefficient value estimated at the previous iteration and $\gamma$ 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 $\ell_1$ problem approaches the solution of the $\ell_0$ problem. We use a homotopy strategy and solve a sequence of weighted $\ell_1$ 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, with $\Psi$ defined as in (3). 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 $\sigma_\eta$ in the representation domain, rough estimate for a baseline above which significant signal components could be identified, serves as a lower bound: $\gamma(t) \geq \sigma_\eta = \sqrt{M/D}\sigma_\eta$, with $\sigma_\eta$ the noise standard deviation in measurement space. As a starting point we set $\hat{x}(0)$ as the solution of the $\ell_1$ problem and $\gamma(0) = \sigma_\eta (\Psi^\dagger \hat{x}(0))$, where $\sigma_\eta(\cdot)$ takes the empirical standard deviation of a signal. The re-weighting procedure 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, $N_{\text{max}}$, is reached. We fix $\eta = 10^{-3}$ and $\beta = 10^{-1}$.

### 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 $256 \times 256$ 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 $\Psi$ 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 BPD8, Curvelet and BPSA for the non reweighted case. The reweighted versions are respectively denoted RW-BPD8, RW-Curvelet and SARA. Additionally, we also compare to the TV prior, where the TV minimization problem is formulated as a constrained problem like (2), but replacing the $\ell_1$ 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 $\hat{x} \in \mathbb{R}_+^N$ 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-

![Algorithm 1 SARA algorithm](algorithm.png)

**Algorithm 1 SARA algorithm**

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

Output: Reconstructed image $\hat{x}$.

1: Initialize $t = 1$, $W(0) = I$ and $\rho = 1$.
2: Compute $\hat{x}(0) = \Delta(y, \Phi, W(0), \epsilon)$, $\gamma(0) = \sigma_\eta (\Psi^\dagger \hat{x}(0))$.
3: while $\rho > \eta$ and $t < N_{\text{max}}$ do
   4: Update $W_i^{(t)}(0) = f\left(\gamma(t-1)\omega_i^{(t-1)}\delta_{ij}\right)$, for $i, j = 1, \ldots, D$ with $\omega_i^{(t-1)} = \Psi^\dagger \hat{x}(t-1)$.
5: Compute a solution $\hat{x}^{(t)} = \Delta(y, \Phi, W(t), \epsilon)$.
6: Update $\gamma(t) = \max\{\beta \gamma(t-1) - \sigma_\eta\}$.
7: Update $\rho = ||\hat{x}^{(t)} - \hat{x}^{(t-1)}||_2/||\hat{x}^{(t-1)}||_2$.
8: $t \leftarrow t + 1$
9: 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).
art methods 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 $256 \times 256$ 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 $\ell_1$-minimization in the Dirac basis (used as benchmark in the field), and BPDb8, constrained analysis-based $\ell_1$-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 artifacts.
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 \( \ell_1 \) 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


