Compressive sensing for spectroscopy and polarimetry

A. Asensio Ramos1,2 and A. López Ariste3

1 Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain
2 Departamento de Astrofísica, Universidad de La Laguna, 38205 La Laguna, Tenerife, Spain
3 THEMIS, CNRS UPS 853, c/vía Láctea s/n, 38200 La Laguna, Tenerife, Spain

e-mail: asensio@iac.es
e-mail: arturo@themis.iac.es

Received 30 July 2009 / Accepted 24 September 2009

ABSTRACT

We demonstrate, through numerical simulations with real data, the feasibility of using compressive sensing techniques for the acquisition of spectro-polarimetric data. This allows us to combine the measurement and the compression process into one consistent framework. Signals are recovered using a sparse reconstruction scheme from projections of the signal of interest onto appropriately chosen vectors, typically noise-like vectors. The compressibility properties of spectral lines are analyzed in detail. The results shown in this paper demonstrate that, thanks to the compressibility properties of spectral lines, it is feasible to reconstruct the signals using only a small fraction of the information that is measured nowadays. We investigate in depth the quality of the reconstruction as a function of the amount of data measured and the influence of noise. This change of paradigm also allows us to define new instrumental strategies and to propose modifications to existing instruments in order to take advantage of compressive sensing techniques.

Key words. techniques: spectroscopic – techniques: polarimetric – methods: observational – magnetic fields

1. Introduction

Our present knowledge of the physical and magnetic properties of solar and stellar plasmas has benefited from the rapid development of spectro-polarimeters in the last decades. These instruments use dispersive elements for spectral analysis. Because visible/infrared detectors are only sensitive to the intensity of light, modulators are used to encode the polarimetric information on intensity variations. Since most detectors are nowadays two-dimensional, the recovery of two-dimensional spectro-polarimetric information is carried out using scanning techniques: spatial in the case of long-slit spectro-polarimeters and spectral in the case of Fabry-Perot-like spectro-polarimeters. According to the Nyquist-Shannon sampling theorem (Nyquist 1928; Shannon & Weaver 1949), the correct sampling of a band-limited signal should be done at a rate equal to twice the bandwidth. In other words, one should sample each resolution element (spectral and/or spatial) with at least two pixels in order to be sure that all the frequencies in the bandwidth can be observed. This theorem has been applied for the last half century but it is only valid for band-limited signals.

During the last few years, the emerging theory of compressive sensing (CS; Candès et al. 2006b; Donoho 2006) has shown that this sampling is indeed too restrictive when some details of the signal structure are known in advance. An interesting point of the new CS paradigm is that, in many instances, natural signals, specifically for space telescopes but also for ground-based telescopes. Through the use of these techniques, we anticipate account in the measurement step, and not only in the post-analysis, thus leading to efficient measurement protocols. Instead of measuring the full signal (wavelength variation of the Stokes profiles in our case), under the CS framework one measures a few linear projections of the signal along some vectors known in advance and reconstructs the signal solving a non-linear problem.

A review of the literature shows us that signals arising in natural phenomena are typically compressible (see JPEG1 compression and wavelet, curvelet or ridgelet decompositions of images, among many others). Since this is also the case for astronomical data (e.g., Mühlmann & Hanslmeier 1996; Flegge & Solanki 1997; Belmon et al. 2002; Polygiannakis et al. 2003; Dollet et al. 2004; Bernas et al. 2004), it has been suggested that CS can be used to alleviate telemetry problems with space telescopes like Herschel (Bobin et al. 2008) or Cassini (Belmont et al. 2002) and to improve existing techniques for the reconstruction of radio interferometric data (Wiaux et al. 2009). Several works demonstrate that this is also true in the field of polarimetry (Socas-Navarro et al. 2001; López Ariste & Casini 2002; Skumanich & López Ariste 2002; Casini et al. 2005). Following this idea, Lites et al. (2002) analyzed the effectiveness of using JPEG compression to reduce the amount of data that needs to be transferred through telemetry for the Hinode space telescope. This option is currently in use in the mission (see Tsuneta et al. 2008). All these results and those of the analysis we carry out in this paper support the interest in investigating the appropriateness of employing CS ideas to measure spectro-polarimetric signals, specifically for space telescopes but also for ground-based telescopes. Through the use of these techniques, we anticipate

---

1 The name JPEG stands for Joint Photographic Experts Group, and is a loss-containing compression format for images based on the application of sparsity-enhancing linear transformations.
an enhanced performance in terms of de-noising and data acquisition rates which eventually may have an impact on the choice of detector technologies and data transfer.

The outline of the paper is the following. Section 2 gives a brief description of the CS paradigm, showing how signals can be recovered from a few linear projections and the properties that such projections need to fulfill. An analysis of the compressibility of spectro-polarimetric signals of interest is shown in Sect. 3. Section 4 presents recovery examples, with an analysis of the influence of noise. Section 5 shows novel instrumental strategies based on CS ideas, while the conclusions are presented in Sect. 6.

2. Compressive sensing

2.1. Theoretical considerations

Because of the innovative character of CS, we give a brief description of the most important points, while more mathematical details are discussed in Appendix A. For a more in-depth description, we refer the reader to recent references (e.g., Baraniuk 2007; Candès & Wakin 2008, and references therein).

The usage of compressive sensing techniques for the measurement of a signal \( x \) represented as a vector of length \( M \) is based on the following two key ideas:

- Instead of measuring the signal itself, one measures the scalar product of the signal with carefully selected vectors:
  \[
  y = \Phi x' + e, \tag{1}
  \]
  where \( y \) is the vector of measurements of dimension \( N \), \( \Phi \) is an \( N \times M \) sensing matrix and \( e \) is a vector of dimension \( N \) that characterizes the noise on the measurement process. Note that the previous equation describes the most general linear multiplexing scheme in which the number of measurements \( M \) and the length of the signal \( N \) may differ. In the standard multiplexing case, the number of scalar products measured equals the dimension of the signal (\( N = M \)). Consequently, it is possible to recover the vector \( x' \) provided that rank(\( \Phi \)) = \( N \), so that the problem is not ill-conditioned. In other words, one has to verify that every row of the \( \Phi \) matrix is orthogonal with respect to every other row.

- The second key ingredient of CS is the assumption that the signal of interest is sparse in a certain basis set (or can be efficiently compressed in this basis set). Any compressible signal\(^2\) can be written, in general, as:
  \[
  x' = W^T x, \tag{2}
  \]
  where \( x \) is a \( K \)-sparse\(^3\) vector of size \( M \) and \( W^T \) is the transpose of an \( M \times M \) transformation matrix associated with the basis set in which the signal is sparse. For instance, \( W \) can be the Fourier matrix if the signal \( x' \) is the combination of a few sinusoidal components. Other transformations of interest are the wavelet matrices or even empirical transformation matrices like those found using principal component analysis.

The combination of the those ingredients leads to the multiplexing scheme:

\[
\begin{equation}
\mathbf{y} = \Phi \mathbf{W}^T \mathbf{x} + \mathbf{e}, \tag{3}
\end{equation}
\]

\(^2\) A signal is said to be compressible (or quasi-sparse) if it is possible to find a basis for which the projection coefficients along the vectors of the basis reordered in decreasing magnitude decay in absolute value like a power-law.

\(^3\) A vector is \( K \)-sparse if only \( K \) elements of the vector are different from zero.

\[
\begin{align*}
\mathrm{min}\ ||x||_1 \quad & \text{subject to } ||y - \Phi \mathbf{W}^T \mathbf{x}||_2 < \epsilon, \tag{4} \\
\mathrm{min}\ ||x||_1 \quad & \text{subject to } ||y - \Phi \mathbf{W}^T \mathbf{x}||_2 < \epsilon \tag{5}
\end{align*}
\]

with the hypothesis that \( x \) is sparse, which renders CS feasible. It has been demonstrated by Candès et al. (2006b) that, even if \( \mathrm{rank}(\Phi \mathbf{W}^T) \ll N \) (we have many fewer equations than unknowns), the signal \( x \) can be recovered with overwhelming probability when using appropriately chosen sensing matrices \( \Phi \).

When the number of equations is less than the number of unknowns, it is usual to solve Eq. (3) using least-squares methods that try to minimize the \( \ell_2 \) norm\(^4\) of the residual. This is usually accomplished using techniques based on the singular value decomposition (see, e.g., Press et al. 1986). However, such minimization is known to return non-sparse results (e.g., Romberg 2008, and references therein). A more appropriate solution to Eq. (3) is to look for the vector with the smallest \( \ell_0 \) pseudo-norm (the number of non-zero elements of the vector) that fulfills the equation (Candès et al. 2006b):

\[
\begin{align*}
\mathrm{min}\ ||x||_1 \quad & \text{subject to } ||y - \Phi \mathbf{W}^T \mathbf{x}||_2 < \epsilon, \tag{4} \\
\mathrm{min}\ ||x||_1 \quad & \text{subject to } ||y - \Phi \mathbf{W}^T \mathbf{x}||_2 < \epsilon
\end{align*}
\]

\(^4\) The \( \ell_0 \) norm of a vector is given by \( ||x||_0 = (\sum |x|)^{1/n} \) if \( n > 0 \). The \( \ell_0 \) pseudo-norm is given by the number of non-zero elements of \( x \).

Fig. 1. Test showing how compressive sensing works to detect a sparse signal. The upper panel shows a sparse signal made of four spikes of very short duration. The lower panel presents 24 measurements built as the scalar product of the signal with Gaussian random noise. The stars show the reconstructed signal, showing that perfect recovery is possible with only a very small fraction of the measurements.
2.2. Practical considerations

The CS framework offers several advantages over standard measurement paradigms. On the one hand, since one only measures linear combinations of the signals, the flux of information that any sensor has to deal with is usually much smaller (thanks to compression). This is probably of secondary interest for ground-based instruments since the infrastructure to cope with such large fluxes of data is available. However, this could be of interest for space-borne instruments, where the flux of data is limited by telemetry. As a sub-product of the simplification of the measurement, the reconstruction of the signal is much more time consuming, but can be done efficiently a-posteriori without affecting the measurement process. On the other hand, if appropriate sensing matrices are chosen, the measurement process can be considered universal and does not depend on the exact basis set in which the data is sparse. In other words, one first measures projections and this assures that the data can be reconstructed a-posteriori if the data is sparse in any (a priori unknown) basis set.

Ground-based instruments may draw advantages from the increased cadence at which data is acquired. An instrument measuring spatial and spectral information with a 2D detector is forced to scan one of the extra dimensions of the data space, either the spectra in filter instruments or one of the spatial dimensions in spectrograph instruments. The CS framework may diminish the number of measurements in the spectrum space thus shortening the time in which the 3D data cube is acquired by the instruments. Boosting data acquisition rates is always advanta-
gerous together with a strong signal produced by the D2 line of Ba II. The spectrum is shown in black solid lines in Fig. 2. The full spectrum was wavelet-decomposed (the wavelet of choice is shown in each panel) and thresholded so that only a certain number of wavelet coefficients survive, while the rest of coefficients are set to zero. This is an efficient way of compressing the signal provided that the thresholding fundamentally cancels noise and leaves the signal unperturbed. The upper left panel of Fig. 2 shows what happens when only 10% of the wavelet coefficients are maintained, while the right panel indicates the behavior after setting 98% of the coefficients to zero. Since the zero coefficients are not necessary in the reconstruction, this thresholding leads to a large compression of the signal. The signal is then reconstructed using the inverse wavelet transform. We note that even in the case of only 2% of the coefficients, the important signals are recovered while the noisy part of the spectrum is largely reduced. It is apparent from the figure that the behavior is very similar for all the wavelet families we have tested, although the computing times are different, being larger for wavelets with a larger number of non-vanishing moments (e.g., Jensen & la Cour-Harbo 2001): the Daubechies standard wavelet mother functions that are widely used in other applications (e.g., Starck et al. 1997; Fligge & Solanki 1997) or Hermite functions (del Toro Iniesta & López Ariste 2003).

We carried out an experiment to characterize the compressibility of the second solar spectrum. We selected a large piece of the spectrum in which many small signals are present, together with a strong signal produced by the D2 line of Ba II. The spectrum is shown in black solid lines in Fig. 2. The full spectrum was wavelet-decomposed (the wavelet of choice is shown in each panel) and thresholded so that only a certain number of wavelet coefficients survive, while the rest of coefficients are set to zero. This is an efficient way of compressing the signal provided that the thresholding fundamentally cancels noise and leaves the signal unperturbed. The upper left panel of Fig. 2 shows what happens when only 10% of the wavelet coefficients are maintained, while the right panel indicates the behavior after setting 98% of the coefficients to zero. Since the zero coefficients are not necessary in the reconstruction, this thresholding leads to a large compression of the signal. The signal is then reconstructed using the inverse wavelet transform. We note that even in the case of only 2% of the coefficients, the important signals are recovered while the noisy part of the spectrum is largely reduced. It is apparent from the figure that the behavior is very similar for all the wavelet families we have tested, although the computing times are different, being larger for wavelets with a larger number of non-vanishing moments (e.g., Jensen & la Cour-Harbo 2001). This can be an issue that should be taken into account depending on the balance between the desired smoothness of the reconstruction and the computing time.

Other examples are shown in the middle and lower panels of Fig. 2, for the case of the Na I doublet at 5890 Å and the Sr I line at 4607 Å, respectively. The first one presents a case in which low-frequency (the large scale quantum interference between the two lines of the doublet) and high-frequency information (the large variability of the profiles close to the core of the line) coexist. This poses an interesting problem to any compression method because it has to retain low and high frequencies simultaneously. Apparently, the wavelet compression does a good job on this multiplet and all important details can be retrieved even with only 2% of the coefficients. The lower panels of Fig. 2 show the case of the Sr I line at 4607 Å, which is a very strong signal embedded in a quasi-flat continuum. In this case, reconstructing only with 2% of the coefficients gives a poor representation of the true underlying signal. Some ripples appear when using Daubechies and Coiflet wavelets on the quasi-continuum, although the amplitude of the signal is still correctly recovered. The reconstruction with the Haar wavelet gives a very good representation of the Sr I line but the depolarizations in the red wing of the line and at 4606.3 Å are not correctly recovered. The reconstruction with 10% of the coefficients is almost perfect.
Fig. 2. Test showing the reconstruction of three widely known domains of the second solar spectrum: the top panels show a region around the Ba II D2 line, the middle panels a region around the Na I doublet at 5890 Å while the lower panel presents the region around the Sr I line at 4607 Å. Reconstruction using different wavelets and different thresholds are displayed. In each panel, the left panel presents how the reconstructed spectrum (red line) compares with the original spectrum using Daubechies-4, Daubechies-8, Coiflet-3 and Haar wavelets when 90% of the wavelet coefficients are set to zero. The right panels show the results when only 2% of the wavelet coefficients are retained. These results demonstrate that the structure of the lines is recovered with a small number of wavelet coefficients. Only in the case of the Sr I do we find spurious ripples (except in the Haar case) due to the loss of information.

We have measured the quality of the reconstruction using the 68% and 95% percentiles of the difference between the original and the reconstructed signal. The value of these quantities versus the percentage of remaining non-zero coefficients is shown in Fig. 3. The left panel has been obtained with the data from the Ba II D2 line, while the right panel is associated with the Na I data. We have verified that the distribution of differences is close to normal except in the cases in which the reconstruction
Fig. 3. Percentiles 68% (lines) and 95% (lines+symbols) of the differences between the true signal and the reconstructed signal when retaining a different percentage of the wavelet coefficients. The left panel corresponds to the data close to the Ba II D₂ line, while the right panel shows the reconstruction of the Na I doublet.

Fig. 4. Percentile 68% (lines) and 95% (lines+symbols) of the differences between the true signal and the reconstructed signal for a Hinode observation. The left panel shows the case using a Daubechies-8 wavelet, the right panel presents the results using the PCA empirical basis set. The noise level in Stokes Q, U and V is close to $1.6 \times 10^{-3}$ in units of the continuum intensity. Therefore, less than 5% of the PCA eigenvectors are needed in order to reconstruct the profiles to this noise level.

is done with too few coefficients. Therefore, the 68% and 95% percentiles are close to the standard deviation and twice the standard deviation, respectively. The lines without symbols present the 68% percentile, showing that differences are in both cases below $10^{-2}$ when retaining just 10% of the profiles. For $Q/I$ signals that are on average at the level of $\sim 0.1$, we find that relative errors are typically below $10^{-3}$. The 95% percentile gives relative errors slightly above $10^{-3}$ for these cases. The results tend to indicate that differences among wavelet families are relatively small.

3.2. Sparsity of Zeeman signals

In the presence of a sufficiently strong magnetic field, the Zeeman effect usually controls the emergent observed polarization. In order to test for the compressibility properties in the Zeeman-dominated case, we use a dataset obtained with the Solar Optical Telescope/Spectro-Polarimeter SOT/SP (Lites et al. 2001) aboard Hinode (Kosugi et al. 2007). The properties of this dataset are typical of what one should encounter in the future if the CS techniques that we propose here are applied to future space missions. In this case, we test two different basis set for compressibility in Fig. 4. The first is the universal Daubechies-8 wavelet (left panel). The results we show are not very sensitive to the specific chosen wavelet and are representative of the general behavior. The second is the empirical basis set obtained using the PCA decomposition (right panel).

The results, that present the 68% and 95% percentiles of the distribution of the difference between the exact and the reconstructed profiles, indicate clearly that signals are again compressible. The best results are, obviously, obtained with the PCA basis set, because the eigenvectors are empirically constructed to maximize the sparsity of the signal (only a few eigenvectors are necessary to reconstruct the signal without noise).

The main problem with the PCA basis set is that it is obtained empirically. Consequently, strictly speaking, one is not able to use the PCA basis set in a CS framework because a-priori the basis set is not known. However, according to Skumanich & López Ariste (2002), some universality properties of the PCA eigenvectors can be demonstrated when many profiles are included in a database. For this reason, we have tested that profiles observed with Hinode can be well compressed with the eigenvectors...
4. Signal recovery

4.1. Examples

Since we have demonstrated that the polarimetric signals are compressible, the CS framework can be used to measure such signals. We give here a few examples using different sensing matrices. The first ones are shown in the left panel of Fig. 5. We present the reconstruction of the Na I and Sr I signals analyzed in Sect. 3 with two different sensing matrices: (i) a Gaussian matrix with elements extracted from the $N(0, 1/K)$ distribution (Candès et al. 2006b), with $K$ being the sparsity of the signal; and (ii) a binary sparse sensing matrix with only 10 non-zero elements per measurement (Berinde & Indyk 2008). Obviously, the binary sparse matrix has two advantages over the Gaussian matrix. First, the number of non-zero elements is very small as compared to the size of the matrix and efficient sparse storage and computational methods can be used (e.g., Press et al. 1986). In our case, the sparse matrix contains less than 2% of the elements different from zero. Second, the binary matrix is easier to implement on hardware using, for instance, micro-mirrors. Only 5% of the elements of the solution vector are allowed to be non-zero for the case of Na I and 10% for the case of Sr I, according to the results presented in Fig. 2. The number of measurements used is 6 K for the Sr I line and 8 K for the Na I doublet, roughly in accordance with Eq. (A.6), while the reconstruction is done using the Daubechies-8 wavelet. The results show that a good recovery is possible in the two cases, with the advantage that, since sparsity is inherent to the reconstruction, noise is largely reduced in the reconstruction. In order to show how the technique behaves with the number of measurements, we show in Fig. 6 the standard deviation of the difference between the reconstructed and original signal versus the number of measurements (normalized to the sparsity of the vector). The horizontal lines indicate an estimation of the noise level in the observations obtained as the standard deviation of a portion of the continuum.
Note that when the number of measurements is not large enough, the reconstruction does not work. However, as soon as condition (A.6) is fulfilled, reconstruction works properly.

Other examples are shown on the right panel of Fig. 5, the Stokes $V$ profiles picked at two positions of an observation carried out with Hinode on February 27, 2007. The Fe i doublet at 630 nm with amplitudes typical of active regions (upper panel) and quiet Sun (lower panel) are shown in black lines. The reconstructed signals using the same sensing matrices as above are shown in red and blue. The universal PCA basis is used and only 11 of such eigenvectors are used, roughly 10% of the full basis set. The number of measurements is 6 K, in accordance with Eq. (A.6). A good recovery is possible even for profiles whose amplitude is close to the noise level. The prior information encoded in the sparse reconstruction produces that noise is slightly reduced with respect to the original profile. This is similar to what one would find after carrying out a PCA filtering of the data, but the filtering is encoded inside the measurement technique (e.g., Martínez González et al. 2008b). The fundamental reason for this is that, while true signals produce sparse signatures, noise destroys the sparsity to some degree. Since the reconstruction is done enhancing sparsity, it is not possible to recover noise and a filtering is carried out as a side effect of the reconstruction.

### 4.2. Sensitivity to noise

To test the influence of noise on our ability to recover the spectrum from the linear combinations given by the sensing matrix, we analyze a synthetic case. We have chosen the spectral line at 6302 Å for its widespread use. A very simplistic Stokes $I$ profile is built using a Voigt function tweaking the width and depth to fit the average profile of the solar atlas (Wallace et al. 1998). The Stokes $V$ profile emerging from a magnetized atmosphere is built under the weak-field approximation, in which it is proportional to the wavelength derivative of Stokes $I$ (e.g., Landi Degl’Innocenti & Landolfi 2004). We used a magnetic flux density of 1 Mx/cm$^2$, resulting in an amplitude of $4 \times 10^{-4}$ in units of the continuum intensity. The small value of the magnetic flux density is used intentionally so that the observation is close to the detection limit of present spectro-polarimeters with relatively short exposure times. We simulate observations with noise added and we use a Gaussian random sensing matrix with zero mean and inverse variance equal to the assumed sparsity. We carry out experiments, shown in Fig. 7, using Daubechies-8 wavelets assuming that only 12 elements of the recovered vector are non-zero (left panel) and a universal PCA basis set assuming that the sparsity of the vector is 5 (right panel). The PCA basis is obtained from a dataset observed with Hinode. The number of measurements is set to six times the sparsity level in each case. The solid line in each plot indicates the standard deviation of the difference between the exact profile and the reconstructed one for each value of the noise level. The dashed line is the same quantity but calculated only for the noise. This curve is used to give an idea of the de-noising abilities of the decomposition on a wavelet/PCA basis. When the signal-to-noise ratio (SNR) is very poor, in both experiments we see a reduction of almost an order of magnitude in the noise, showing that the signal is well below the noise level. Only when the SNR approaches ∼0.4 do we see that the CS detects signal and gives a much better behavior than just the direct measure of the profile using standard techniques. The reason lies in the fact that sparsity is promoting the detection of signals, contrary to the detection of noise. Noise is not sparse in any of the basis sets used and the reconstruction is made assuming sparsity in the solution. This is important information that we are including as a prior in the CS recovery, something that is not done in standard measurements.

### 4.3. CS with polarimetric modulation

Since existing instruments are not directly sensitive to polarization in the optical and infrared spectral domains, it is customary to use modulation schemes to measure the Stokes parameters. In some sense, modulation is another form of multiplexing which is carried out using retarders and polarizers. The monochromatic Stokes vector entering the telescope $S$ is modulated $N_{mod}$ times to generate intensities that are linear combinations of the Stokes parameters of $S$:

$$I_{out} = O'S,$$

(6)

where each row of the $N_{mod} \times 4$ matrix $O$ equals the first row of the Mueller matrix of each modulation state.

The obvious question is whether it is possible to apply CS to compress the wavelength information while still carrying out the modulation to detect the four Stokes parameters. The answer is that it is possible since both multiplexing operations “act” on different spaces: polarimetric modulation applies to monochromatic Stokes parameters and the sensing matrix of CS applies to the wavelength variation of a single Stokes parameter. Therefore, under the assumption that the polarimetric modulation is achromatic over the observed spectral range, their effect can be interchanged and one ends up solving four problems of the kind:

$$y_i = O W^T x_i + e,$$

(7)

where $x_i$ is the demodulated wavelength variation of the $i$th Stokes parameter (del Toro Iniesta & Collados 2000) and we have assumed that the same CS matrix is used for all the modulation states of the polarimeter. Consequently, the procedure to follow is to obtain the $N$ linear measurements for each position of the polarimeter, thus leading to a set of $N N_{mod}$ measurements. Then, following del Toro Iniesta & Collados (2000), the pseudo-inverse of the modulation matrix $O$ is applied to the $N_{mod}$ modulation states of each linear measurements. At the end, the four CS problems of Eq. (7) are solved.
4.4. Fringes and other spurious signals

Among others, non-polarized and polarized fringes are undesirable contaminations present in many spectro-polarimetric observations (e.g., Semel 2003). Observations are partially cleaned of these fringes using flat-fielding techniques. The remaining fringes are filtered out at the end of the reduction process with the disadvantage of having a large subjective component. Under the compressive sensing scheme, these spurious signals are measured together with the real signal. In the reconstruction we try to avoid introducing them in the final result, be it by rejecting or reconstructing them together with the true signal. Obviously, the ideal situation is to employ a fringe-free spectro-polarimeter to obtain a better reconstruction. We defer the deep investigation of this issue to a later study. However, preliminary experiments indicate that the reconstruction can efficiently reduce the amplitude of periodic fringes if the basis set used is not able to reproduce them. The drawback is that the ensuing reconstruction is less accurate than in the absence of fringes. Such a situation arises when applying a PCA basis set that is able to efficiently describe the spectro-polarimetric signals but not the periodic fringes. Another possibility of investigation is to assume that the fringes are sparse in the Fourier domain, carrying out the reconstruction merging the Fourier basis set and the basis set for the signal.

5. Applications

5.1. Efficient spectro-imagers

Recent interest in Hadamard techniques for spectro-polarimetry (see Harwit & Sloane 1979, for more details) can greatly benefit from CS techniques. Hadamard techniques allow us to condense spectral information inside single detector pixels through multiplexing. Before going into the use of these techniques in the framework of CS, it is advisable to describe these Hadamard techniques in spectro-polarimetry and the advantages they bring through two illustrative examples. In traditional spectroscopy, a certain number of pixels (often a full dimension of a detector array) are dedicated to measure intensities at different wavelengths for the same point in an image. In Hadamard techniques, this is substituted by a temporal modulation over the Hadamard cyclic mask in a single detector pixel. Such an exchange has two interesting applications: in long-slit spectroscopy, the spectrum can be multiplexed in a single pixel and instead of using a 2-dimensional detector array one can use a one dimensional array with increased acquisition cadences that allow us to freeze the seeing during the modulation cycle of polarimetry with the resulting improvement on polarimetric sensitivity (see, e.g., ZIMPOL; Povel 2001).

In double-pass subtractive spectroscopy (Mein 2002), the resulting image has been filtered by a narrow spectral slit that selects a single wavelength per pixel. Due to the dispersion of the first pass over the diffraction grating, the selected wavelength changes as one moves over the spatial image. To reconstruct the full 3D data cube with both spatial dimensions and spectral covering, every pixel has to be scanned over a range of wavelengths. Through the use of Hadamard cyclic matrices one can have several wavelengths sampled simultaneously in every pixel. As a result the temporal coherence of the recovered spectra is increased as several wavelengths over the spectral domain are detected simultaneously. Also, as a side effect, the raw images do not show any evident spectral features, which makes them more suitable for image reconstruction techniques.

Such applications of Hadamard techniques to spectro-polarimetry are however hindered by the known fact that, in the presence of a multiplicative noise such as photonic noise, the Hadamard transformation results in a reduction of SNRs with respect to the case of equivalent exposures without multiplexing (Harwit & Sloane 1979). The reduction in the SNR can be limited with the use of appropriate binary masks (Wuttig 2005), although it is never too large for the usual cases in solar spectro-polarimetry.

Compressive sensing can help mitigate the problem. Since the Hadamard technique is applied to the spectral information, one can make use of the fact that there is prior information about the spectrum to be measured and that, as a consequence, the space of spectral profiles (with polarization included) is sparse, as demonstrated in the previous sections of this work. The use of the CS techniques illustrated above allows the recovery of the full spectral information with just a few spectral measurements. In the terminology of Hadamard techniques, this
If each exposure lasts for $t$ signal (perhaps the optimal value the number of CS measurements but it su-
not very large. An example of this is shown in Fig. 8. A
for our aims, since we are typically interested in cases where
From the tests of previous sections we can conclude that
With a factor $M$ one can imagine a spectrograph able to measure the wave-
When prior information about the expected signals is available it
5.2. Sub-Nyquist spectrograph
translates into the fact that not all the data acquisitions attached
to the cyclic Hadamard masks are required for the recovery of the spec-
If traditionally a Hadamard mask of dimension $N$
would require $N$ cyclic measurements to solve the multiplexing
linear system, CS techniques may be used to solve the system
is possible to imagine a spectrograph able to measure the wave-
length variation of the Stokes parameters using resolution ele-
ments (pixels in the camera) larger than the spectral sampling. In
is defined in the frequency domain. This is the case of signals
for which the power associated with frequencies above a certain
threshold is associated with noise. In such a case, one can use
this new threshold, using the Nyquist-Shannon theorem, to estimate the number of pixels per resolution element.
Reconstruction should be done using the appropriate prior in-
formation. We have verified, although not shown here, that good
reconstructions can be obtained using 1/2 and 1/4 of the original
information by employing the Fourier basis set and forcing the
signal in the Fourier domain to be sparse. This has the advantage
over PCA eigenvectors that they are fully universal.

5.3. Hadamard-PCA magnetometer
The combination of CS ideas and standard PCA techniques can be
applied to develop an efficient magnetometer. To this end, let
us assume that the Stokes $V$ profile is well represented using the
first PCA eigenvector $v_1(\lambda)$ obtained empirically from a set of
previous observations:

$$V(\lambda) \approx A v_1(\lambda).$$

This equation assumes, therefore, that the observed Stokes pro-
files are 1-sparse in the PCA basis set, so that one can recover the
signal, according to the CS theory, using of the order of 4–6 mea-
measurements. Using a sensing matrix $\Phi$, we end up with the follow-
ing measurement process:

$$y_i = \sum_j \Phi_{ij} V(\lambda_j), \quad \forall i = 1, \ldots, N_{\text{meas}}.$$

If the sparsity constraint is used, the proportionality constant $A$
can be obtained from the observations with a linear fit:

$$A \approx \frac{\sum_i y_i f_i}{\sum_i f_i^2},$$

where $f_i = \sum_j \Phi_{ij} v_1(\lambda_j)$. Note that a similar result could have
been obtained if we measured projections of the data over the
PCA eigenvectors. However, this presents the difficulty of its
practical implementation because the PCA eigenvectors con-
tain negative values. Measuring with sensing matrices like a
binary (1/0) Hadamard matrix that is incoherent with the PCA
eigenvectors leads to a universal sensing process. We point out
that only one measurement gives a rough estimation of the mag-
etometer signal, although it should be used with care. An exam-
ple of the capabilities of the method is shown in Fig. 9. A scan
of a sunspot observed on 27 February 2007 with the SOT/SP
onboard Hinode has been used. The vertical axis presents the
value of the Stokes $V$ profile at a fixed wavelength corresponding
to the blue $c_r$ component of the 6302.5 Å line. The horizontal
axis presents the value inferred from the CS measurements.
The black dots show the scatter when only one measurement is made,
while the red dots show what happens with 8 measurements.
If enough measurements are taken, it is possible to include more eigenvectors in the decomposition of Eq. (9) and obtain information about the projection along them from the observations. If Stokes $I$ is decomposed, continuum images and velocities can be inferred from the first and second eigenvectors, respectively. Likewise, if Stokes $V$ is used, magnetic flux and magnetic velocities can be inferred.

6. Conclusions

This paper demonstrates the feasibility of applying compressive sensing techniques to measure the wavelength variation of the Stokes parameters observed in stellar atmospheres. We have shown that spectro-polarimetric signals are, in general, compressible on universal basis sets. However, in general, it is more advantageous to use empirical basis sets like that obtained from principal component analysis because data can be more efficiently reproduced on such a basis. We stress that the results presented in this paper are extensible to standard spectroscopic observations and not only to linear and circular polarization profiles, so that any day-time and night-time spectrograph can take advantage of these techniques.

According to our results, it is possible to measure Stokes parameters using less than a half of the measurements one would need when strictly applying the Nyquist-Shannon sampling theorem. Compressive sensing leads to several interesting effects. Since fewer measurements are made, an inherent reduction in the exposure time is present. Such a reduction can be used to do more measurements in the same total time, thus allowing less noisy observations. The a-priori information encoded in the sparsity condition results in the fact that the reconstructed signal is much less noisy than one would expect. The reason is that filtering is applied simultaneously while measuring. For instance, several of the examples shown in this paper produce signals that are automatically filtered with the principal component analysis because data can be more efficiently developed.

We have proposed potential applications of CS to the field of spectro-polarimetry, testing them numerically on real data. Some of these techniques can be straightforwardly applied to existing instruments, while other proposals need more profound modifications. The future of observational spectro-polarimetry, at least in solar physics, has to be rooted in the development of two-dimensional spectro-polarimeters. Since detectors are only two-dimensional at the moment, scanning schemes have to be used. We consider that double-pass subtractive spectroscopy constitutes a very appealing technique for two-dimensional spectro-polarimetry if combined with multiplexing techniques using Hadamard masks. Compressive sensing will help significantly reduce the total exposure time, thus allowing an increase in the final SNR for a fixed integration time.

Because of the natural physical interpretation of projections of the observed Stokes profiles along PCA eigenvectors (Skumanich & López Ariste 2002), it would be desirable to directly measure such projections. Using compressive sensing techniques, we have proposed a technique that, thanks to Hadamard masks, is able to retrieve such projections from a universal multiplexing. The advantage is that this multiplexing is binary and easy to build.

Finally, we analyze the plausibility of a spectro-polarimeter that does not fulfill the Nyquist-Shannon sampling theorem. Some a-priori information about the expected signals is available and compressive sensing techniques can take full advantage of this information to recover the signals from a reduced set of measurements. We show with an example that it is possible to reconstruct Stokes profiles using the information obtained from adding the signal in consecutive pixels. This can be understood as a super-resolution scheme in which one knows the basis set in which the high-resolution signal can be efficiently developed.

Acknowledgements. We thank Rafael Manso Sainz and María Jesús Martínez González for illuminating discussions and carefully reading the manuscript. Financial support by the Spanish Ministry of Science and Innovation through project AYA2007-63881 is gratefully acknowledged.

Appendix A: Compressive sensing

As noted in the main text, the multiplexing scheme for a sparse signal reads:

$$y = \Phi W^T x + \epsilon,$$

(A.1)

with the condition that $x$ is sparse. When the number of measurements is much smaller than the size of the signal, the sparsest solution fulfills:

$$\min_x \| x \|_0 \text{ subject to } \| y - \Phi W^T x \|_2 < \epsilon,$$

(A.2)

where $\| x \|_0$ is the $\ell_0$ pseudo-norm that equals the number of non-zero elements of the vector $x$. Since solving this problem is, in general, not feasible, it has been demonstrated by Candès et al. (2006b,a) that, if the matrix $\Phi W^T$ fulfills the restricted isometry property (RIP; Candès et al. 2006a), the solution to the problem

$$\min_x \| x \|_1 \text{ subject to } \| y - \Phi W^T x \|_2 < \epsilon,$$

(A.3)

is equivalent to that of Eq. (A.2). The RIP condition is a sufficient condition and it is often too restrictive. The advantage of the last problem is that it can be easily solved using linear programming techniques.

Intuitively, the RIP condition states that the action of the $\Phi W^T$ operator on the sparse vector $x$ does not excessively modify its $\ell_2$ norm. Mathematically:

$$(1 - \delta_K) \| x \|_2^2 \leq \| \Phi W^T x \|_2^2 \leq (1 + \delta_K) \| x \|_2^2$$

(A.4)

for all $K$-sparse vectors $x$ and $\delta_K < 1$.

In spite of the mathematical importance of the RIP condition, it is more intuitive to think in terms of coherence between the sensing matrix and the transformation matrix. In general, for
a sensing matrix to be considered good, it should be as incoherent as possible with the transformation matrix. Every row of the sensing matrix should be able to obtain as much information as possible from the sparse vector \(x\) in order to facilitate its reconstruction with as few measurements as possible. This is achieved when the sensing matrix and the transformation matrix are as incoherent as possible. The coherence between the two matrices is defined as (Candès & Romberg 2007):

\[
\mu(\Phi, W) = \max_{\phi \in \Phi, w \in W} \|\phi, w\|,
\]

where \(\phi\) and \(w\) are, respectively, columns and rows of the matrices \(\Phi\) and \(W\). Generally (Candès & Romberg 2007), for a sensing matrix of size \(N \times M\), it is possible to recover an \(s\)-sparse vector using a number of linear combinations that fulfills:

\[
M \geq C \mu(\Phi, W)^2 K \log N,
\]

where \(C\) is a constant of order 1. According to this equation, if one is able to find sensing matrices with small coherence with respect to the basis set of interest, one should be able to recover the sparse signal using a number of measurements that is proportional to \(K \log N\). As we have shown in the main text, the proportionality constant is typically between 4 and 6.

### A.1. Recovery algorithm

For the recovery problem in our experiments, different methods have been developed during the last few years. After testing several methods, we found that the recent algorithm presented by Blumensath & Davies (2008) is very efficient in terms of computing time and shows state of the art performance. The method uses the simple iterative procedure:

\[
x^{t+1} = H_t \left[ x^t + \mu W \Phi^T (y - \Phi W^T x^t) \right],
\]

where \(H_t(\cdot)\) is a non-linear thresholding operator that leaves as non-zero the \(s\) elements of the vector \(t\) with the largest absolute value, setting to zero the rest of the elements. The method is guaranteed to be stable thanks to the re-scaling quantity \(\mu\). Indeed, according to our experience, its stability is remarkable, converging to the solution in almost all experiments. The method is initialized by \(x^0 = 0\). The main drawback of the method (usually common to all recovery methods) is that the sparsity of the solution, \(s\), has to be chosen in advance. The advantage is that only multiplications with the matrices \(\Phi\) and \(W\) (and their transposes) are needed. Many sparsity-promoting basis sets are accompanied by fast multiplication algorithms (e.g., fast Fourier transform, fast wavelet transform, etc.). In such a case, the computing time of the multiplication with the \(W\) and \(W^T\) matrices scales as \(O(n)\) for the fast wavelet transform and as \(O(n \log n)\) for the fast Fourier transform, instead of scaling as \(O(n^2)\) like in a standard matrix-vector product.

### References


Candès, E., & Romberg, J. 2007, Inverse Problem, 3, 969

Candès, E., Romberg, J., & Tao, T. 2006a, IEEE Transactions on Information Theory, 52, 489


Donoho, D. 2006, IEEE Trans. on Information Theory, 52, 1289


Nyquist, H. 1928, Trans. AIEE, 41, 617


Wallace, L., Hinkle, K., & Livingston, W. 1998, An atlas of the spectrum of the solar photosphere from 13500 to 28000 cm⁻¹ (3570 to 7405 Å), ed. L. Wallace, K. Hinkle, & W. Livingston (Tucson: NWO)
