Polarimetry: From the Sun to Stars and Stellar Environments
Proceedings IAU Symposium No. 305, 2014
K. N. Nagendra, S. Bagnulo, R. Centeno &
M. J. Martínez González, eds.
© International Astronomical Union 2015
doi:10.1017/S1743921315004792

Pattern recognition techniques in Polarimetry

Arturo López Ariste

IRAP, Toulouse, France
email: Arturo.LopezAriste@irap.omp.eu

Abstract. Sparsity is a property of data by which it can be represented using a small number of patterns. It is the key concept behind an evergrowing list of mathematical techniques for handling data and recover from it signals or information in conditions previously thought impossible. The application of those techniques to spectropolarimetric data is relatively straightforward. We present three examples of such application: the use of Principal Component Analysis to invert the magnetic field in solar prominences from spectropolarimetry of the He D_3 line, the removal of fringes from spectropolarimetric data with Relevance Vector Machines, and the retrieval of high resolution spectra from low resolution data with Compressed Sensing.

Keywords. methods: data analysis; polarization

1. Introduction

We are witnessing a revolution in information technology spanning from simple aspects of everyday's life to our more or less specialized problems in astrophysics. The key to this revolution lays in the mathematical treatment and analysis of the available, and evergrowing, amount of, data to extract critical information. The bottom line is that data that contains information presents patterns, and the absence of patterns means that data is made of noise. Hence our ability to learn something new from the data relies on our ability to detect patterns. By *pattern* we mean correlations of one kind or another among the different elements (let's call them pixels) in the data set. These correlations can yet be understood in a different manner: the value of many pixels in the data set is not independent but depends on the value of the other pixels. Thus the dimension of the data set is not N, the number of pixels, but a much smaller number n with often $n \ll N$. Our data set is *sparse*. Sparsity is the mathematical property of datasets with information we are interested in. Sparsity is the key to extract that information and assess its validity.

While the previous introduction could have been applied to many different subjects on which these techniques are being succesfully applied nowadays, we are here concerned with applications in the observation and analysis of polarization in solar and stellar light. As a showcase we will show in this contribution three different applications of pattern recognition or sparsity-based techniques in the analysis of spectropolarimetric data: the inversion and inference of magnetic fields in solar prominences with Principal Component Analysis, the removal of fringes from polarized spectra with Relevance Vector Machines, and the zoom into larger spectral resolutions through Compressed Sensing.

2. Inversion of the He D_3 polarization spectra in solar prominences.

Principal Component Analysis (PCA) decomposes the data as a linear combination of just a few principal components. Given those principal components, the data can be

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

reproduced by just the few coefficients of the linear combination: the data is sparse in the basis of the principal components. PCA was first used to reproduce spectropolarimetric data by Rees *et al.* (2000), who demonstrated that less than 10 principal components sufficed to reproduce the Stokes profiles of Zeeman-sensitive atomic lines. These principal components were calculated by computing the eigenvectors (or singular vectors to be mathematically rigorous) of a correlation matrix built with a large set of Stokes profiles. Skumanich & López Ariste (2002) later showed that these few principal components captured the essence of the Stokes profiles: given an average atomic line in absorption over the continuum the two main shape changes are a Doppler shift and a broadening. If both are small enough to accept to be represented to first order, then it appears reasonable that most profiles can be reproduced as a linear combination of the average profile plus its first derivative (to describe spectral shifts) plus its second derivative (to describe broadenings). Skumanich & López Ariste (2002) showed that PCA did capture those 3 patterns as principal components.

The dimensionality reduction provided by the description of any individual Stokes profile as a set of a few coefficients (less than 10 in general, compared to spectra made of several hundred pixels) allowed Socas-Navarro *et al.* (2001) to propose an inversion code that used a lookup table of known cases represented by those PCA coefficients. The principal components form an algebraic basis and they are orthogonal, an Euclidean distance over the PCA coefficients allows to determine which particular model in the lookup table is the most similar to the one observed, and then propose the magnetic field of the nearest model in the lookup table as solution to the inverse problem of measuring magnetic fields. López Ariste & Casini (2002) applied the same idea to the measurement of magnetic fields in solar prominences from spectropolarimetry of the He D_3 line. Fig.1 shows in the left plots the 2 main principal component profiles of the 4 Stokes parameters of this line emitted in prominence conditions. In the right images it shows a recent example of a prominence observed with the telescope THEMIS and inverted with this code. The lookup table is made up of 90000 different cases, all of them computed by solving the statistical equilibrium equation including quantum coherences among levels of a He atom model under known thermodynamic and magnetic conditions. The computed profiles were stored in the lookup table using just 4 PCA coefficients. Each profile observed by THEMIS was then compared to this database to find the nearest model and map the corresponding magnetic field as solution in Fig. 1. In this inversion code sparsity is used as a means to describe the observations in a condensed manner without losing the information. Whatever information about prominence magnetic fields there is in the observed Stokes profiles is contained in those few PCA coefficients. The lookup table is filled with Stokes profiles computed for known magnetic and thermodynamic conditions, but these may or may not correspond to those present in solar prominences. We may encounter the situation in which we can reproduce the observed Stokes profiles with a small set of PCA coefficients but there is no similar set of coefficients in our lookup table: in the absence of a pre-known case we do not know what magnetic field is behind the observed profile. Pattern recognition thus condenses the information, but does not necessarily unveil it. A striking case appears when doing the statistics of the magnetic fields measured over 138 prominences observed by THEMIS in 2014. The histogram in Fig. 2 gives the inclination of the inferred magnetic field. There are two kinds of distributions in the figure: one is represented with blue bars, the other with red bars. The first case of blue bars corresponds to cases where the error bar is smaller than 10 deg. This error is considered to be the typical error of the inversion code arising from the typical signal-to-noise ratios of the data and the finite size of the lookup table. This histogram represents those observed profiles for which a good solution was found in the lookup

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

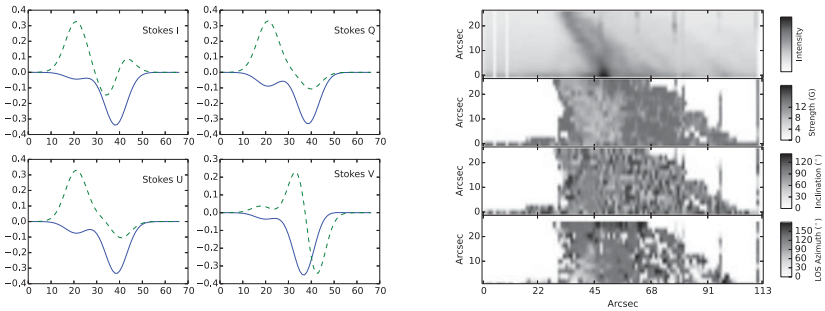

Figure 1. Left: First two eigenprofiles for the Stokes parameters of the He D_3 line in prominences. Right: Results of the inversion of the He I D_3 line by Principal Component Analysis on a prominence observed in September 2013 with the telescope THEMIS. From top to bottom line intensity, field strength, inclination in the local reference frame (90° being horizontal), and azimuth in the plane of the sky (adapted from Schmieder *et al.* 2014; see also Schmieder *et al.* 2015, this issue).

table. They are crowded around the 90° inclination mark with a standard deviation of 10°, as expected. The 90° inclination corresponds to horizontal fields in the prominence in line with what is expected from previous observations (Labrosse *et al.* 2010; López Ariste 2015) and theory (Aulanier & Démoulin, 1998; Mackay *et al.* 2010). The second distribution in red corresponds to cases where the error bar is larger than 30°, that is, well beyond the nominal error of the inversion code. These are cases for which no resembling profile was found in the lookup table. Even if not alike, there is always one profile that resembles more than the others and this is given as solution. The curious thing is that these solutions with large error bars cluster around inclinations of 60° and 120°. This means that the observed profiles had similar shapes among them and that generic shape found its best match in synthetic profiles with those inclinations. Those inclinations correspond nicely to the Van Vleck's angles around the vertical so this gives us a clue towards identifying what kind of magnetic fields may be at the origin of those particular profiles. Schmieder *et al.* (2014) identified a prominence region full of those unmatched profiles where imaging showed a highly dynamic plasma moving in almost any possible direction at high speeds. Could a turbulent field be the explanation? Not exactly, since a turbulent field would show the absence of signal in Stokes U, which was not the case. But we could try a model in which a horizontal field shared the (spatial and temporal) resolution pixel with a turbulent field with different fill fractions. How would the PCA code interpret the Stokes profiles emitted by such a model? The answer in Fig. 2 shows inversions with large error bars clustering at inclinations of 60° and 90° as the red histogram from observations. This does not mean that the observed prominences respond to this model, but at least we can propose a model which fits the observation, a model in which the background horizontal magnetic field of the prominence is mixed with a turbulent field. This is an interesting result for the physics of solar prominences, but it is also illustrative on how pattern recognition can condense the information and help separate the actual amount of information in the observed data from the models we use to interpret it. We found a case in which the inversion code was not giving as a measurement of the magnetic field, but telling us *this is not the right model, try with something else* (see Schmieder *et al.* 2015, this issue).

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

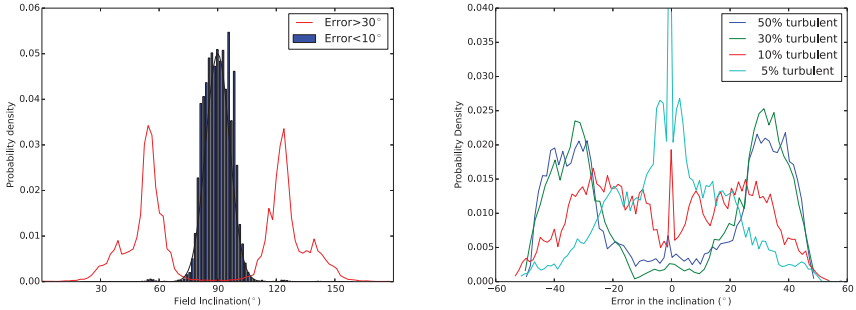

Figure 2. Histograms of the inclination of the magnetic field. Left, from the inversion of 138 prominences observed with THEMIS in 2014, bars are for solutions with error bar smaller than 10° while the continuous line is for solutions with errors larger than 30°. Right, from simulations of models combining a horizontal magnetic field with a turbulent field at ratios of 5, 10, 30 and 50%.

3. Fringe removal

One of the most ubiquitous patterns found in spectroscopic observations are fringes. Fringes in most spectroscopic instruments are not part of the intended result, but rather undesired side-effects. Instrument design tries to eliminate them to undetectable visibility levels but, in reality, fringes will reappear as soon as an instrument is taken to the limit. Thus fringes can be seen as intrinsic side-effects of instruments pushed to the limit. In this perspective it is just a question of time and effort before the data reduction codes of a spectroscopic instrument face the necessity of removing fringes from the data. And since we detect them as very characteristic patterns, why not to try pattern recognition on them?

In a sense, the usual Fourier filtering technique is a pattern recognition approach, one focused on finding sinusoidal patterns. But even when fringes are simple sinusoidal patterns, our spectral lines are not. This means that the Fourier decomposition of spectral lines may have a non-zero power at the frequency of the fringe pattern and any filtering will eliminate some information from the lines as well. Fourier filtering may work in some cases and may not work in others. Casini *et al.* (2012) tried an approach based upon PCA. In this approach, fringes are assumed to be a constant pattern present in the data, in particular during the flat field calibrations. When PCA is applied to these calibrations it will identify the predominant fringe pattern and associate a principal component to it, while spectral lines will end up in another principal component. If this is the case, it will suffice to decompose the data as a linear combination of those principal components and make zero the one corresponding to fringes before reconstructing the data as a linear combination of all the others. There is no assumption here about the actual pattern associated with fringes, sinusoidal or other. However, there is the assumption that the fringe pattern will be separable from the other principal components related to spectral lines. Unfortunately, this is not always the case and, sometimes, no principal component will contain just fringes, while the others contain just spectral lines. In these situations cancelling the contribution of one principal component will, as in the case of Fourier filtering, eliminate spectral information.

One implicit assumption behind Fourier filtering is sparsity: since we assume that fringes will present a sinusoidal pattern, decomposing the data in a Fourier series harbingers a few non-zero frequencies. The sparsity assumption is clearly evident in the PCA algorithm: it is based on the hope that just one or a few principal components of the image will contain the whole fringe pattern. Sparsity is what made them work in those

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

cases in which they did work. The drawback was the imposition of the space in which we assumed the signals were sparse. In the case of Fourier filtering, the fringes may be sparse in the Fourier space, but the spectral lines may not and this is a serious drawback. In the case of PCA filtering, both the fringes and the spectra were assured to be sparse in the series of principal components, but they may not be sparse in separable subspaces. Could one take advantage of sparsity without the drawback of the imposed space on which fringes or spectra had to be sparse? This is what is attempted with Relevance Vector Machines (RVM; Tipping 2000; Asensio Ramos & Manso Sainz, 2012). In RVMs we drop the requirement of an algebraic basis of the space of data. This means that the number of basic patterns has no upper limit. We will next build a dictionary (the equivalent of an algebraic basis but without the requirements of orthogonality or completitude) of the characteristic patterns we expect in the data. For example, to describe fringes we can build a dictionary of many sine and cosine patterns at a range of frequencies that generously includes the rough frequencies of the fringes in the data. And to describe spectral lines we can build a dictionary of gaussians of several widths and positions to roughly match all the atomic lines present in the spectra. The two dictionaries together must describe the data, but the number of entries in the dictionary is certainly enormous. In the example below, the dictionary has several thousand sine, cosine and gaussians, compared to the few principal components we normally use in classic pattern recognition techniques. Obviously we cannot just try to describe the data as a combination of all those entries: they are not orthogonal and they are too many. But we can impose sparsity: for a given spectral profile with fringes and through the use of bayesian techniques we can try combinations of all the entries, compute the posterior distributions and eliminate from the list those dictionary entries whose posterior distribution has near-zero average. The process is restarted with all other entries and we prune the dictionary until we find the smaller set of dictionary entries able to describe the data. Once this is achieved we will have a small set of dictionary entries, some of which will belong to the description of fringes and some others to the description of spectral lines plus some residual that, hopefully, is just noise. We can reconstruct the data by using just the gaussian description of the lines in the dictionary plus the noise.

An example of this is shown in Fig. 3 on THEMIS data of the He D_3 line in prominences, the same kind of data that will be inverted with a PCA algorithm as described in the previous section. A new polarimeter was used in this data set that introduced polarized fringes with frequencies similar to the range of frequencies of the Stokes signals. In this case we could nevertheless reduce the dictionary to just 420 entries, 400 for the fringes and just 20 for the line profiles. Although small, this number is 100-fold the number of principal components used for the description of this line during the inversion. The RVM pruned that dictionary down to 12 entries on average, or just 3% of the dictionary. Different profiles in the image used different entries of the dictionary, but on average just 12 sufficed. Sparsity, made explicit in that 3% figure, is the key for this scheme to work. In RVMs the dictionary is sparse for both the fringes and the spectral lines, unlike in Fourier filtering, and fringes and spectral lines are in separated parts of the dictionary, unlike PCA techniques. This makes of RVM a more flexible tool for fringe removal. However it is reliant on the ability of the dictionary to reproduce the signal. The dictionary should be big enough that all patterns in the data have a counterpart. Since dictionaries have to be built beforehand this may be difficult in some cases. López Ariste & Asensio Ramos (2015) explored the consequences of incomplete dictionaries. As expected when the dictionary is not big enough, the algorithm will try to fit the overseen patterns as weird combinations of the fringe and spectral line dictionary entries. The result will be that fringes will not be completely removed but some line information

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

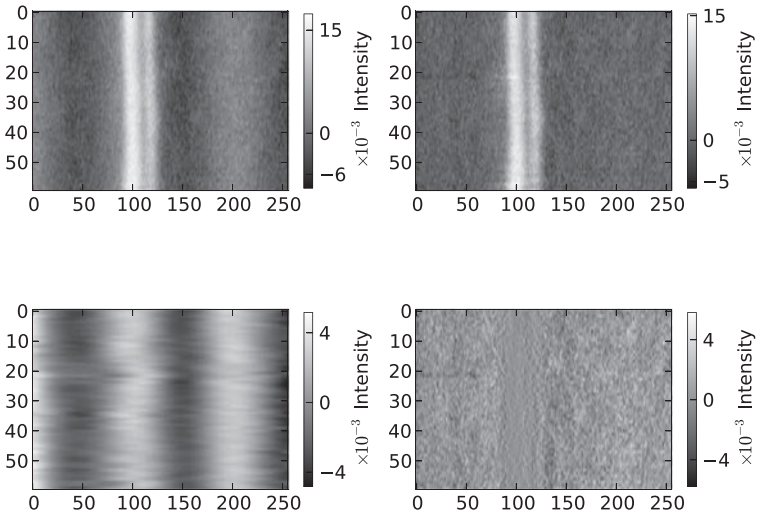

Figure 3. Example of defringing of linear polarization spectra in the He D_3 line observed in solar prominences with the THEMIS-MTR instrument. Wavelength is in the abscissas, which are left in pixel units since the actual units do not matter for defringing. From top to bottom and left to right: the original data, the defringed data, the extracted fringes, and the residual. From López Ariste & Asensio Ramos (2015).

may be deleted. Fortunately, López Ariste & Asensio Ramos (2015) showed that, in the explored cases, the code successfully removed fringes well beyond the limits foreseen in advance. Exploiting sparsity is once again in this example the key to solve a difficult problem.

4. *A posteriori* increase of spectral resolution

The two previous examples of the use of pattern recognition techniques in spectropolarimetry imply the computation of a sparse dictionary of patterns (principal components, dictionary entries, Fourier series, etc). Although the success of the techniques used stemmed from the sparsity of the signal when represented using those patterns, the critical point in the use of the technique was the ability to build or compute the dictionary of patterns. This was done by computing correlations among a representative set of data (in the PCA inversion code) or by our knowledge of the different patterns that may appear in the data (when building the dictionaries for fringe removal). Determining the dictionary of patterns on which the data is sparse is the most difficult aspect of the application of these techniques and, indeed, a good share of the lines written in the different references in the literature is dedicated to assure the reader that the dictionary encompased all possible cases. In this last example this aspect is even more critical.

Let us suppose that we are concerned with a certain spectral region and that we know in advance the principal components of that spectrum at a spectral resolution of 600000. Our instrument, however, can only provide us with a spectral resolution of merely 100000. Can we recover the spectra at high spectral resolution? Beyond deconvolution limits the classic answer is negative since there is an ambiguity of (infinite) high resolution spectra resulting in the same low-resolution signal. This answer is usually proposed as a corollary of Shannon's theorem saying that 2 pixels is the resolution limit and there is no power at Fourier frequencies above the Nyquist frequency. However, this argument relies on the fact that there is no other information about the high resolution spectrum than the

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

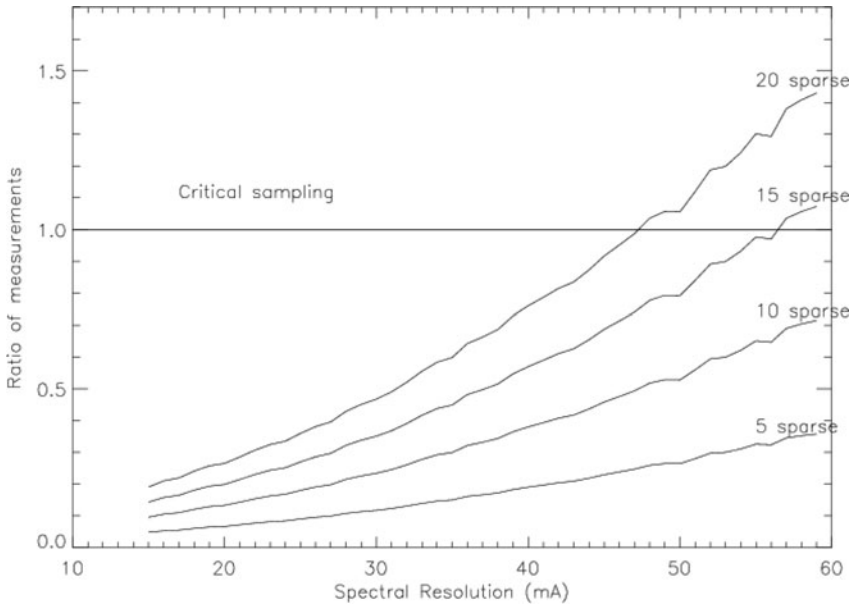

Figure 4. Compressed sensing condition (Eq. 4.1) for the reconstruction of Stokes V spectra of the Fe lines at 630 nm at a resolution of 10 mÅ starting from spectra at the resolutions indicated in abscissas. The number of measurements is given as the ratio respect to the critical sampling value at each resolution, corresponding to 2 pixels per resolution element. The condition equation is computed for 4 cases with different sparsities.

low-resolution signal. And this assumption is wrong. We know a few more things. On the trivial side we can be certain that even at high resolution the spectrum is a continuous semi-positive function. This eliminates many possible high-resolution spectra resulting in the low-resolution signal observed, but still we are left with infinite possible solutions.

In the last 20 years a new approach has appeared called *compressed sensing* (Candès *et al.* 2006) that takes advantage of the sparsity of a signal in an *a priori* known space. The knowledge of this space on which the spectrum is sparse is often enough to increase spectral resolution beyond the 2-pixel Nyquist limit†. Asensio Ramos & López Ariste (2010) showed that a simple application of this formalism allowed to recover a factor 4 in resolution. The condition for this resolution zoom to work is that the inequality

$$M \geqslant C\mu(\phi, W)^2 K \log N \tag{4.1}$$

is satisfied. In that equation, M is the number of measurements made at low resolution, that is, the number of pixels of the low resolution spectrum. For the zoom to high resolution to work the number of measurements, the number of pixels M, has to be larger than a certain threshold. This threshold depends on N, the number of pixels in the high resolution spectrum, on K, the sparsity of the signal in a given space W, and on $\mu(\phi, W)$, the correlation between the PSF (Point Spread Function) of the instrument ϕ (the degradation of the signal) and the principal components of W where we assumed that the signal is sparse. C is just a constant roughly equal to 1.

This equation makes a lot of sense. For example, if we expect the signal to be always proportional to a gaussian at a given place, only one measurement is enough to determine its amplitude, but if the gaussian can shift position, then two measurements will be

† To be rigorous, the information content increases and a new Nyquist frequency appears much larger than the one from the 2-pixel argument.

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

necessary: the minimum number of measurements required has to be proportional to the sparsity. It also makes sense that any correlation between the PSF that degrades the signal and the signal itself will mask the information, so the smaller this correlation, the smaller the number of measurements required. Eventually, and perhaps surprisingly at first view, a random PSF would be the ideal measurement setup. Random masks are an interesting tool to measure signals at very low resolution. Fig. 4 shows an illustration of this formula in increasing the spectral resolution of Stokes profiles in the domain of the Zeeman-sensitive Fe I line at 630 nm. This spectral domain was synthesized in a simulated photosphere (Rempel 2012; Borrero *et al.* 2014) with the SIR code (Ruiz Cobo & del Toro Iniesta 1992) at steps of 10 mÅ. The resulting profiles were used to compute the principal components of the data set and then they were convolved with gaussian PSFs to resolutions of 20 through 60 mÅ. In Fig. 4 we show the computation of Eq. 4.1 for the cases in which we accept that just 5, 10, 15 or 20 principal components reproduce the data. Starting from the spectral sampling in abscissas we compute how many measurements M will at least be required to recover the spectral resolution of 10 mÅ. This number M is not given directly but as a ratio compared to the case of 2 pixels per resolution element (critical sampling). So when in ordinates we find the value 1 for the spectra at 60mÅ it means that the number of pixels in the low resolution spectrum at 60 mÅ is such that each pixel samples 30 mÅ. Values above 1 in ordinates mean that we are oversampling the low resolution spectrum, while values under 1 mean that we are undersampling it. If 5 principal components are sufficient to describe the data, our low resolution instrument can afford to severely undersample the spectra, and even at 60 mÅ resolution it can afford to have pixels that sample larger than 60 mÅ regions and still recover the high resolution data through compress sensing. In less sparse conditions more and more measurements are required. If 20 principal components are needed, at 50 mÅ resolution our instrument has to have at least 25 mÅ pixels if we want to recover the high resolution spectra, and at 60 mÅ it has to clear over-sample the spectrum. Sparsity is therefore a critical aspect for this spectral resolution zoom to work. But even then it appears that a factor 6 increase in spectral resolution is possible if just we can find a set of principal components on which our high resolution spectra has sparsities smaller than 20.

This kind of results can have a direct impact in instrument design: if we are able to find or compute a set of principal components such that the data is sparse when decomposed in that space, then we do not need to make the instrument extremely resolutive. Our spectrographs can have smaller spectral resolutions, and hence be overall smaller, if we can figure out the space on which that spectra is sparse. In our tests it is clear that even a factor 6 beyond the Nyquist frequency is achievable meaning, for example, a spectrograph 6 times shorter.

5. Conclusion

Pattern recognition rests upon sparsity, the property of a data set to depend on just a few coordinates when represented in the right space. Spectropolarimetric data is often a sparse data set, and a few eigenprofiles or principal component suffice to describe all data. This allows us to use a full set of recent mathematical techniques based upon sparsity. In this contribution we have illustrated the potential of these techniques with 3 examples. In the first one, sparsity was used just to describe data and models with a small number of coefficients that can be easily and quickly compared to determine which model better represents data and give us a magnetic field in solar prominences. In the second case, sparsity was the key to constrain the description of fringes and spectral lines in raw

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press

data to a few entries of a hand-made dictionary thanks to relevance vector machines. Without sparsity the solutions to the problem are infinite, in the presence of sparsity a unique solution is found that allowed us to remove fringes. In the third example, the knowledge of the sparse set of principal component allowed us through compress sensing to greatly reduce our observables in number and resolution, yet to be confident that a high resolution spectrum could be recovered, holding the promise of simpler instruments without loss of precision.

Sparsity is therefore a property of observations that is worth exploring and taking full profit of. New mathematical techniques and tools based upon sparsity appear almost daily, techniques that can be often applied to our data handling problems in spectropolarimetry almost straightforwardly.

Acknowledgements

Fig. 2 was made using data acquired at THEMIS in 2014 by P. Levens and N. Labrosse (U. of Glasgow), B. Schmieder (Obs. de Paris), S. Gùnar (U. of St Andrews) and the author with support by the SOLARNET project (www.solarnet-east.eu), funded by the European Commission's FP7 Capacities Programme under the Grant Agreement 312495.

References

Aulanier, G. & Démoulin, P. 1998, *ApJ* 703, 114
Asensio Ramos, A. & López Ariste, A. 2010, *A&A* 509, A49
Asensio Ramos, A. & Manso Sainz, R. 2012, *A&A* 547, A113
Borrero, J. M., Lites, B. W., Lagg, A., Rezaei, R., & Rempel, M. 2014, *A&A* 572, A54
Candes, E., Romberg, J., & Tao, T. 2006, *IEEE Transactions on Information Theory* 52, 489
Casini, R., Judge, P. G., & Schad, T. A. 2012, *ApJ* 756, 194
Labrosse, N., Heinzel, P., Vial, J.-C., Kucera, T., Parenti, S., Gunár, S., Schmieder, B., & Kilper, G. 2014, *Space Science Review* 151, 243
López Ariste, A. & Casini, R. 2002, *ApJ* 575, 529
López Ariste A., Asensio Ramos A., Manso Sainz R., Derouich M., & Gelly, B. 2009, *A&A* 501, 729
López Ariste, A. 2015, in: J.-C. Vial & O. Engvold (eds.), Solar Prominences, Astrophysics and Space Science Library 415. (Springer International Publishing Switzerland), p. 179.
López Ariste, A. & Asensio Ramos, A. 2015, *A&A*. In preparation
Mackay, D. H., Karpen, J. T., Ballester, J. L. , Schmieder, B., & Aulanier, G. 2010, *Space Science Review* 151, 333
Rees, D. E., López Ariste, A., Thatcher, J., & Semel, M. 2000, *A&A* 355, 759
Rempel, M. 2012, *ApJ* 750, 62
Ruiz Cobo, B. & del Toro Iniesta, J. C. 1992, *ApJ* 398, 375
Schmieder, B., Tian, H., Kucera, T. *et al.* 2014, *A&A* 569, A85
Skumanich, A. & López Ariste, A. 2002, *ApJ* 570, 379
Socas-Navarro, H., López Ariste, A., & Lites, B. W. 2001, *ApJ* 553, 949
Tipping, M. E. 2001, *J. Mach. Learn. Res.* 1, 211

https://doi.org/10.1017/S1743921315004792 Published online by Cambridge University Press