Journal of Physics: Conference Series **699** (2016) 012010 doi:10.1088/1742-6596/699/1/012010

Sparsity and inverse problems in astrophysics

Jean-Luc Starck
Laboratoire AIM, UMR CEA-CNRS-Paris 7,
Irfu, Service d'Astrophysique,
CEA Saclay, F-91191 Gif-Sur-Yvette CEDEX, France

Abstract. Sparsity constraints are now very popular to regularize inverse problems. We review how sparsity can be used to solve inverse problems such as inpainting, deconvolution or blind source separation.

1. Introduction

Many data processing problems in astrophysics can be formalized as a linear inverse problem,

$$Y = HX \bullet N \ , \tag{1}$$

where Y are a set of noisy measurements, N is the unknown noise/error term, and the $\bullet$ operator translates the way the noise contaminates the data (e.g. "+" for additive noise). This noise/error can be either a stochastic measurement noise induced by the sensor (e.g. additive Gaussian noise, Poisson noise, or mixed Gaussian and Poisson), or a deterministic perturbation due, for example, to an imperfect forward model. X is the solution of our problem, and H is a linear operator. Finding X is an inverse problem.

Typical cases of such inverse problems are:

- *Deconvolution*: If the optical system is linear and shift-invariant, the relation between the data Y and the unknown X in the same coordinate frame is a convolution by a blurring kernel. H is the point spread function, PSF, of the optical system. We want to determine X knowing Y and H. This deconvolution problem has led to a large amount of work, the main difficulties being the existence of: (i) a cut-off frequency of the point spread function, and (ii) the additive noise.
- *Blind Deconvolution*: It is similar to the deconvolution problem, except the point spread function, i.e. matrix H, is unknown. Hence. we want to determine X and H knowing Y.
- *Myopic deconvolution*: It is an intermediate problem between deconvolution and blind deconvolution, the PSF H is partially known.
- *Superresolution*: H is the convolution by a blurring kernel followed by a sub-sampling. The goal is to recover spatial frequency information outside the spatial bandwidth of the data formation system.
- *Inpainting*: Missing data is a recurrent problem in astronomical data processing. H is a pixelwise multiplication by a binary mask; 1 for the pixels to be kept and 0 otherwise.

- *Radio-Interferometry Image Reconstruction:* this is a specific case of the deconvolution problem. Indeed the observed data, called visibilities, correspond to a subset of the Fourier component of the unknown X. The PSF, called *Dirty Beam*, is not compact in the direct space, and contains many zero in the Fourier domain.
- *Blind Source Separation:* In the blind source separation (BSS) setting, the instantaneous linear mixture model assumes that we are given m observations $Y = [y_1^T, \cdots, y_m^T]^T$ where each $(y_i)_{i=1,\cdots,m}$ is a row-vector of size t; each measurement is the linear mixture of n source processes . As the measurements are m different mixtures, source separation techniques aim at recovering the original sources $X = [x_1^T, \cdots, x_n^T]^T$ by taking advantage of some information contained in the way the signals are mixed in the observed data. The linear mixture model is rewritten in matrix form, $Y = HX + N$, where Y is the $m \times t$ measurement matrix (i.e. observed data), X is the $n \times t$ source matrix and H is the $m \times n$ mixing matrix. H defines the contribution of each source to each measurement. An $m \times t$ matrix N is added to account for instrumental noise or model imperfections. Similarly to the blind deconvolution, the H matrix is not known.
- *Image reconstruction:* an image is formed from projections.

When the inverse problem has not a unique and stable solution, it is an *ill-posed problem*, and a regularization is necessary to reduce the space of candidate solutions, and we need to incorporate some prior knowledge on the typical structure of the original signal X.

This prior information accounts typically for the "smoothness" class of the solution and can range from the uniform smoothness assumption to more complex knowledge of the geometrical structures of the solution. A very popular regularization in astronomy is the well-known Bayesian maximum entropy method (MEM), which is based on the principle that we want to select the simplest solution which fits the data. Sparsity is an alternative to describe the *simplest solution*, and has emerged as very powerful approach for regularization [1].

Sparse recovery involves four main ingredients which are the keys for its success. These ingredients are:

- **Data representation though the use of a dictionary:** this will be discussed in next section.
- **The noise modeling:** having the right dictionary is only a part of the story and is not enough. In most inverse problems, one needs to disentangle the useful signal/information from the noise. This requires to take properly into account the noise behavior in the observed data. Different kinds of noise models such as Gaussian, Poisson, mixture of Gaussian and Poisson, correlated noise, etc., can be used to derive the distribution of the coefficients in the chosen dictionary and therefore to detect, with a given confidence level, significant coefficients that should be used to reconstruct the signal of interest.
- **The variational problem and minimization algorithm:** piecing together the first ingredient and the second one (i.e. regularization and data fidelity), solving the inverse problem at hand as solving a variational (optimization) problem. To solve the obtained optimization problem, very general class of optimization schemes can be used, namely proximal splitting algorithms, that are fast, easy to implement and enjoy provable convergence. A very efficient proximal minimization method is presented in Section 5.3.
- **The choice of parameter(s):** variational formulation of inverse problems involve parameters that must be chosen. The scientist does not have to choose these parameters based on a tedious trial-and-error, by running many times the algorithm with a range of parameters. Furthermore, one definitively does not want to make scientific conclusions based on an ad-hoc parameter. We show in Section 5.4 how these parameters can be derived from the noise modeling.

A brief tour of sparsity is given in Section 2, and the different kind of disctionaries are discussed in Section 3. Specific inverse problems will be addressed, such as denoising and deconvolution in Sections 4 and 5, inpainting in Section 6, and blind source separation in Section 7.

2. Sparsity

A signal $X = (X[1], \ldots, X[N])$ considered as a vector in $\mathbb{R}^N$, is said to be sparse if most of its entries are equal to zero. If k number of the N samples are equal to zero, where $k \ll N$, then the signal is said to be k-sparse. In the case where only a few of the entries have large values and the rest are zero or close to zero the signal is said to be weakly sparse (or compressible). With a slight abuse of terminology, in the sequel, we will call compressible signals sparse. Generally signals are not sparse in direct space, but can be sparsified by transforming them to another domain. Think for instance of a purely sinusoidal signal which is 1-sparse in the Fourier domain, while it is clearly dense in the original one. In the so-called sparsity synthesis model, a signal can be represented as the linear expansion

$$X = \Phi\alpha = \sum_{i=1}^{T} \phi_i \alpha[i] \quad , \tag{2}$$

where $\alpha[i]$ are the synthesis coefficients of X, $\Phi = (\phi_1, \ldots, \phi_T)$ is the dictionary, and ϕ_i are called the atoms (elementary waveforms) of the dictionary Φ. In the language of linear algebra, the dictionary Φ is a $N \times T$ matrix whose columns are the atoms normalized, supposed here to be normalized to a unit ℓ_2-norm, i.e. $\forall i \in [1, T], \|\phi_i\|_2^2 = \sum_{n=1}^{N} |\phi_i[n]|^2 = 1$ [1]. A function can be decomposed in many dictionaries, but the best dictionary is the one with the sparsest (most economical) representation of the signal. In practice, it is convenient to use dictionaries with fast implicit transform (such as Fourier transform, wavelet transform, etc.) which allow us to directly obtain the coefficients and reconstruct the signal from these coefficients using fast algorithms running in linear or almost linear time (unlike matrix-vector multiplications). The Fourier, wavelet and discrete cosine transforms provide certainly the most well known dictionaries.

Sparsity as a very flexible model to represent data

Most natural signals however are not exactly sparse but rather concentrated near a small set. Such signals are termed *compressible* or *weakly sparse* in the sense that the sorted magnitudes $|\alpha_{(i)}|$, i.e. $|\alpha_{(1)}| > |\alpha_{(2)}|, ..., > |\alpha_{(t)}|$, of the sequence of coefficients α decay quickly according to a power law, i.e. $|\alpha_{(i)}| \leq Ci^{-1/r}$, $i = 1, \ldots, t$, where C is a constant. The larger r is, the faster the amplitudes of coefficients decay, and the more compressible the signal is. In turn, the non-linear ℓ_2 approximation error of α (and x) from its M largest entries in magnitude also decrease quickly. One can think, for instance, of the wavelet coefficients of a smooth signal away from isotropic singularities, or the curvelet coefficients of a piecewise regular image away from smooth contours. A comprehensive account on sparsity of signals and images can be found in [2].

3. Dictionaries

Obviously, the best dictionary is the one which leads to the sparsest representation. Hence we could imagine having a huge overcomplete dictionary (i.e. $T \gg N$), but we would be faced with prohibitive computation time cost for calculating the α coefficients. Therefore there is a trade-off between the complexity of our analysis step (i.e. the size of the dictionary) and the

[1] The l_p-norm of a vector X, $p \geq 1$, is defined as $\|X\|_p = \left(\sum_i |X[i]|^p\right)^{1/p}$, with the usual adaptation $\|X\|_\infty = \max_i X[i]$.

computation time. Some specific dictionaries have the advantage of having fast operators and are very good candidates for analyzing the data. The Fourier dictionary is certainly the most famous, but many others have been proposed in the literature such as wavelets [3], ridgelets [4], curvelets [5, 6] or shearlets [7], to name only a few.

Toward Morphological Diversity
The morpholocal diversity concept was introduced in [8] in order to model a signal as a a finite linear mixture, each component of the mixture being sparse in a given dictionary. The idea is that a single transformation may not always represent an image well, especially if the image contains structures with different spatial morphologies. For instance, if an image is composed of edges and a locally oscillating texture, we can consider edges to be sparse in the curvelet domain while the oscillating texture is better sparsified in the local DCT domain. It has been shown by several authors that choosing a dictionary as a combination of several sub-dictionaries, each sub-dictionary having a fast transformation/reconstruction, allows us to enjoy the advantages of all sub-dictionaries, still having fast and efficient algorithms.

Adaptive representations
Different approaches have also been proposed in order to build a dictionary directly from the data. This is the case in learned dictionaries [9], for instance using e.g. the KSVD algorithm [10], the grouplet decomposition [11] or the GMCA method for multichannel/hyperspectral data [12].

4. Sparse regularization for inverse problems

In the following, for a vector z we denote $\|z\|_p^p = \sum_i |z_i|^p$ for $p \geq 0$. In particular, for $p \geq 1$, this is the p-th power of the ℓ_p norm, and for $p = 0$, we get the ℓ_0 pseudo-norm which counts the number of non-zero entries in z. The ℓ_0 regularized problem amounts to minimizing

$$\tilde{\alpha} \in \mathrm{argmin}_\alpha \, \|Y - H\Phi\alpha\|_2^2 + \lambda\|\alpha\|_0, \tag{3}$$

where λ is a regularization parameter. A solution $\tilde{X}$ is reconstructed as $\tilde{X} = \Phi\tilde{\alpha}$. Clearly, the goal of Eq. 3 is to minimize the number of non-zero coefficients describing the sought after signal while ensuring that the forward model is faithful to the observations.

Solving Eq. 3 is however known to be NP-hard. The ℓ_1 norm has been proposed as a tight convex relaxation of Eq. 3 leading to the minimization problem

$$\tilde{\alpha} \in \mathrm{argmin}_\alpha \, \|y - H\Phi\alpha\|_2^2 + \lambda\|\alpha\|_1, \tag{4}$$

where λ is again a regularization parameter different from that of Eq. 3. There has been a tremendous amount of work where researchers spanning a wide range of disciplines have studied the structural properties of minimizers of Eq. 4 and its equivalence with Eq. 3. Eq. 4 is computationally appealing and can be efficiently solved, and it has also been proved that under appropriate circumstances, Eq. 4 produces exactly the same solutions as Eq. 3, see e.g., [13].

5. Astronomical Image Restoration

5.1. Denoising

The denoising problem corresponds to the case where H is the identity (i.e. $Y = X + N$). The classical modus operandi is then to first apply the transform operator to the noisy data Y, then apply a nonlinear estimation rule to the coefficients (each coefficient individually or as a group of coefficients), and finally compute the inverse transform to get an estimate $\tilde{X}$. In brief, A classic approach is to derive the threshold t from the noise modeling. Many thresholding or shrinkage rules have been proposed. Among them, hard and soft thresholding are certainly the most well known.

5.2. Hard and Soft Thresholding

Once coefficients $\alpha = \Phi^t Y$ have been calculated, *Hard thresholding* [14, 15] consists of setting to zero all coefficients whose magnitude is less than a threshold t:

$$\tilde{\alpha}_k = \text{HardThresh}_t(\alpha_k) = \begin{cases} \alpha_k & \text{if } |\alpha_k| \geq t, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Hard thresholding is a keep-or-kill procedure.

Soft thresholding [16, 17] is defined as the kill-or-shrink rule:

$$\tilde{\alpha}_k = \begin{cases} \text{sign}(\alpha_k)(|\alpha_k| - t) & \text{if } |\alpha_k| \geq t, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The coefficients above the threshold are shrunk toward the origin. This can be written in the compact form

$$\tilde{\alpha}_k = \text{SoftThresh}_t(\alpha_k) = \text{sign}(\alpha_k)(|\alpha_k| - t)_+ \ , \tag{7}$$

where $(\cdot)_+ = \max(\cdot, 0)$.

Thresholding as a Minimization Problem Suppose that Φ is orthonormal. It can be proved that equations Eq. 6 and 7 are the unique closed form solution to the following minimization problems for a hard or soft thresholding with a threshold t:

$$\tilde{\alpha} = \underset{\alpha}{\text{argmin}} \ \|Y - \Phi\alpha\|^2 + t^2 \|\alpha\|_0 \quad \textbf{Hard thresholding} \tag{8}$$

$$\tilde{\alpha} = \underset{\alpha}{\text{argmin}} \ \tfrac{1}{2} \|Y - \Phi\alpha\|^2 + t \|\alpha\|_1 \quad \textbf{Soft thresholding} \tag{9}$$

Therefore, hard and soft thresholding have been used as building blocks to derive fast and efficient iterative thresholding techniques to minimize more complex inverse problems with sparsity constraints, such as inpainting and deconvolution.

For the orthogonal wavelet transform, hard thresholding is known to result in a larger variance estimate, while soft thresholding with the same threshold level creates undesirable bias because even large coefficients lying out of noise are shrunk. In practice, hard thresholding is generally preferred to soft thresholding.

5.3. Deconvolution

In a deconvolution problem, when the sensor is linear, H is the block Toeplitz matrix. A first iterative thresholding deconvolution method was proposed in [18] which consists in the fowollowing iterative scheme:

$$X^{(n+1)} = X^{(n)} + H^T \left(\mathbf{WDen}_{\Omega^{(n)}} \left(Y - HX^{(n)} \right) \right) \tag{10}$$

where **WDen** is an operator which performs a wavelet thresholding, i.e. applies the wavelet transform of the residual $R^{(n)}$ (i.e. $R^{(n)} = Y - HX^{(n)}$), threshold some wavelet coefficients, and applies the inverse wavelet transform. Only coefficients that belong to the so called *multiresolution support* $\Omega^{(n)}$ [18] are kept, while the others are set to zero. At each iteration, the multiresolution support $\Omega^{(n)}$ is updated by selecting new coefficients in wavelet transform of the residual which have an absolute value larger than a given threshold. The threshold is automatically derived assuming a given noise distribution such as Gaussian or Poisson noise.

More recently, it was shown [19, 20, 21] that a solution of (4) can be obtained through the following iteration:

$$\alpha^{(n+1)} = \text{SoftThresh}_\lambda \left(\alpha^{(n)} + \Phi^T H^T \left(Y - H\Phi\alpha^{(n)} \right) \right) \ , \tag{11}$$

with $\|H\| = 1$. In the framework of monotone operator splitting theory, it is called Forward-Backward (FB) splitting algorithm and is classical gradient projection method for constrained convex optimization. It was shown that for frame dictionaries, it converges to the solution [21]. In the last ten years, many new optimization methods have been proposed to perform a sparse regularization in a very more efficient way. For instance, the primal-dual splitting algorithm of [22] can handle properly both sparse and positivity constraints, and leads to the following scheme:

(i) **Initialization:** Choose $\tau > 0$ and $\sigma > 0$ such that $1 - \tau\sigma \|\Phi\|^2 > \|H\|^2 /2$; A sequence of relaxation parameters $\mu_n \in]0, 1]$; (x_0, u_0).

(ii) **Main iteration**

- $p_{n+1} = \left(x_n - \tau\Phi u_n + \tau H^T(y - Hx_n)\right)_+$;
- $q_{n+1} = (\mathbf{I} - \text{SoftThresh}_\lambda)\left(u_n + \sigma\Phi(2p_{n+1} - x_n)\right)$;
- $(x_{n+1}, u_{n+1}) = \mu_n(p_{n+1}, q_{n+1}) + (1 - \mu_n)(x_n, u_n)$.

The parameter μ_n is a relaxation one that can be used to accelerate the algorithm. In the unrelaxed case, i.e. $\mu_n = 1$ for all n, we obtain $(x_{n+1}, u_{n+1}) = (p_{n+1}, q_{n+1})$.

5.4. Regularization parameters

Once the dictionary and the minimization method are chosen, a final problem remains: choosing parameters that are needed to control the algorithm. Most minimization methods, using l_0 and l_1, have a single thresholding step, where coefficients in the dictionary have to be soft or hard-thresholded using a threshold value, which is a value λ, common to projection in Φ. This parameter controls the trade-off between the fidelity to the observed visibility and the sparsity of the reconstructed solution. In real applications however, the sought-after coefficients are likely to have very different scalings, and it would be wiser to adapt the regularization parameter to each coefficient. This corresponds to solving

$$\min_\alpha \ \frac{1}{2} \|y - H\Phi\alpha\|^2 + \sum_{i=1}^{T} \lambda_i \left|\alpha[i]\right| \ . \tag{12}$$

where T is the size of the vector α (i.e. the number of coefficients).

5.4.1. Noise driven strategy Recall that in the iterative schemes discussed earlier, the noise impacts the solution through the residual $R^{(n)} = Y - H\Phi\alpha^{(n)}$ and its backprojected coefficients $\alpha^{(R)} = \Phi^T H^T R^{(n)}$. If we see the regularization as a denoising step, then we can consider that λ_i should be related to the noise in $\alpha^{(R)}$, $\lambda_i = k\sigma_i$, where k is generally chosen between 3 and 5, and σ_i is the noise standard deviation on the coefficient $\alpha[i]$. If the noise is zero-mean additive with known covariance matrix Σ, then σ_i is the square root of the ith entry on the diagonal of the covariance matrix $\Phi^T H^T \Sigma H \Phi$. Alternatively, if we know how to simulate a set of N_r realistic noise realizations, we can calculate the coefficients $\alpha^{(r)} = \Phi^T H^T Z_r$ for each realizations Z_r, and we have:

$$\sigma_i = \sqrt{\frac{1}{N_r} \sum_{r=1}^{N_r} \alpha^{(r)}[i]^2} \ .$$

5.4.2. Residual-driven strategy for subband transforms In [23], it is proposed to fix the threshold only from the noise distribution at each wavelet scale j. A similar approach can be used for multiscale geometric transforms such as the curvelets.

Suppose that H is a convolution operator, Φ corresponds to a band-pass transform with J bands, and the noise N is stationary an Gaussian. Then $\alpha_j^{(N)} = (\Phi^T H^T N)_j$ in each band j is also Gaussian and stationary. We can therefore only consider one regularization parameter per band rather than one per coefficients, and Eq. 12 becomes

$$\min_{\alpha} \frac{1}{2} \|y - H\Phi\alpha\|^2 + \sum_{j=1}^{J} \lambda_j \|\alpha_j\|_1 , \tag{13}$$

where α_j are the coefficients in the band j and J is the number of bands. We could then use the $\alpha_j^{(R^{(n)})}$ to estimate standard deviation σ_j. As $R^{(n)}$ may contain also structure and not only noise, a more robust approach is to use the MAD (Median of Absolute Deviation) estimator. Then we have

$$\sigma_j = \mathrm{MAD}(\alpha_j^{(R^{(n)})})/0.6745 = \mathrm{median}(\left|\alpha_j^{(R^{(n)})} - \mathrm{median}(\alpha_j^{(R^{(n)})})\right|)/0.6745 ,$$

6. Inpainting

Missing data are a standard problem in astronomy. They can be due to bad pixels, or image area we consider as problematic due to calibration or observational problems. These masked areas lead to many difficulties for post-processing, especially to estimate statistical information such as the power spectrum or the bispectrum. The inpainting technique consists in filling the gaps. The classical image inpainting problem can be defined as follows. Let X be the ideal complete signal, Y the observed incomplete data and M the binary mask (i.e. $M_i = 1$ if we have information at pixel i, $M_i = 0$ otherwise). In short, we have: $Y = MX$. Inpainting consists in recovering X knowing Y and M. We thus want to minimize:

$$\min_X \|\Phi^T X\|_0 \quad \text{subject to} \quad Y = MX . \tag{14}$$

It was shown in [24] that this optimization problem can be efficiently solved through an iterative thresholding algorithm [24, 25] :

$$X^{(n+1)} = \Delta_{\Phi,\lambda_n}(X^{(n)} + Y - MX^{(n)}) . \tag{15}$$

where the nonlinear operator $\Delta_{\Phi,\lambda}(Z)$ consists in i) decomposing the signal Z in the dictionary Φ to derive the coefficients $\alpha = \Phi^T Z$, ii) thresholding the coefficients: $\tilde{\alpha} = \rho(\alpha, \lambda)$, where the thresholding operator ρ can either be a hard thresholding or a soft thresholding, and iii) reconstructing $\tilde{Z}$ from the thresholded coefficients $\tilde{\alpha}$.

The threshold parameter λ_n decreases with the iteration number and it plays a role similar to the cooling parameter of the simulated annealing techniques, i.e. it allows the solution to escape from local minima. More details on optimization in inpainting with sparsity can be found in [25, 2]. The case where the dictionary is a union of subdictionaries $\Phi = \{\Phi_1, ..., \Phi_K\}$ where each Φ_i has a fast operator has also been investigated in [24, 25].

Sparse inpainting has been used for different applications in astrophysics:

- high order statistics on weak lensing and Cosmic Microwave Background (CMB) data: for second order statistics such as power spectrum estimation, robust methods exist and it is generally not necessary to use more sophisticated methods. For higher order statistics (bispectrum, etc), dealing with the mask of missing data is a challenging task. It was shown that sparse inpainting can very well fill the gap, and keep a very high accuracy on the statistical estimators for both weak lensing [26] and CMB data [27]

- asteroseismology: Sparse inpainting has been included in the official CoRoT pipeline to correct the missing data in both asteroseismic and exoplanet channel [28, 29].
- Large scale studies on spherical maps for ISW (integrated Sachs-Wolfe effect) signal reconstruction [30] and CMB anomalies analysis [31].

7. Blind Source Separation

In the blind source separation (BSS) setting, the instantaneous linear mixture model assumes that we are given m observations $\{Y_1, \cdots, Y_m\}$ where each $\{Y_i\}_{i=1,\cdots,m}$ is a row-vector of size t; each measurement is the linear mixture of n source processes. As the measurements are m different mixtures, source separation techniques aim at recovering the original sources $\mathbf{X} = [X_1^T, \cdots, X_n^T]^T$ by taking advantage of some information contained in the way the signals are mixed in the observed data. The linear mixture model is rewritten in matrix form,

$$Y = HX + N, \tag{16}$$

where Y is the $m \times t$ measurement matrix (i.e. observed data), X is the $n \times t$ source matrix and H is the $m \times n$ mixing matrix. H defines the contribution of each source to each measurement. An $m \times t$ matrix $\mathbf{N}$ is added to account for instrumental noise or model imperfections.

It has been shown that sparsity is a very robust regularization to solve BSS for both underdetermined (i.e. we have less observations than unknown, $m < n$) [32] and overdetermined mixtures ($m \geq n$) [33, 12]. The GMCA (Generalized Morphological Component Analysis) algorithm [33] finds the sparsest solution through the following iterative scheme:

$$\begin{aligned} X^{(k+1)} &= \Delta_{\Phi,\lambda_k}(H^{+(k)}Y) \ , \\ H^{(k+1)} &= YX^{(k+1)^T}(X^{(k+1)}X^{(k+1)^T})^{-1} \ , \end{aligned} \tag{17}$$

where $H^{+(k)}$ is the pseudo-inverse of the estimated mixing matrix $H^{(k)}$ at iteration k, λ_k is a decreasing threshold and Δ_{Φ,λ_k} is the nonlinear operator which consists in decomposing each source X_i on the dictionary Φ, threshold its coefficients and reconstruct it. Finally, for hyperspectral data, it was advocated to impose sparsity constraints on the columns of the mixing matrix to enhance source recovery [12]. GMCA has shown to be very impressive to remove Galactic and extragalactic foregrounds in order to detect the Epoch of Reionization 21-cm signal [34].

8. Conclusion

Signal processing in the 20th century was mainly considering band limited signals with an acquisition system based on the Shannon-Nyquist sampling theorem, and inverse problems were solved using least square estimators. The last ten years have seen emerging a new paradigm, based on sparse signals and the compressed sensing sampling theorem, where sparse regularization is used to solve inverse problems. We have presented how sparse representations can be used in different applications such denoising, deconvolution, inpainting or blind source separation. More details can be found in [1].

Acknowledgments

This work was funded by the PHySIS project (contract no. 640174) and the DEDALE project (contract no. 665044), within the H2020 Framework Program of the European Commission.

References

[1] Starck J -L, Murtagh F and Fadili M J 2015 *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis* Cambridge University Press
[2] Starck J -L, Murtagh F and Fadili M J 2010 *Sparse Image and Signal Processing* Cambridge University Press
[3] Mallat S 1998 *A Wavelet Tour of Signal Processing* Academic Press
[4] Candès E J and Donoho D L 1999 Ridgelets: the key to high dimensional intermittency? *Philosophical Transactions of the Royal Society of London* A **357** 2495–2509
[5] Starck J -L, Candès E and Donoho D L 2002 The curvelet transform for image denoising *IEEE Transactions on Image Processing* **11**(6) 131–141
[6] Candès E J, Demanet L, Donoho D L and Ying L 2006 Fast discrete curvelet transforms *Multiscale Modeling and Simulation* **5**(3) 861–899
[7] Kutyniok G and Lim Wang-Q 2011 Compactly supported shearlets are optimally sparse *Journal of Approximation Theory* **163**(11) 1564–1589
[8] Starck J -L, Elad M and Donoho D L 2004 Redundant multiscale transforms and their application for morphological component analysis *Advances in Imaging and Electron Physics* 132
[9] Olshausen B A and Field D J 1996 Emergence of simple-cell receptive-field properties by learning a sparse code for natural images *Nature* **381**(6583) 607–609
[10] Aharon M, Elad M and Bruckstein A 2006 K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation *IEEE Transactions on Signal Processing* **54**(11) 4311–4322
[11] Mallat S 2009 Geometrical grouplets *Applied and Computational Harmonic Analysis* **26**(2) 161–180
[12] Bobin J, Starck J -L, Moudden Y and Fadili M J 2008 Blind source separation: The sparsity revolution *Advances in Imaging and Electron Physics* **152** 221–306
[13] Donoho D L 2006 For most large underdetermined systems of linear equations, the minimal ℓ_1 solution is also the sparsest solution *Communications on Pure and Applied Mathematics* **59**(7) 907–934
[14] Starck J -L and Bijaoui A 1994 Filtering and deconvolution by the wavelet transform *Signal Processing* **35** 195–211
[15] Donoho D L, Johnstone I M, Kerkyacharian G and Picard D 1995 Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society Series* B **57** 301–369
[16] Weaver J B, Yansun X, Healy D M and Cromwell L D 1991 Filtering noise from images with wavelet transforms *Magnetic Resonance in Medicine* **21**(2) 288–295
[17] Donoho D L 1995 De-noising by soft-thresholding *IEEE Transactions on Information Theory*, **41**(3) 613–627
[18] Starck J -L, Bijaoui A and Murtagh F 1995 Multiresolution support applied to image filtering and deconvolution *CVGIP: Graphical Models and Image Processing* **57** 420–431
[19] Figueiredo M A and Nowak R 2003 An EM algorithm for wavelet-based image restoration *IEEE Transactions on Image Processing* **12**(8) 906–916
[20] Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint *Communications on Pure and Applied Mathematics* **57** 1413–1541
[21] Combettes P L and Wajs V R 2005 Signal recovery by proximal forward-backward splitting *Multiscale Modeling and Simulation* **4**(4) 1168–1200
[22] Vũ B 2013 A splitting algorithm for dual monotone inclusions involving cocoercive operators *Advances in Computational Mathematics* **38**(3) 667–681
[23] Lanusse F, Paykari P, Starck J -L, Sureau F, Bobin J and Rassat A 2014 PRISM: Recovery of the primordial spectrum from Planck data *Astron. Astrophys.* **571** L1
[24] Elad M, Starck J -L, Donoho D L and Querre P 2005 Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA) *Applied and Computational Harmonic Analysis* **19** 340–358
[25] Fadili M J, Starck J -L and Murtagh F 2009 Inpainting and zooming using sparse representations *The Computer Journal* **52** 64–79
[26] Pires S, Starck J -L, Amara A, Réfrégier A and Teyssier R 2009 Cosmological model discrimination with weak lensing *Astron. Astrophys.* **505** 969–979
[27] Perotto L, Bobin J, Plaszczynski S, Starck J -L and Lavabre A 2010 Reconstruction of the cosmic microwave background lensing for Planck *Astron. Astrophys.* **519** A4
[28] García R A, Mathur S, Pires S, Régulo C, Bellamy B, Pallé P L, Ballot J, Barceló Forteza S, Beck P G, Bedding T R, Ceillier T, Roca Cortés T, Salabert D and Stello D 2014 Impact on asteroseismic analyses of regular gaps in Kepler data *Astron. Astrophys.* **568** A10
[29] Pires S, Mathur S, García R A, Ballot J, Stello D and Sato K 2015 Gap interpolation by inpainting methods: Application to ground and space-based asteroseismic data *Astron. Astrophys.* **574** A18
[30] Dupé F -X, Rassat A, Starck J -L and Fadili M J 2011 Measuring the integrated Sachs-Wolfe effect *Astron. Astrophys.* **534** A51
[31] Starck J -L, Fadili M J and Rassat A 2013 Low-ℓ CMB analysis and inpainting *Astron. Astrophys.* **550** A15

[32] Gribonval R, Rauhut H, Schnass K and Vandergheynst P 2008 Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms *Journal of Fourier Analysis and Applications* **14**(5) 655–687
[33] Bobin J, Starck J -L, Fadili M J and Moudden Y 2007 Sparsity and morphological diversity in blind source separation *IEEE Transactions on Image Processing* **16**(11) 2662–2674
[34] Chapman E, Zaroubi S and Abdalla F 2014 Foreground Removal vs. Foreground Avoidance: Contamination of the EoR Window ArXiv e-prints arXiv:1408.4695