

# Analysis of spectro-imaging data using the Karhunen-Loève expansion: Application to NGC 7027 in the infrared

D. Mékarnia<sup>1,\*</sup>, A. Bijaoui<sup>1</sup>, C. Delle Luche<sup>2</sup>, and J. P. Maillard<sup>3,\*</sup>

<sup>1</sup> Dpt Cassiopée, CNRS UMR-6202, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France

<sup>2</sup> ESO, Alonso de Cordova 3107, Vitacura Santiago 19, Chile

<sup>3</sup> Institut d'Astrophysique de Paris, 98 bis Bd Arago, 75014 Paris, France

Received 22 January 2003 / Accepted 15 January 2004

**Abstract.** A method, based on the Karhunen-Loève (KL) transformation, to improve the signal-to-noise ratio of a multispectral data cube is presented. As an illustration we have re-analyzed near-infrared multi-spectral images of the young planetary nebula NGC 7027 previously observed by Cox et al. (2002). Our KL analysis clearly shows an enhancement of both spectral and spatial information, particularly in the case of low signal-to-noise ratio images.

**Key words.** methods: data analysis – planetary nebulae: individual: NGC 7027

## 1. Introduction

Imaging spectroscopy, or integral field spectroscopy as it is also called, is the ability to record spectral information for each spatial sample of an astronomical field. Spectrophotometric information of the observed object is recorded in the form of a data cube, with spatial information in two dimensions and the spectral information in the third dimension. Because the detectors are of a two-dimensional nature, recording the third, missing dimension requires two classes of instruments: those for which time scanning is used and those for which spatial sampling is used. To the first class belong two interferometric systems, the scanning Fabry-Pérot (e.g., Ryder et al. 1998) and the imaging Fourier transform spectrometer or IFTS (Maillard 1995, 2000). The FP is well adapted for the study of a single emission line profile while an IFTS is more suitable for extended, rich emission line objects. To the second class belong the slit spectrometers with lenslet arrays, fiber bundle or image slicers as foreoptics to feed the entrance slit (Ren & Allington-Smith 2002). The latter can record the spectra for every point of a field in a single exposure but suffer from the loss of spatial resolution if a large field of view is observed. An imaging Fourier transform spectrometer gives simultaneously, over a large enough spectral range, the spectrum of all the resolved points of an extended object. It makes it possible to obtain wide field diffraction-limited imaging for visible to infrared wavelengths, with flexible spectral resolution. Due to its ability to achieve high

spectral resolution, this instrument provides the capability of measuring the gas kinematics from strong emission lines, e.g., the molecular hydrogen H<sub>2</sub> 1–0 S(1) emission line at 2.121 μm, over the full molecular envelope of young PNs. Significant astronomical results were obtained in the near infrared with an IFTS called BEAR on the 3.6-m Canada-France-Hawaii Telescope (Paumard et al. 2001; Cox et al. 1997, 2002, 2003). However, observations with an IFTS require data acquisition of a large number of frames for a given bandpass, particularly for high spectral resolution. Consequently deep spectroscopic investigations of many astronomical objects could be difficult to achieve because of the large total integration time needed for these observations. This parameter competes with the achievable signal-to-noise ratio.

In this paper we show how a data processing method based on the Karhunen-Loève (KL hereafter) transformation (Karhunen 1947; Loève 1960) can increase the signal-to-noise ratio of a spectrophotometric data cube making possible deep integral field observations. This method uses a linear transformation to translate and rotate the data cube in a new coordinate system that maximizes the variance. This approach has been successfully used for a variety of astronomical applications including data compression (e.g. Bijaoui 1974; Pelat 1974) and noise reduction, leading to identification of variable phenomena (Bijaoui & Doazan 1979), stellar and galaxy classifications (e.g. Connolly et al. 1995).

To enhance the spatial and spectral components of a multi-spectral data cube, the cross-correlation matrix of the data cube is first computed and a singular value decomposition is carried out, then a small number of basis vectors are used to reproduce a first approximation for each spectral image. As will be shown later, images reconstructed from a limited number of basis

Send offprint requests to: D. Mékarnia,  
e-mail: mekarnia@obs-nice.fr

\* Visiting Astronomer, Canada-France-Hawaii Telescope, operated by the National Research Council of Canada, le Centre National de la Recherche Scientifique de France and the University of Hawaii, USA.

vectors are less noisy than the original ones. A small and possibly significant amount of information is lost in this process. This information is retrieved by re-adding denoised residuals between the original and the restored images to the reconstructed ones. Denoising was obtained in this paper by using a wavelet method which takes into account both i) the cross-correlation between images or between spectra (KL expansion) and ii) the spatial and spectral local correlations (wavelet denoising).

We applied the KL expansion to increase the signal-to-noise ratio of the near-infrared spectro-imaging observations of the young planetary nebula NGC 7027. The data cube, consisting of 840 planes of  $128 \times 128$  pixels, was obtained by means of near-infrared Fourier transform spectro-imaging observations with high spatial and high spectral resolutions using a narrow-band filter (bandpass 4705–4745  $\text{cm}^{-1}$ ) which contains  $\text{H}_2$  1–0 S(1) in its bandpass (Cox et al. 2002).

In the second section of this paper, we describe the processing method based on the KL expansion. In Sect. 3 we outline the processing of the BEAR spectro-imaging data cube of NGC 7027 we have used. Results of the method applied to this data cube are presented in Sect. 4 and discussed in Sect. 5.

## 2. Karhunen-Loève processing

### 2.1. The KL basis

The KL expansion scheme has been widely applied for processing astronomical data (see for examples Pelat 1974; Bijaoui 1974; Bijaoui & Doazan 1979; Connolly et al. 1995; Tegmak et al. 1997; Vogeley & Szalay 1996). Let us call  $V_n(k, l)$ , ( $n \in (1, N)$ ;  $k$  and  $l$  the pixel coordinates), the set of images to process. We consider that for a given image  $n$  the intensities are the realization of a stochastic process with a probability density function (PDF)  $p_n(V)$ . A correlation may exist between the intensities of two different images at the same pixel,  $V_m(k, l)$  and  $V_n(k, l)$ . Then it is possible to derive a set of  $N$  linear combinations  $S_i = \sum_i b_{ni} V_n$  such that the values  $S_i$  and  $S_j$  are de-correlated. This can be done by applying the Principal Component Analysis (PCA) (e.g. Murtagh & Heck 1987) on the intensities. Then, the KL scheme is the following:

- compute the mean of each image  $\bar{V}_n$ , getting image  $V'_n$  after subtraction;
- compute the variance-covariance matrix  $C$  of the intensities.  $C$  is also the cross correlation matrix of images  $V'_n$ ;
- compute the  $C$  eigenvectors  $u_i$  of components  $u_{in}$  by decreasing eigenvalues  $\lambda_i$ ;
- compute the images  $S_i = \frac{1}{\sqrt{\lambda_i}} \sum_n u_{in} V'_n$ .

The image noise increases the  $C$  diagonal terms. On the hypothesis of uniform noise, its variance has to be subtracted from these terms.

$S_i$  is called the  $i$ th basis image. Generally only  $I < N$  images are computed.  $V'$  and  $S$  are the image (reduced from the mean) and the basis vectors respectively. We have the relation:

$$S = B.V' \quad (1)$$

with:

$$B = \Lambda^{-\frac{1}{2}} U \quad (2)$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & \dots & \ddots & 0 \\ 0 & \dots & \dots & \lambda_I \end{pmatrix} \quad (3)$$

and:

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1N} \\ u_{21} & u_{22} & \dots & u_{2N} \\ \dots & \dots & \ddots & \dots \\ u_{I1} & \dots & \dots & u_{IN} \end{pmatrix}. \quad (4)$$

If  $I < N$ , the intensities  $V'_n$  are not exactly restored. The approximation  $\tilde{V}'_n$  is:

$$\tilde{V}' = A.S \quad (5)$$

where:

$$A = \tilde{U} \Lambda^{\frac{1}{2}} \quad (6)$$

and  $\tilde{U}$  is the  $U$  transpose matrix.

The distance between  $V'$  and its approximation  $\tilde{V}'$  is equal to the sum of the eigenvalues not used in the approximation:

$$\sum_n \sum_{k,l} |V'_n(k, l) - \tilde{V}'_n(k, l)|^2 = KL \sum_{j=J+1}^N \lambda_j \quad (7)$$

where  $K$  and  $L$  are respectively the number of columns and lines of the images.

The choice of eigenvectors with decreasing eigenvalues allows one to concentrate the energy in a basis having the minimum of components.

### 2.2. The noise of the basis

Using few basis images the KL scheme allows one to restore images having a slight difference with the original ones. The restored images look less noisy than the original ones. From Eqs. (1) and (5) we get:

$$\tilde{V}' = AB.V' \quad (8)$$

$R = AB$  has a dimension  $N \times N$  and its terms are:

$$r_{nm} = \sum_{i=1, I} u_{in} u_{im}. \quad (9)$$

If  $I = N$ ,  $R$  is the identity matrix, else the sums are incomplete. For the restored image, the variance due to the noise is:

$$\sigma^2 (\tilde{V}'_n(k, l)) = \sigma^2 \sum_m r_{nm}^2 \quad (10)$$

where  $\sigma^2$  is the noise variance, admitted to be the same for each image value  $V_n(k, l)$ .

If  $I \approx N$  no gain is obtained for the signal-to-noise ratio. In our KL application  $I \ll N$  the sums are reduced and then the noise is significantly reduced. A faint residual noise appears in the first basis images, and it increases with the order. Denoising can be done on each image.

### 2.3. Denoising by the wavelet transform

Many papers have been dedicated to the application of discrete wavelet transforms for denoising images. The reader can get a general overview of these methods in Mallat (1999) or Starck et al. (1998). The general framework of the algorithms is the following:

- Application of a discrete wavelet transform: redundant transforms allowing shift invariance reduce the artifacts (Starck & Bijaoui 1994; Coifman & Donoho 1995).
- Thresholding: different thresholding rules were proposed.
- Restoration: the restoration is done by the inverse transform for un-redundant transforms. If it is not the case regularized inversion can be done.

The relation between the measured and the restored wavelet coefficients constitutes the core of the algorithm. Many rules were proposed. In this paper we choose Donoho's thresholding for which the threshold  $t$  is given by Donoho & Johnstone (1994):

$$t = \sigma_b \sqrt{2 \log(n)} \quad (11)$$

where  $\sigma_b$  is the standard deviation of the wavelet coefficient and  $n$  the number of linearly independent coefficients, which decreases by a factor 2 (1D) or 4 (2D) with the scale. To reduce the artifacts, the thresholding is then softened by subtracting the threshold  $t$  from the absolute value of the measured coefficient  $w$ . The restored coefficient  $\hat{w}$  is set to 0 if  $|w| \leq t$ .

A redundant wavelet transform based on the *à trous* algorithm (Starck & Bijaoui 1994) was applied. The denoising may take away the orthogonality between the basis images. Therefore, a Schmidt orthogonalization is carried out on them.

### 2.4. Multiscale denoising of the components

A given component  $a_{ni}$  is computed by the scalar product of an image  $V'_n$  with a basis function  $S_i$  considered as denoised. Then  $a_{ni}$  is a random variable with a variance  $\sigma^2$ .

The component set  $A_i = \{a_{in}\}$ ,  $n \in (1, N)$ , is associated to the basis image  $i$ .  $A_i$  can be considered as its spectrum which is a noisy set. It can also be processed using a 1-D wavelet denoising method.

### 2.5. Multiscale denoising of the residuals

Each image  $V'_n$  is restored from the denoised basis images and components. A difference exists between the original and the restored image  $\tilde{V}'_n$  due to:

- The limited number of basis images;
- the basis denoising;
- the component denoising.

The residual  $R_n = V'_n - \tilde{V}'_n$  may contain significant information. It can also be denoised. For this operation two possibilities do exist: i) image per image denoising at given wavelengths ii) spectrum per spectrum denoising at given pixels. In the first operation significant wavelet coefficients are computed from a

pixel environment while it is the spectral neighborhood which is used for the second denoising. This is not completely equivalent. Then the residual cube is denoised by applying successively the two operations. The denoised residuals are added to  $\tilde{V}'_n$  to obtain the final restored image.

## 3. The data set: Multispectral images of NGC 7027

### 3.1. Observations and data processing

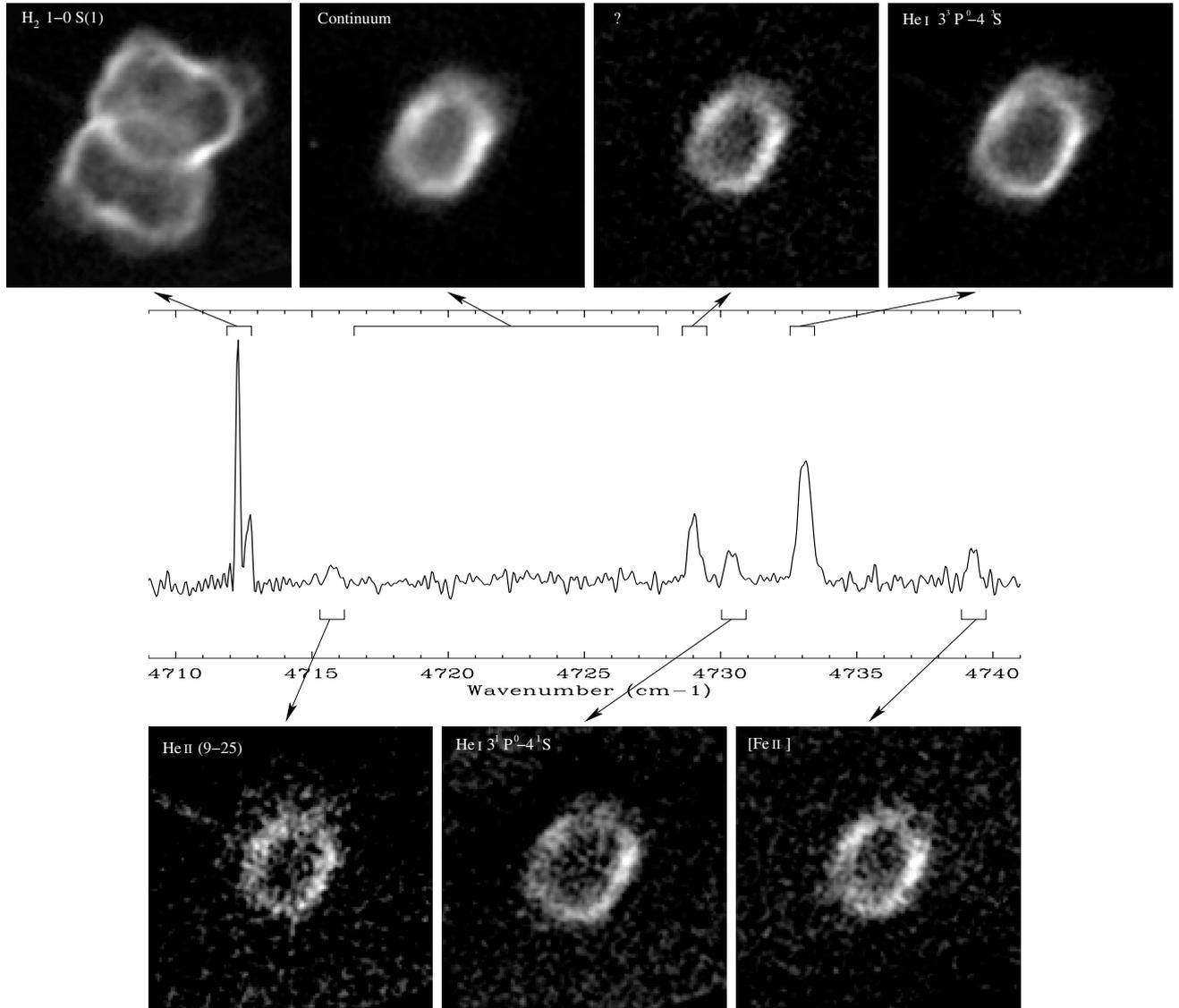
The 3-D data of the young planetary nebula NGC 7027 used in this analysis were obtained on October 24, 1999 during a run with the BEAR imaging spectrometer at the f/35 focus of the 3.6-m CFH Telescope at Mauna Kea (Hawaii). This instrument, based on a dual input/dual output Fourier Transform Spectrometer (FTS) coupled to a  $256 \times 256$  NICMOS camera, simultaneously produces spectra at each point of an extended object over a circular field of view of  $24''$  with a scale of  $0.353''/\text{pixel}$  at seeing-limited spatial resolution ( $\sim 0.5''$ ), in any spectral range between 1 and  $2.5 \mu\text{m}$ . A detailed description of this instrument and its properties have been given by Simons et al. (1994), Maillard (1995) and Maillard (2000).

A narrow-band filter (bandpass  $4705\text{--}4745 \text{ cm}^{-1}$ ) which contains the  $\text{H}_2$  1–0 S(1) line at  $4712.9 \text{ cm}^{-1}$  ( $2.12 \mu\text{m}$ ) in its bandpass was used to reach higher spectral resolution in a minimum number of frames by a strict application of the Shannon sampling theorem. A cube of 1150 planes with an integration time of 9 s per image was obtained, an image being taken at each interferometer step. The optical path difference which was reached corresponds to a spectral resolution of  $0.137 \text{ cm}^{-1}$ , i.e., a velocity resolution of  $8.9 \text{ km s}^{-1}$  which is the maximum spectral resolution obtained in this mode, without being the limit of the instrument. During the observations, the seeing was typically  $0.5''$ . Analysis of these data was presented in Cox et al. (2002).

Prior to the application of the KL analysis a reprocessing of the data was performed, using the package *bearprocess* whose main functions are described in Maillard (2000), plus some additional procedures. The raw frames were spatially oversampled by a factor of 4 to take full advantage of the quality of the seeing and to accurately register each frame relatively to the first one. To fully reconstruct the line profiles and to apply the instrumental phase correction through the field, an over-sampled cube of 840 frames over the useful part of the free spectral range (from  $4700$  to  $4741 \text{ cm}^{-1}$ ) was computed leading to a mean velocity resolution of  $3.1 \text{ km s}^{-1}$ , i.e., an over-sampling by a factor of 2 from the initial cube.

### 3.2. Contents of the data set

Figure 1 shows a typical high resolution spectrum extracted at a bright spot of the NGC 7027 nebula from the  $\text{H}_2$  data cube, and the images in the main emission lines identified on the spectrum. Besides the strong  $\text{H}_2$  line, the spectrum includes the continuum and a series of weaker emission lines which have been identified by Cox et al. (2002) as HeII (9–25) ( $4716.19 \text{ cm}^{-1}$ ), HeI  $3^1\text{P}^0\text{--}4^1\text{S}$  at  $4730.80 \text{ cm}^{-1}$ , HeI  $3^3\text{P}^0\text{--}4^3\text{S}$



**Fig. 1.** Multispectral observations of NGC 7027 using the BEAR imaging Fourier transform spectrometer (Cox et al. 2002). In the middle, spectrum at an emission peak of the ionized region in NGC 7027 as measured in the 4700 to 4741  $\text{cm}^{-1}$  filter. At the top and in the lower rows, images of the nebula in all the detected lines, integrated over their full width, plus image of the continuum emission integrated from 4716 to 4728  $\text{cm}^{-1}$ . This continuum is subtracted from each emission line image. The field of view is of  $22.6 \times 22.6$  arcsec<sup>2</sup>. North is up and East is left.

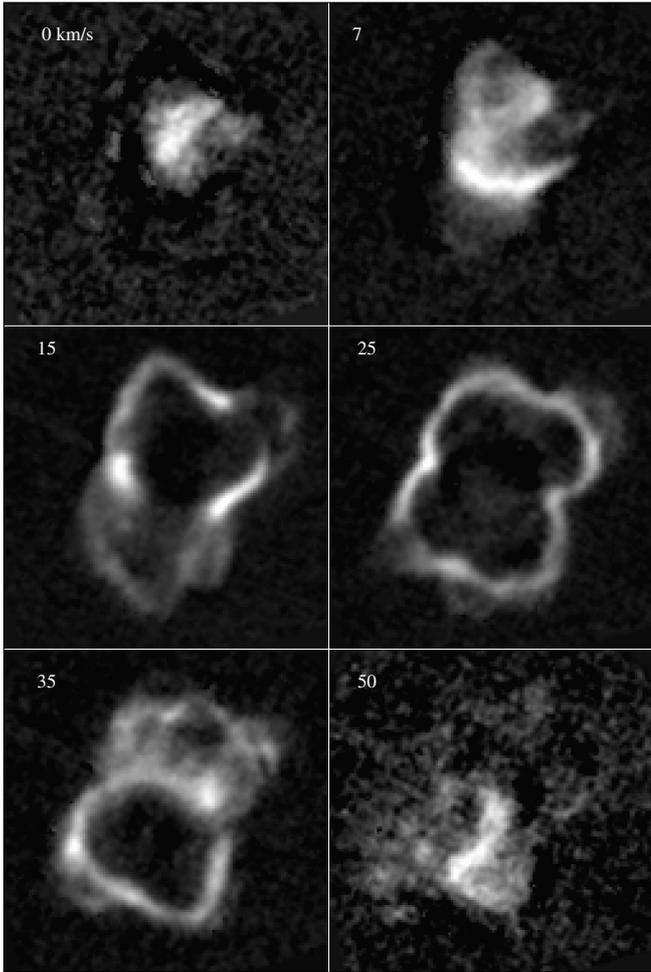
at 4733.52  $\text{cm}^{-1}$ , [FeII] at 4739.72  $\text{cm}^{-1}$  and a line which could be identified as OI at 4729.47  $\text{cm}^{-1}$ . All these lines are spectrally resolved. The distribution and the kinematics of each emission line could be derived together with the spatial distribution of the continuum emission. An important property of these high-resolution spectro-imaging data is that each spectral image contains no other contribution, either from the continuum or from nearby emission lines from the ionized gas. The general aspect of the continuum-subtracted H<sub>2</sub> image is complex, with multiple lobes and weak H<sub>2</sub> emission linking the polar caps along the major axis. This bi-conical shell, which was described in previous studies (Graham et al. 1993; Kastner et al. 1994; Cox et al. 1997; Latter et al. 2000) and recently modeled by Cox et al. (2002), shows holes in the brightness distribution, noticeably to the north-west and south-east where weak H<sub>2</sub> emission extends beyond the main limb-brightened structure. The kinematics of the H<sub>2</sub> emission line is shown

in Fig. 2. At red-shifted velocities ( $v > 25$  km s<sup>-1</sup>), the H<sub>2</sub> emission shifts towards the south while it shifts towards the north at blue-shifted velocities, indicative of a tilted, expanding envelope. The atomic lines trace the inner ionized region showing an elliptical shape similar to the continuum emission distribution.

Figures 1 and 2 show that, despite the high quality of the data set, improvement of the spatial contrast is needed to analyze the faint spatial and spectral structures of the planetary nebula. To increase the signal-to-noise ratio of these data, we have applied the KL method to the data cube resulting from the processing described in this section.

#### 4. KL multispectral data cube analysis

The first step of the KL analysis is the computation of the cross-correlation matrix of the multispectral data cube



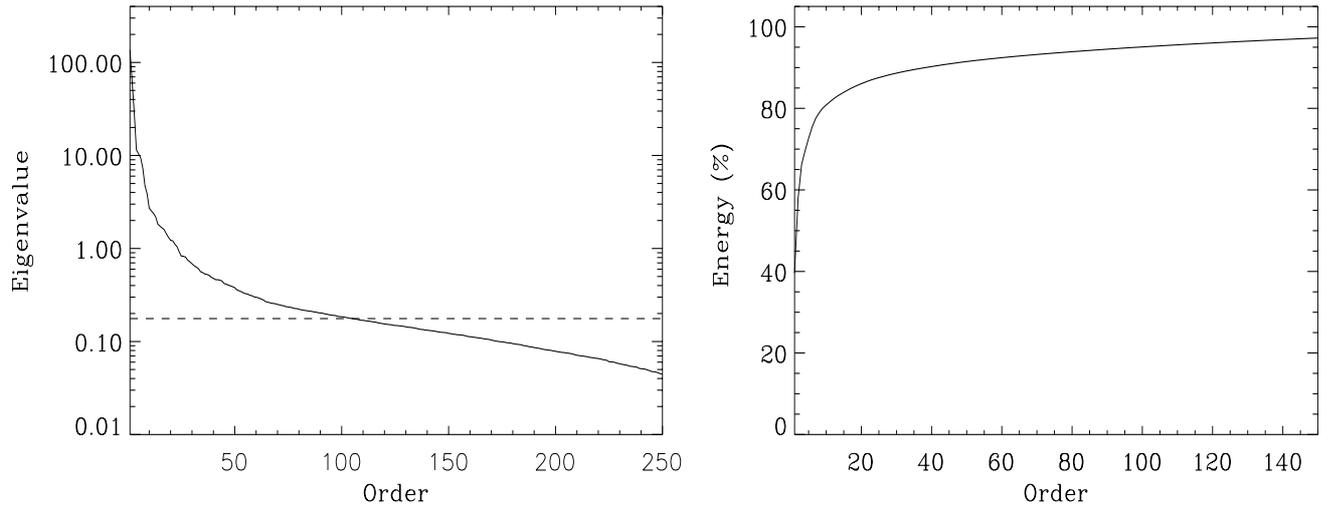
**Fig. 2.** Velocity channels maps (one spectral element of  $3.1 \text{ km s}^{-1}$ ) of NGC 7027 in the 1–0 S(1)  $\text{H}_2$  emission line at  $4712.9 \text{ cm}^{-1}$  ( $2.12 \mu\text{m}$ ). The central velocity ( $v_{\text{LSR}}$ ) is given in the upper left of each panel and the systemic velocity is  $\sim 25 \text{ km s}^{-1}$ . The field of view is  $22.6 \times 22.6 \text{ arcsec}^2$ . North is up and East is left.

( $128 \times 128 \times 840$ ) resulting from the data analysis presented in Sect. 3.1. Then, a diagonalization is carried out.

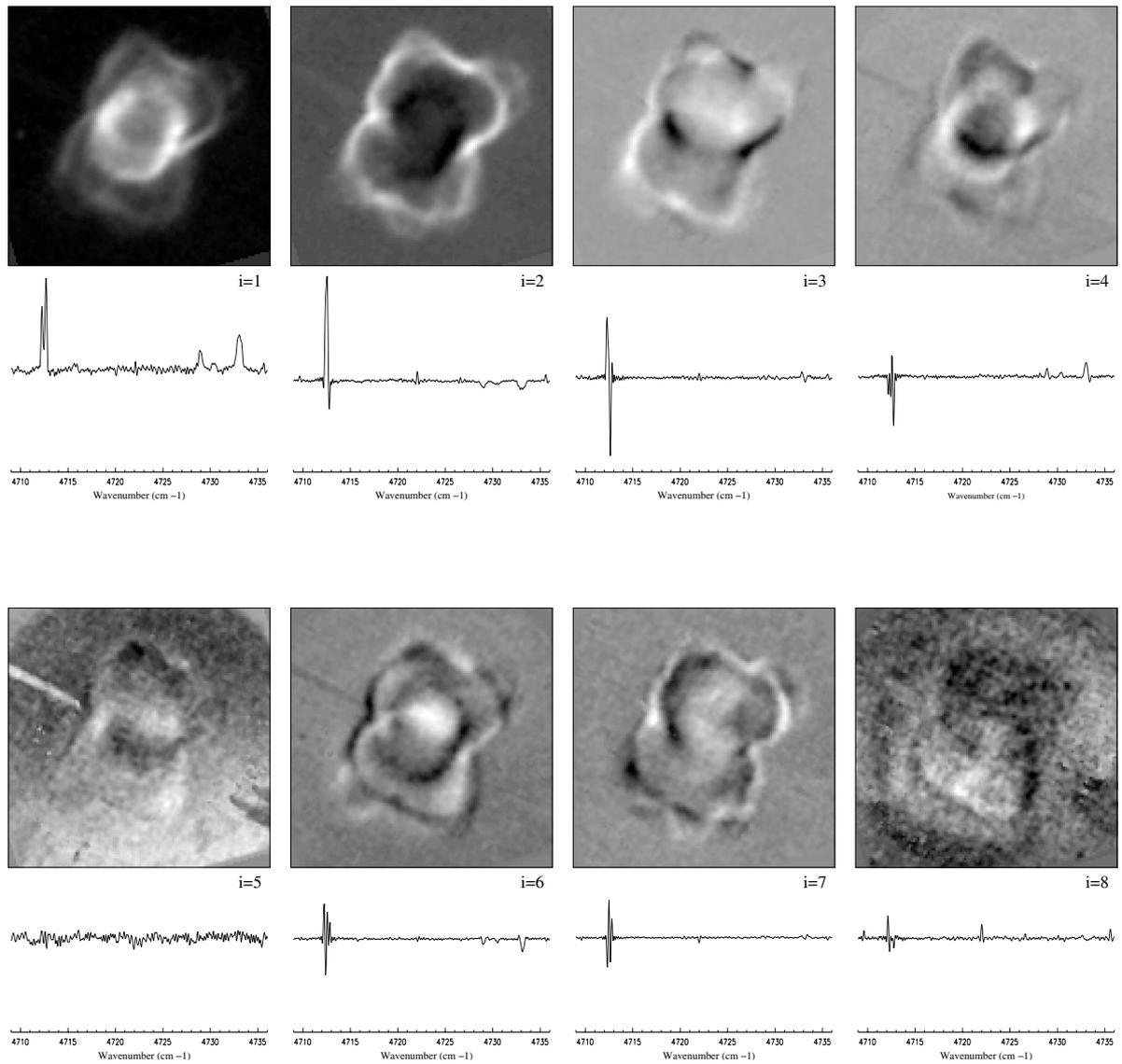
The total number of eigenvectors corresponds to the number of spectral elements of the data cube, i.e. 840. Eigenvalues, sorted in decreasing order, and the corresponding integrated energy distribution are plotted in Fig. 3. The eigenvalues decrease rapidly up to order  $\sim 10$  and then slope down slowly up to order  $\sim 250$ . At this order the eigenvalue is less than 20% of the noise variance which is reached at order  $\sim 100$ . In addition, Fig. 3 shows that 85% of the energy is contained in the first eight eigenvalues while the noise contributes  $\sim 50\%$  of the total energy. To avoid a decomposition on non-significant vectors, the KL expansion has to be truncated before reaching the noise level, i.e. beyond the first 10–20 eigenvalues. In a first step, the process was done using 30 vectors. It appeared that, beyond order  $i = 8$ , basis images did not show visually clear information. Hence, we chose to work with a short expansion, i.e. using only 8 eigenvectors, taking into account that the processing of the residuals between the original and restored images allow us to avoid any loss of significant spatial or spectral information.

These first 8 basis images and their corresponding component sets are displayed in Fig. 4. Component variations of each basis are related to the image index, i.e. the wavenumber variations. The processed sequence contains 840 images of  $128 \times 128$  pixels. Close examination of each basis and its corresponding component variations allow one to understand the decomposition process. The basis image  $i = 1$  represents the “broad-band” image (i.e. integrated over the whole spectral band between  $4705\text{--}4745 \text{ cm}^{-1}$ ).  $\text{H}_2$  and continuum emission is dominant while contributions of other emission lines are less significant. It is a classical representation for KL analysis, where more information is present in the basis number that corresponds to the maximum variance. The basis  $i = 2$  corresponds to the contribution of the  $\text{H}_2$  emission line which is the most prominent line in our spectra while the contributions of other spectral lines are negative. No continuum is present in this basis number and beyond. The basis  $i = 3$  represents the first order of the velocity field over the  $\text{H}_2$  line and a faint contribution of the second prominent emission line (HeI  $3^3\text{P}^0\text{--}4^3\text{S}$ ) of our spectra. Other basis numbers represent high orders of the velocity field over the  $\text{H}_2$  line together with a faint contribution of the other emission lines. In basis  $i = 6$ , the velocity field contribution of HeI, FeII and of the line identified as OI are negative. An artifact is clearly seen in the basis image  $i = 5$  as a bright, linear spike. This artifact, which is also present with a fainter intensity in the basis images  $i = 4$  and  $i = 6$ , is probably due to a local phase shift in the process of the zero path difference (ZPD) map determination which occurred in the standard data reduction presented in Sect. 3.1. The ZPD is determined for each position in the field in order to be able to compute the spectra directly from the cosine FFT of the interferograms. The process starts from the ZPD determination at a continuum intensity peak. The ZPD map is reconstructed step-by-step from this origin over all the pixels of the field, each new determination being used as a best guess of the ZPD position for the next pixel. Phase shifts of half a fringe may occur in regions where only a weak interferometric signal is present. Hence, the corresponding artifact in the KL process does not affect our analysis since no significant signal is present in this region. In addition, no spectral information seems to be carried by the component variations corresponding to the basis image  $i = 5$ . Residual telluric OH lines are present at  $4709.5 \text{ cm}^{-1}$ ,  $4722 \text{ cm}^{-1}$ ,  $4726.7 \text{ cm}^{-1}$  and  $4735.5 \text{ cm}^{-1}$  in the spectral component variations of the basis image  $i = 8$ . Of course, these spectral features, which were not totally subtracted in the data acquisition, are not related to any spectral physical properties of the nebula. In addition, little spatial and spectral information is present in the basis image  $i = 8$  and in its corresponding component variations. Hence, we only have used the first eight basis images in the KL reconstruction process, considering that the basis images and the component variations beyond this value are not necessary to perform a complete data cube restoration.

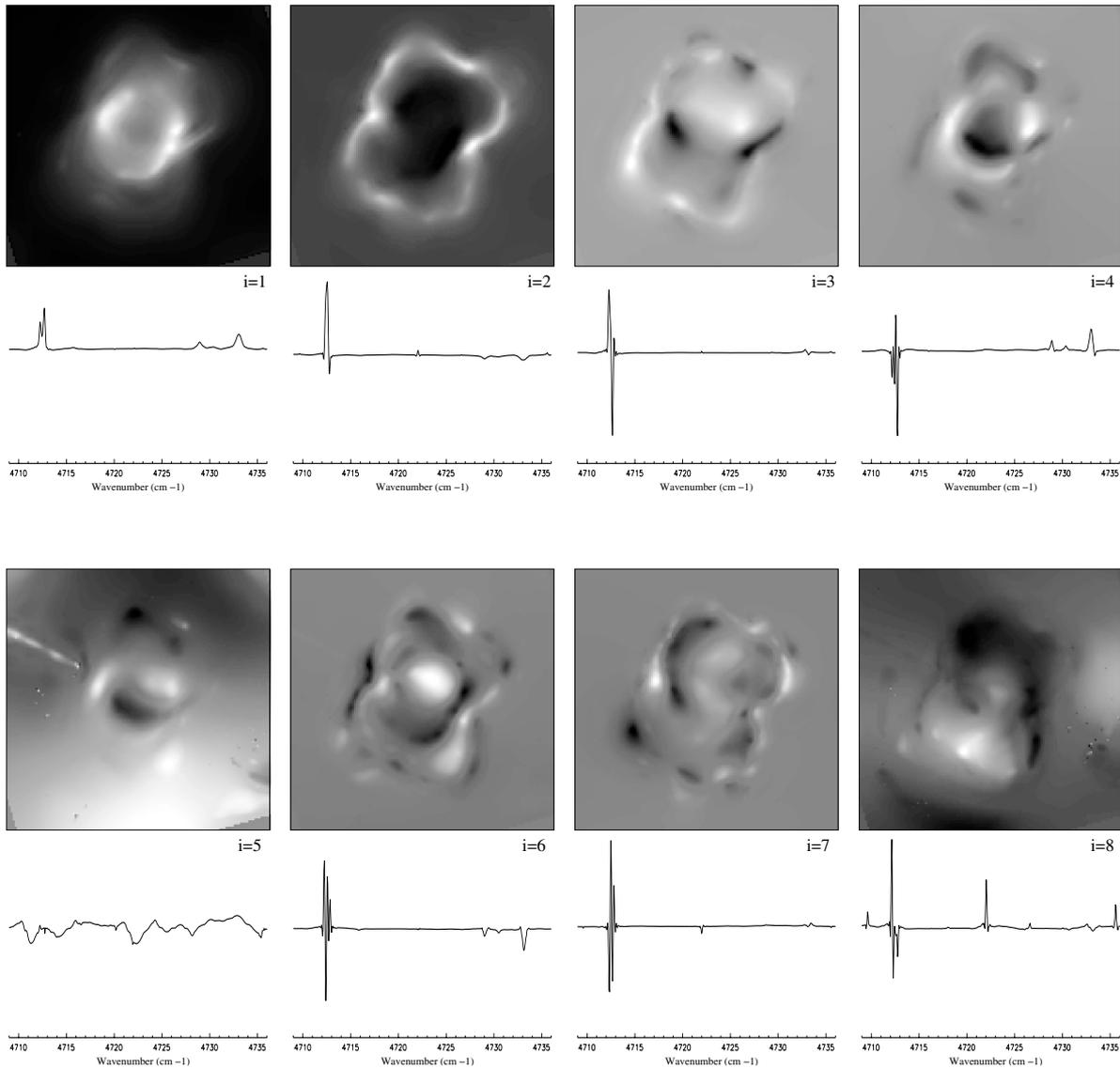
The eight basis images were denoised by the wavelet transform and re-orthonormalized by the Schmidt method. The resulting images and the corresponding denoised component variations are displayed in Fig. 5. We obtain significant enhancement of the signal-to-noise ratio. However, evident loss in spatial resolution due to the drastic wavelet filtering process is



**Fig. 3.** Eigenvalues of the KL expansion. *Left panel:* eigenvalues distribution. *Right panel:* integrated energy distribution. The noise level is represented by the dotted line. Note that the first eight eigenvalues represent 85% of the total energy contained within the KL expansion.



**Fig. 4.** The 8 first basis images and their corresponding component variations.



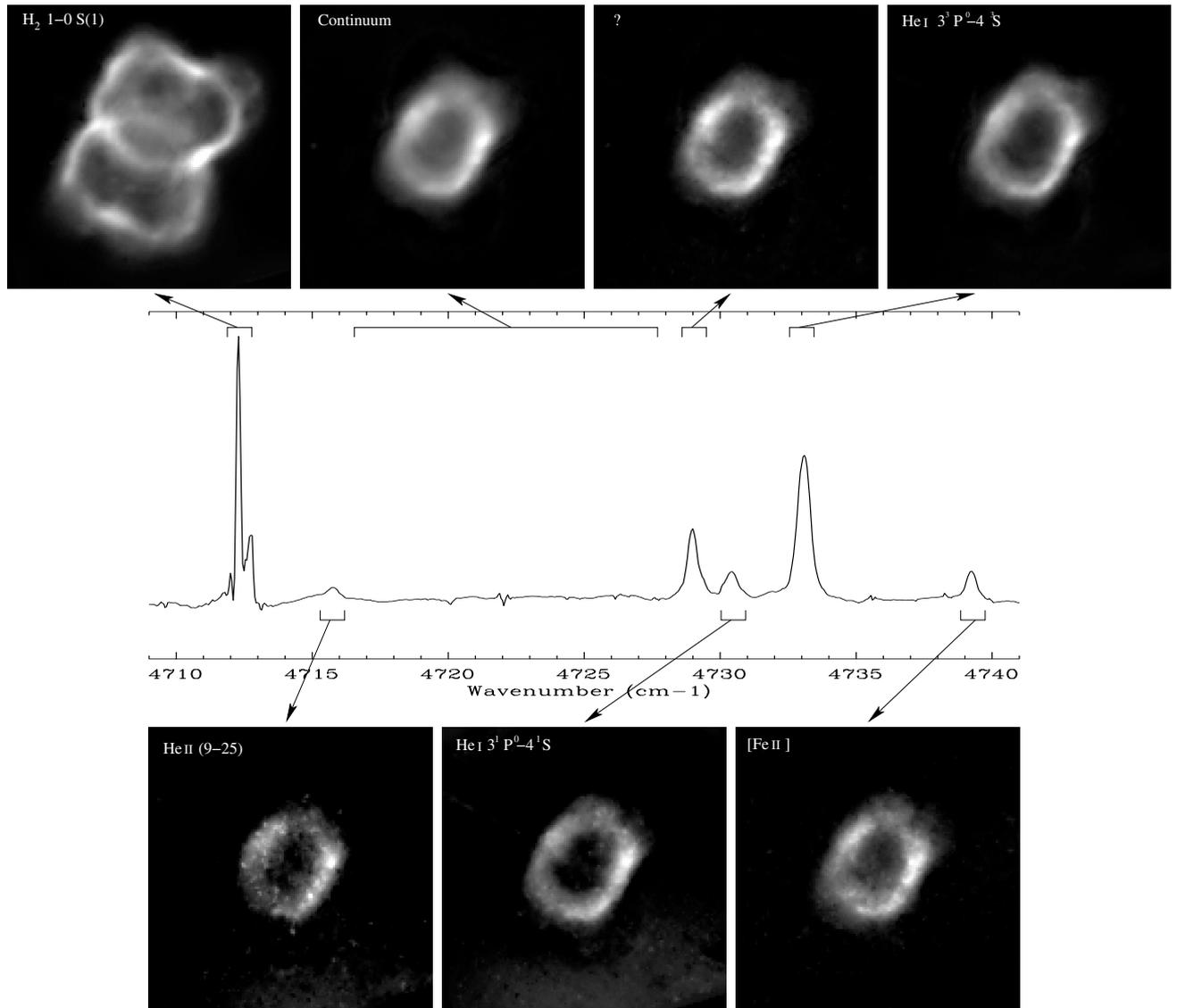
**Fig. 5.** The 8 first basis images and their corresponding component variations after denoising and ortho-normalization.

clearly seen in these images. The filtering we applied is needed to perform a noiseless KL first approximation of the process and to retain first only the significant information present in the data cube. The remaining information is contained in the residual data cube which is the difference between the original and the KL-restored data cube by using the first eight basis images. This data cube, which contains significant information, is denoised image per image and then added to the first KL approximation of the reconstructed data cube. Taking into account the chosen threshold of the wavelet filtering process only a small amount of residual noise remains.

The KL reconstructed images corresponding to Figs. 1 and 2 are displayed in Figs. 6 and 7. These images seem to be smoothed compared to the previous ones. The signal-to-noise ratio of these images, made at the peak of the emission lines or extracted at various positions in the  $\text{H}_2$  line profile, is increased because the signal-to-noise ratio of all the spectra in the new cube has been increased by a factor of  $\sim 4$ . This factor is estimated by taking the ratio of the rms noise in the two

spectra from the same position in the field, before and after the KL analysis. As an example, Fig. 8 shows the comparison between a low signal-to-noise-ratio spectrum measured towards one pixel, at a position of faint emission in the ionized zone of NGC 7027 and the KL restored spectrum. This illustration clearly shows that the KL process does not modify either the line positions and profiles or significantly the peak intensity. If the line positions were changed, a negative and positive peak typical of a line shift would appear in the difference. No such effect is detected. Consequently, the kinematics of the gas traced by the line positions is fully conserved after the KL analysis. Table 1 presents the minor changes of intensities and linewidths between the original and the reconstructed spectrum.

It is interesting to note that features retained by this process do not result from one image per image analysis, but from the use of information contained in the whole data cube. Taking into account the correlation between the images improves the quality of the restoration even if a residual is re-introduced



**Fig. 6.** KL reconstruction of the multispectral observation of NGC 7027 presented in Fig. 1. Significant improvement of the signal-to-noise ratio is obtained. The rms noise of the spectrum has been divided by a factor of  $\sim 4$  (see text).

after the KL process. Another illustration is given by Fig. 9 which shows a comparison between a single-element spectral image of the continuum taken from the raw data cube set and its corresponding KL image. The signal-to-noise ratio is highly increased after the KL reconstruction. By measuring the rms noise on both images outside the emission region the gain in signal-to-noise ratio is equal to 16. On the other hand, comparison between a continuum integrated image (from 4716 to 4727  $\text{cm}^{-1}$ ) before and after the KL process (Fig. 10) does not show a similar enhancement as in the case of a single element spectral image, which is natural, since the integrated image already used a large fraction of the information on the continuum component contained in the original data cube.

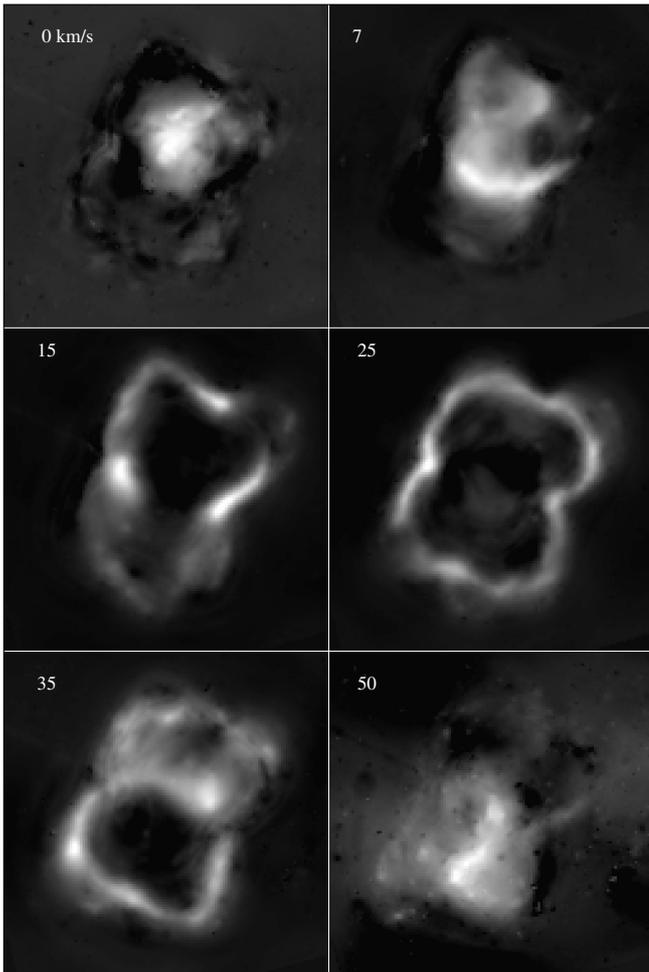
## 5. Discussion

We have proposed a method which improves the spatial and spectral contents of an imaging spectroscopy data set. This

method uses spatial and spectral correlations among adjacent pixels, and is particularly effective in the case of a large number of spectral images.

We show that a 3D-data set can be restored with a small number, eight in our case, of basis images. Only a small fraction of spectral and spatial information is present beyond this position. This information is not lost because it is present in residuals which are filtered and added to the restored KL data cube. Working with a larger number of basis images leads to the introduction of noise which degrades the final signal-to-noise ratio.

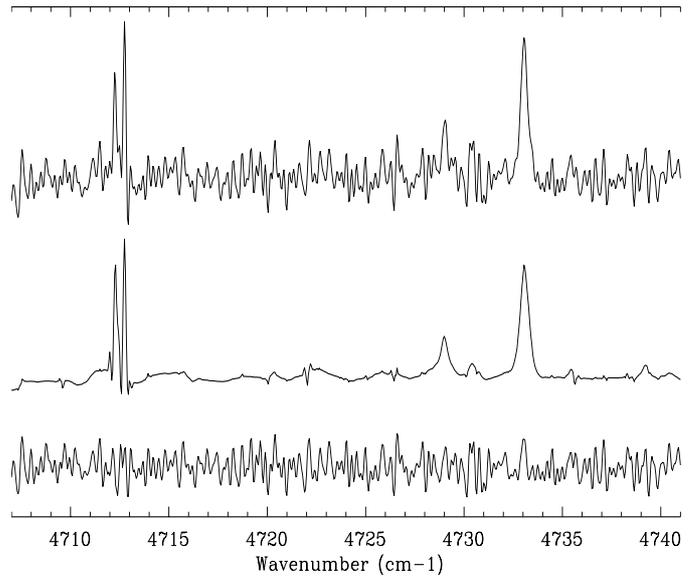
The KL method is more efficient in the case of a low signal-to-noise ratio where an improvement by a factor of 4 is obtained, as illustrated in Figs. 8 and 9. In this case, the restoration process takes advantage of the redundant information contained in the whole data set. As a consequence, analysis at the optimum spatial resolution offered by the data becomes possible. There is no need to average the signal of several



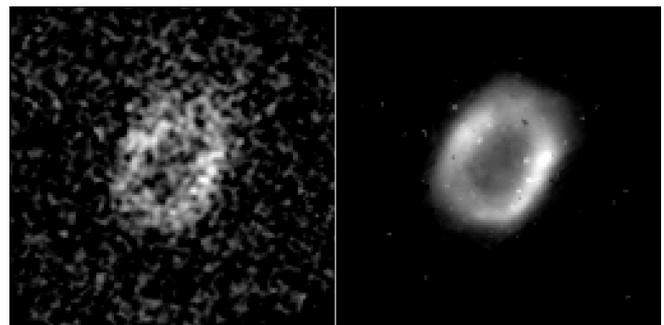
**Fig. 7.** Velocity channel maps in the 1-0 S(1) H<sub>2</sub> emission line at 4712.9 cm<sup>-1</sup> (2.12 μm) after application of the KL reconstruction. These images have to be compared to those of Fig. 2. Note the significant improvement of the contrast of faint structures.

pixels to increase the signal-to-noise ratio of the images. On the other hand there is no significant gain when the signal-to-noise ratio is high because of the small contribution of the redundant information. A drastic wavelet filtering allows one to begin the KL process with noiseless initial information. The data set contains images that have variable signal-to-noise ratio. For high signal-to-noise ratio, e.g. the H<sub>2</sub> emission line, the KL process is insufficient and the contribution of the residual is high. Nevertheless, for low signal-to-noise-ratio images, the whole information is contained in the KL process and the residual contribution is low. Note that the redundant information is present both in spectral and spatial space and contributes to increase the signal-to-noise ratio of faint spectro-imaging data.

Starck et al. (2001) proposed another way of coupling the KL expansion with the wavelet transform. First, the wavelet transform is separately applied to each image. Second, KL is done scale-by-scale. Each KL basis is then smoothed taking its corresponding noise. The original images are restored scale-by-scale taking into account the denoised KL basis. In our approach the KL expansion is useful only for getting an image quite acceptable but fully denoised. The lost information is



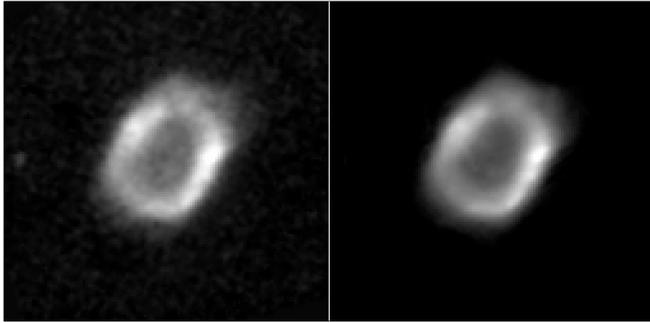
**Fig. 8.** Spectrum measured towards one pixel position of a faint emission of the ionized region in NGC 7027 (upper signal) and its corresponding KL reconstruction. The bottom line is the difference between the original and the KL reconstructed signals.



**Fig. 9.** One spectral element image in the continuum taken in the data cube at spectral position 4715.48 cm<sup>-1</sup> (left) and its corresponding KL image (right). The signal-to-noise ratio in the KL image is highly increased (see text).

contained in residuals and it is mainly restored after their denoising. In order to get sufficiently smoothed images after the KL reconstruction, few components are used. Increasing the number of components leads to an increase of the noise in all images reconstructed by KL. The daily experiments show that beyond a given order, the KL components only carry information for few pixels of few images. By construction, residuals from one image to another are quite de-correlated. Thus, their denoising needs an individual wavelet transform process. In the case of multicolor imaging, for which few images are to be processed, all KL components are used for the reconstruction. The process of Starck et al. (2001) is then useful. For integral field spectroscopy, the size is too large. Few components are computed which makes the proposed procedure necessary.

The KL decomposition allows one to detect some spectral parasitic lines and spatial artifacts which are not related to any spatial or spectral properties of the observed object (e.g. Figs. 4 and 5). These features could be removed from the whole data



**Fig. 10.** Comparison between a continuum integrated image (from 4716 to 4727  $\text{cm}^{-1}$ ) before and after the KL process. No significant improvement is obtained in KL process of high signal-to-noise ratio image.

**Table 1.** Peak intensity and linewidth ratios of a few lines from the original spectrum and the KL reconstructed spectrum shown in Fig. 8.

Line	Peak intensity ratio	Linewidth ratio
H <sub>2</sub> red-shifted component	1.0	1.2
H <sub>2</sub> blue-shifted component	0.9	1.0
OI at 4729.47 $\text{cm}^{-1}$	0.8	1.5
HeI at 4733.52 $\text{cm}^{-1}$	0.8	1.2

set thanks to the application of the Independent Component Analysis (ICA) process. In this case, the KL analysis is used as the first step of the whole process.

## 6. Conclusion

We have examined a method, based on the Karhunen-Loève (KL) transformation, to improve the signal-to-noise ratio of a multi-spectral data cube. We have applied this analysis on near-infrared multi-spectral images of the young planetary nebula NGC 7027. Significant improvement of signal-to-noise ratio in the spectra and the images is obtained, particularly for low signal-to-noise ratio data, for which an improvement by a factor of 4 is obtained, while for high signal-to-noise ratio images no improvement is obtained. The efficiency of the KL analysis depends on the quality of information contained in the data set. The higher the quality, the better the determination of the KL basis and the reconstruction of faint structures present in the data set. However, the KL analysis did not contribute to the detection of new spectral or spatial features in the NGC 7027 data set, which was already a good signal-to-noise ratio data cube. The same analysis should be attempted on a similar data set of a weak object, containing low signal-to-noise-ratio information, to demonstrate the potential capability of detecting new spectral and spatial features. In addition, we

have shown that the KL decomposition process makes it possible to detect spatial residual artifacts and parasitic lines which can be removed from the data. Removing these artifacts from data could be done thanks to the application of the Independent Component Analysis (ICA). This application will be examined in a next paper.

*Acknowledgements.* The authors wish to thank the anonymous referee for many helpful comments and suggestions. Part of this work has been performed using the computing facilities provided by the program “Simulations Interactives et Visualisation en Astronomie et Mécanique (SIVAM)”.

## References

- Bijaoui, A. 1974, *A&A*, 30, 199  
 Bijaoui, A., & Doazan, V. 1979, *A&A*, 70, 285  
 Biegging, J. H., Wilner, D., & Thronson, H. A. Jr. 1991, *ApJ*, 379, 271  
 Coifman, R. R., & Donoho, D. 1995, Translation invariant denoising, Technical report 475, Dpt of Statistics, Stanford University  
 Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., & Calzetti, D. 1995, *AJ*, 110, 1071  
 Cox, P., Maillard, J.-P., Huggins, P. J., et al. 1997, *A&A*, 321, 907  
 Cox, P., Huggins, P. J., Maillard, J.-P., et al. 2002, *A&A*, 384, 603  
 Cox, P., Huggins, P. J., Maillard, J.-P., et al. 2003, *ApJ*, 586, 87  
 Donoho, D. L., & Johnstone, I. M. 1994, *Biometrika*, 81, 425  
 Graham, J. R., Herbst, T. M., Matthews, K., et al. 1993, *ApJ*, 408, 105  
 Kastner, J. H., Gatley, I., Merrill, K. M., Probst, R., & Weintraub, D. 1994, *ApJ*, 421, 600  
 Karhunen, K. 1947, *Ann. Acad. Scie. Fenn. Ser. A1: Math. Phys.*, 37  
 Latter, W. B., Dayal, A., Biegging, J. H., et al. 2000, *ApJ*, 539, 783  
 Loève, M. 1960, *Probability theory*, Van Nostrand, Princeton, 478  
 Maillard, J.-P. 1995, in *3-D Optical spectroscopy methods in astronomy*, ed. G. Comte, & M. Marcellin, *IAU Coll.*, 149, *ASP Conf. Ser.*, 71, 316  
 Maillard, J.-P. 2000, in *Imaging the Universe in 3 Dimensions*, ed. E. van Breughel, & J. Bland-Hawthorn, *ASP Conf. Ser.*, 195, 185  
 Mallat, A. 1999, *A wavelet tour of the signal processing* (New-York: Academic Press)  
 Murtagh, F., & Heck, A. 1987, *Multivariate data analysis*, 13 (Dordrecht: Reidel)  
 Paumard, T., Maillard, J. P., Morris, M., & Rigaut, F. 2001, *A&A*, 366, 466  
 Pelat, D. 1974, *A&A*, 33, 321  
 Péquignot, D. 1997, *IAUS*, 180, 167  
 Ren, D., & Allington-Smith, J. 2002, *PASP*, 114, 866  
 Ryder, S. D., Sun, Y.-S., Ashley, M. C. B., et al. 1998, *PASA*, 15, 228  
 Simons, D. A., Clark, C. C., Smith, S., et al. 1994, in *Instrumentation in Astronomy VIII*, *Proc. SPIE*, 2198, 185  
 Starck, J. L., & Bijaoui, A. 1994, *Signal Process.*, 35, 195  
 Starck, J.-L., Murtagh, F., & Bijaoui, A. 1998, *Image Processing and Data Analysis, The Multiscale Approach* (Cambridge: Cambridge University Press), 21  
 Starck, J.-L., Murtagh, F., Querre, P., & Bonarel, F. 2001, *A&A*, 368, 730  
 Tegmak, M., Taylor, A. N., & Heavens, A. F. 1997, *ApJ*, 480, 22  
 Vogeley, M. S., & Szalay, A. S. 1996, *ApJ*, 465, 34