

# Singular value decomposition: A tool to separate elementary contributions in ISOCAM\* spectral maps

P. Boissel<sup>1</sup>, C. Joblin<sup>2</sup>, and P. Pernot<sup>1</sup>

<sup>1</sup> Laboratoire de Chimie Physique, Université Paris Sud, Bât. 350, 91405 Orsay Cedex, France

<sup>2</sup> Centre d'Étude Spatiale des Rayonnements, CNRS-UPS, BP 4346, 31028 Toulouse Cedex 04, France

Received 18 January 2001 / Accepted 7 May 2001

**Abstract.** Singular value decomposition (SVD) is applied to a mid-infrared ISOCAM spectral map of NGC 7023. As a first result, this decomposition provides a mathematical analysis of the map in terms of a linear combination of elementary spectra. After further processing, it is shown that the spectrum observed on each pixel can be described as the physical superposition of four components. Three of them can be attributed to the intrinsic spectra of polycyclic aromatic hydrocarbons (PAH), very small grains and larger dust grains respectively. The fourth component is a differential spectrum that could trace the ionisation state of PAHs.

**Key words.** methods: data analysis – ISM: dust – ISM: molecules – infrared: ISM: lines and bands – infrared: ISM: continuum – ISM: individual objects: NGC 7023

## 1. Introduction

The aromatic infrared bands (AIBs) at 3.3, 6.2, 7.7, 8.6, 11.3 and 12.7  $\mu\text{m}$  have been observed for more than 20 years. Although their attribution to polycyclic aromatic hydrocarbon species (PAH) is widely accepted, the detailed nature of their carriers is not precisely known. Thanks to the instruments on board the European Infrared Space Observatory (ISO), a lot of information is now available to get a deeper insight into this question. In particular, the combined imaging and spectroscopic capabilities of the camera ISOCAM can be used on extended objects to study the spatial distribution of the different AIBs and their evolution with the physical conditions.

A limitation to these studies is that the amount of data contained in a single map is already very large and not easy to handle. The simpler analysis technique consists in mapping the relative band intensities and the band-to-continuum ratio. The decomposition of the observed spectra using Lorentzian functions is an improved version of this technique, allowing a more precise and less operator-dependent evaluation of the integrated intensities (Boulanger et al. 1998). However, in these approaches, the data are taken on a line-by-line basis, losing sight that a full spectrum is required to characterize a given species.

It would be more appropriate to consider the spectrum derived for each pixel as a superposition of elementary spectra associated to different carriers. The spatial variations can then be interpreted in terms of variations of the abundances induced by changing physical conditions (UV flux, density, ...). Access to these elementary spectra is therefore fundamental to derive information on the photo-physical and chemical properties of the various emitters.

Laboratory experiments using time-resolved spectroscopy are often faced with the same problem of separating elementary component spectra. In that case, singular value decomposition (SVD), also called principal component analysis (PCA), has proven to be a valuable tool (Henry & Hofrichter 1992). In astronomy, this technique was applied to galaxy spectra (Folkes et al. 1999). When a set of  $m$  spectra measured at  $n$  wavelengths is combined in a  $m \times n$  matrix  $A$ , SVD provides an estimate of the number of independent contributions that are needed to describe the rows as linear combinations of separate spectra. Although not directly related to physical components, these individual spectra provide a mathematical reduced basis which is useful to compare different sets of data. Further analysis of this basis leads to spectra that can be attributed to physical species.

In the present work, SVD is applied to an ISOCAM map of the reflection nebula NGC 7023. After a presentation of the method, the direct results of the decomposition are given in Sect. 4. We show then that four distinct contributions can be separated in the spectra, each of them having a clear physical meaning as discussed in Sect. 6.

Send offprint requests to: P. Boissel,  
e-mail: pierre.boissel@lcp.u-psud.fr

\* Based on observations with ISO, an ESA project with instruments funded by ESA Member States (especially the PI countries: France, Germany, The Netherlands and the UK) with the participation of ISAS and NASA.

## 2. SVD and its application to ISOCAM data

Assuming that the emission originates from a relatively small number of species  $k$ , the spectrum observed on a given pixel  $p$  of the map is the superposition of  $k$  elementary spectra  $e_{i\lambda}$ , weighted by a set of coefficients  $c_{pi}$ . The observed  $m \times n$  matrix  $A$  is then the product of the  $m \times k$  matrix of weight coefficients  $C$  by the  $k \times n$  matrix  $E^T$  whose rows are the elementary spectra  $e_i$ .

$$A = C \times E^T. \quad (1)$$

From a purely mathematical point of view, using SVD,  $A$  is decomposed as the product:  $A = U \times S \times V^T$  where  $U$  ( $m \times n$ ) and  $V$  ( $n \times n$ ) are orthogonal normalized matrices (i.e.  $U \times U^T = I$  and  $V \times V^T = I$ ) and  $S$  a diagonal  $n \times n$  matrix whose elements, the singular values, are non-negative and sorted in decreasing order. This decomposition appears closely related to Eq. (1) if the product  $U \times S$  is gathered in a single matrix that we call  $US$ . The rows of  $V^T$ , that are the equivalent of the elementary spectra, are called SVD vectors.

An important property of SVD is that the number of non-zero elements of  $S$  gives the rank of the matrix. If  $A$  results from linear combinations of only  $k$  independent spectra, all the singular values  $s_i$  with  $i > k$  will be zero in an ideal noiseless observation. When noise is added, an estimate of  $k$  can still be obtained by comparing the noise level and the standard deviation obtained when the map is reconstructed with a SVD set truncated at rank  $k$ .  $V^T$  is then reduced to  $V_r^T$  containing the first  $k$  rows and  $US$  to  $US_r$  with  $k$  columns.  $A$  is now the product of a  $m \times k$  matrix by a  $k \times n$  matrix as in Eq. (1).

$$A = US_r \times V_r^T. \quad (2)$$

The set of spectra measured on the ISOCAM pixels (rows of  $A$ ) can then be described in a  $k$  vector basis, using either the rows  $V_i$  of  $V_r^T$  or the rows  $E_i$  of  $E^T$ .

## 3. Input data

The main goal of this paper is to test the relevance of the SVD method for the analysis of the AIB spectrum. At this stage, it is not necessary to get the best reduced data and we have directly retrieved ISOCAM maps from the ISO archive. The ‘‘Quick Look’’ products were used without further processing except the usual linear interpolation for the missing Col. 24. The pixels whose spectrum is obviously aberrant after visual inspection were masked.

SVD has been tested on several objects and detailed comparative results will be presented in a forthcoming paper. We focus here on the ISOCAM map of NGC 7023 (Cesarsky et al. 1996; Ref. 10702605 in the ISO archive). NGC 7023 is a well-studied bright reflection nebula where the AIBs are strong. It can be simply described by two photodissociation regions viewed edge-on that are created by the interaction of the central Herbig Be star HD 200775 with its parent cloud (Gerin et al. 1998 and references

**Table 1.** Singular values  $S$  and standard deviation  $\sigma$  as function of the rank  $k$ .  $\sigma$  in units of  $10^{-3}$  Jansky/pixel<sup>2</sup>.

$k$	$S_k$	$\sigma(k)$	$k$	$S_k$	$\sigma(k)$	$k$	$S_k$	$\sigma(k)$
0		138.0	6	1.52	6.1	12	0.46	4.0
1	55.97	25.0	7	1.11	5.5	13	0.44	3.9
2	8.47	14.4	8	0.99	5.0	14	0.38	3.8
3	3.72	11.3	9	0.80	4.6	15	0.35	3.7
4	3.13	8.4	10	0.61	4.3	16	0.34	3.6
5	1.78	7.2	11	0.50	4.2			

therein). This is the best situation to follow the photo-physical evolution of the various species when going from the more irradiated regions close to the star to the more shielded environment of the molecular cloud.

## 4. Analysis of SVD vectors

The set of dominant singular values is reported in Table 1 as a function of the rank  $k$ . The value of  $\sigma(k)$  is also given. It is the root mean square of the difference between the observed map and the map which is calculated with a set of SVD vectors truncated at the rank  $k$ . It is expressed in flux units and can be directly compared to the value for the brightest non-masked pixel: 1.70 Jansky/pixel<sup>2</sup>. Looking at the list of values, a rapid decrease is observed up to rank 4, then the improvement is slower down to a value of  $3.5 \times 10^{-3}$ , which can be considered as the noise level. The reconstruction of the data with only four vectors leads to a standard deviation of  $\sigma(k) = 8.3 \times 10^{-3}$ , which is close to twice the rms noise. Adding six more vectors up to rank 10 gives a nearly perfect reconstruction.

Knowing the SVD vectors, the entire map can then be described with a good accuracy by a set of four to ten numbers by pixel. This is a drastic simplification compared to a description by a set of lorentzian profiles (Boulanger et al. 1998) which requires eighteen plus two parameters to fit six bands and the underlying continuum. In such case, a fully automatic treatment is generally not possible because of convergence problems. On the opposite, starting with a given set of SVD vectors, the weight coefficients needed to reproduce a given spectrum are readily found by a simple linear regression. This allows to use the same set of vectors not only for the map on which the decomposition has been performed, but also for any map dominated by the AIBs, as will be shown in a forthcoming paper.

Figure 1 presents the spectra of the SVD vectors  $V_1$  to  $V_4$ . As can be seen,  $V_1$  gives an average AIB spectrum. Its contribution to the map is largely dominant. It is shown in gray scale in Fig. 2 and is directly comparable to the maps of the 6.2 and 11.3  $\mu\text{m}$  bands (Figs. 2 and 3 of Cesarsky et al. 1996).

Analysis of the physical meaning of the other vectors is hindered by the fact that SVD is only a mathematical decomposition. Due to the orthogonality condition, all vectors must exhibit positive and negative components, excluding an interpretation in terms of emission spectra. However, owing to the hierarchy in the singular values,

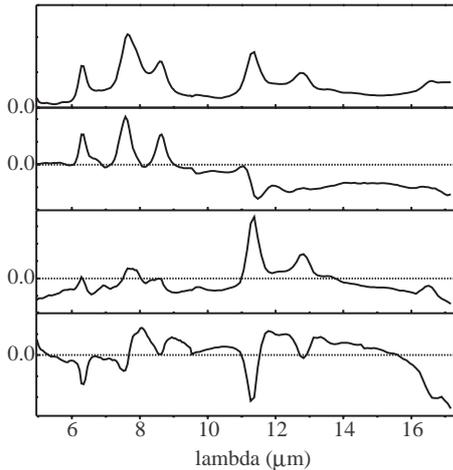


Fig. 1. SVD vector spectra  $V_i$ ,  $i = 1$  to 4 from top to bottom.

they can be viewed as variations of the average spectrum given by  $V_1$ .

$V_2$  exhibits a negative continuum at wavelengths longer than  $10 \mu\text{m}$ . The bands at  $6.2$ ,  $7.7$  and  $8.6 \mu\text{m}$  appear with positive amplitudes and a reduced width compared to  $V_1$ . A negative weight of this vector corresponds then to an increase of a  $10$ – $16 \mu\text{m}$  continuum and a decrease of the  $6$ – $9 \mu\text{m}$  bands. Looking at the map in Fig. 2, this happens in regions located deeper inside the molecular cloud compared to the maximum of  $V_1$ . On the opposite, positive weights observed closer to the star are associated with an increase and a sharpening of the bands and a decrease of the continuum. This result, already mentioned by Cesarsky et al. (1996), naturally comes out of SVD. Its meaning will appear more precisely in the next section. In particular, we will see that the origin of the continuum emitted in the close vicinity of the star differs from that of the continuum present in the molecular cloud.

## 5. Separation of physical components

Given the assumption that the emission originates from  $k$  distinct species, the ultimate goal is to find the transformation matrix  $T$  linking the elementary spectra  $E_i$  to the SVD vectors ( $E^T = T \times V_r^T$ ). The matrix  $T$  gives simultaneously the spectra and the maps of their weight coefficients using the inverse matrix  $T^{-1}$  ( $C = US_r \times T^{-1}$ ). With normalized vectors,  $k^2 - k$  coefficients have to be determined. One straightforward constraint is that all the elements of  $E$  and  $C$  must be positive since emission spectra are observed. Finding the set of coefficients that insures positivity is a non trivial problem that is currently under investigation. Likely, this problem does not have a unique solution and other criteria have to be implemented. Fortunately, the decreasing importance of the successive SVD vectors makes possible a step-by-step procedure by progressively increasing the number of parameters as described below. In the present work, the matrix  $T$  has been built by manually tuning the parameters and simultaneously observing the effect on the maps and on the spectra.

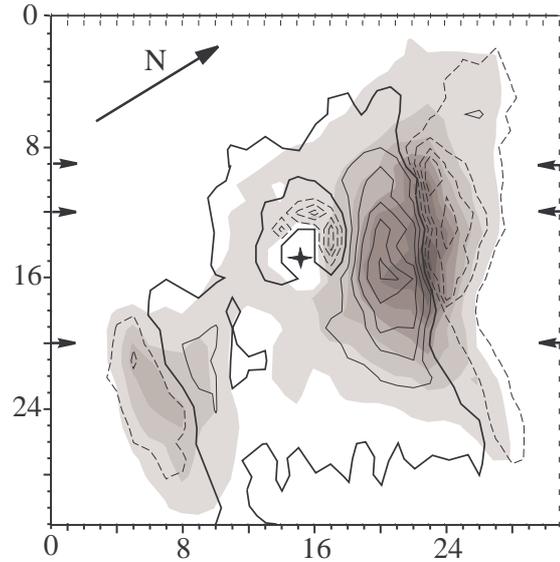
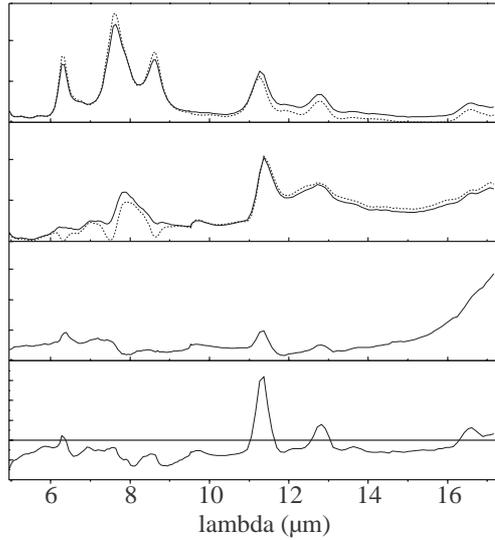


Fig. 2. Map of the weight of SVD vectors:  $V_1$  (gray scale; white = 0) and  $V_2$  (contours; solid, bold and dashed lines are for positive, zero and negative values respectively). Axes are pixel numbers (PFOV =  $6''$ ). Saturated pixels around the star have been set to zero. Arrows mark the cuts given in Fig. 4.

The contribution of the dominant vectors  $V_1$  and  $V_2$  was first considered. A reconstruction in this subspace already accounts for 90% of the data. It is very tempting to separate the bands (the molecular contribution) from the continuum. This result happens to come out of the calculations with a very low level of uncertainty. The positivity constraint leads to a small allowed range for the two independent coefficients of  $T$ . The spectra  $E_1$  and  $E_2$  calculated with the extreme values are reported in Fig. 3 as solid and dashed lines. As can be seen the overall shape of both spectra is fully determined. Their spectral features can be only slightly modified by mixing with lower rank vectors.

Keeping  $E_1$  and  $E_2$  fixed, we can increase the subspace to dimension 4 in order to include the contribution of  $V_3$  and  $V_4$ . As can be seen in Fig. 1, both of them exhibit the C-H out-of-plane bending bands at  $11.3$  and  $12.7 \mu\text{m}$  and a long-wavelength rise is observed in  $V_4$ . A linear combination of these two vectors leads to the separation of the rising continuum in  $E_3$  from the C-H bands in  $E_4$ . Further combination of these spectra with  $V_1$  and  $V_2$  is then performed in order to maximize both the spectral and spatial decorrelations. From a spectral point of view, the intensities of the  $6.2$  and  $7.7 \mu\text{m}$  bands were minimized in  $E_3$  and  $E_4$ . The last one then becomes a pure difference spectrum tracing the intensity of the C-H out-of-plane bending modes relative to the other components. From a spatial point of view, adding a proper fraction of  $V_2$  to  $E_3$  leads to a complete spatial decorrelation between  $E_3$  and  $E_2$ . This is illustrated by the cuts displayed in Fig. 4:  $E_3$  is restricted to the vicinity of the star (cut  $y = 12$ ) whereas  $E_2$  peaks inside the molecular cloud.

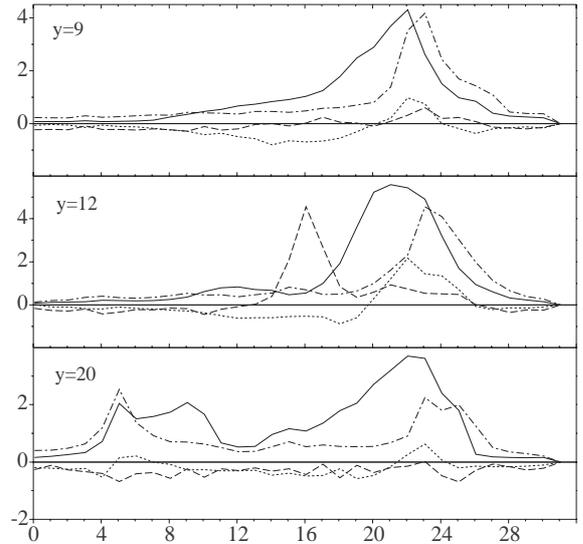


**Fig. 3.** Spectra of the four physical components derived from the SVD vectors shown in Fig. 1 and proposed identification of the carriers. From top to bottom:  $E_1$  (PAHs) and  $E_2$  (VSGs) with solid and dashed lines for two extreme values in the calculations;  $E_3$  (Dust) and  $E_4$  (differential spectrum related to the charge state of PAHs).

## 6. Discussion and perspectives

This paper presents the use of SVD and more generally of global decomposition methods to analyse the AIB spectrum. These methods appear to be very promising as illustrated by the following preliminary results. First, two dominant contributions to the emission could be separated. The main one,  $E_1$ , is a pure band spectrum without continuum and can be attributed unambiguously to molecular species such as PAHs. The second one,  $E_2$ , is representative of the generally called “underlying continuum” attributed to larger species. An important result of this paper is that the structure of this spectrum is revealed. Broad bands appear in the 8.0 and 11–13  $\mu\text{m}$  regions although their detailed shape is not yet fully constrained. The maximum emission is located quite deep inside the cloud which implies that the carriers cannot be at thermal equilibrium. Stochastic heating of very small particles such as the very small grains (VSGs) first introduced by Désert et al. (1990), has therefore to be invoked. In contrast, the emission  $E_3$ , which is only observed in the vicinity of the star, can be attributed to continuous heating of larger grains. Another interesting result is that the maximum of the PAH emission is always linked to a sharp decrease of the  $E_2$  intensity (cf. Fig. 4). The study of the maps of other objects indicates that this correlation is quite general. This is also confirmed by the recent work of Abergel et al. (2001). This strongly suggests that PAHs might be produced by the destruction of the  $E_2$  carriers.

The last spectrum,  $E_4$  traces variations of the intensity of the C-H out-of-plane bending modes. Joblin et al. (1996) have shown in the reflection nebula NGC 1333-SVS3 that the 8.6/11.3  $\mu\text{m}$  intensity ratio decreases with



**Fig. 4.** Amplitudes of the four spectra in three horizontal cuts:  $E_1$  (solid line),  $E_2$  (dash-dot),  $E_3$  (dash), and  $E_4$  (dots). The intensity was scaled by a factor of 2 for  $E_3$  and  $E_4$ .

increasing angular distance from the star. This was interpreted as changes in the ionised fraction of PAHs, the ionisation yield increasing closer to the star. The variation of  $E_4$  (cf. cuts of Fig. 4) clearly shows the same trend. The negative values observed near the star correspond to an increase of the 8.6/11.3  $\mu\text{m}$  intensity ratio.  $E_4$  appears therefore to be a tracer of the ionisation of PAHs. Furthermore, this analysis confirms the prediction made by the authors that the 6.2/11.3  $\mu\text{m}$  and the 7.7/11.3  $\mu\text{m}$  band intensity ratios should behave like the 8.6/11.3  $\mu\text{m}$  ratio. Finally, it can be noticed that the 11.3, 12.7 and 16.4  $\mu\text{m}$  bands appear simultaneously both in  $E_1$  and  $E_4$  with almost the same ratio, which puts new constraints for the identification of the emitters.

Detailed information on the position and width of the bands cannot come out the present study limited to four spectra. However, using a larger basis set, difference spectra describing shift and broadening can be constructed. More detailed results and further discussion will be reported in a forthcoming paper.

## References

- Abergel, A., Bernard, J. P., Boulanger, F., et al. 2001, in prep.
- Boulanger, F., Boissel, P., Cesarsky, D., & Ryter, C. 1998, *A&A*, 339, 194
- Cesarsky, D., Lequeux, J., Abergel, A., et al. 1996, *A&A*, 315, L305
- Désert, F. X., Boulanger, F., & Puget, J. L. 1990, *A&A*, 237, 215
- Folkes, S., Ronen, S., Price, I., et al. 1999, *MNRAS*, 308, 459
- Gerin, M., Phillips, T. G., Keene, J., Betz, A. L., & Boreiko, R. T. 1998, *ApJ*, 500, 329
- Joblin, C., Tielens, A. G. G. M., Geballe, T. R., & Wooden, D. H. 1996, *ApJ*, 460, L119
- Henry, E. R., & Hofrichter, J. 1992, *Methods in Enzymology*, 210, 129