

# Extinction law classification and lens redshift estimate by means of the principal component analysis

C. Jean<sup>1,2</sup> and J. Surdej<sup>1</sup>

<sup>1</sup> Institut d'Astrophysique et de Géophysique, Université de Liège, Allée du 6 août 17, 4000 Liège, Belgium  
e-mail: jean@astro.ulg.ac.be

<sup>2</sup> Departament d'Astronomia i Astrofísica, Universitat de València, Edifici d'Investigació Jeroni Muñoz, C/ Dr. Moliner 50, 46100 Burjassot (València), Spain

Received 9 November 2006 / Accepted 11 June 2007

## ABSTRACT

**Aims.** We propose a method based on the Principal Component Analysis (PCA) to classify and estimate the redshift of an extinction law in a distant gravitational lens galaxy. Such extinction laws are very poorly known and an efficient method to characterize them is badly needed.

**Methods.** We first compute the principal axes of an exhaustive collection of redshifted theoretical extinction laws. Then, we project on these new axes the extinction law we wish to classify. The position of its projection among those redshifted extinction laws from the collection allows us to characterize it and to estimate its redshift.

**Results.** Monte Carlo simulations show that the method is efficient and relatively precise for reasonably good signal-to-noise ratio data. The application of the method to a real case, the gravitational lens system SBS 0909+532, leads to very encouraging results.

**Key words.** dust, extinction – methods: data analysis

## 1. Introduction

Extinction laws of remote galaxies (i.e. with the exception of the Magellanic Clouds) are very poorly known. They have been studied for a small number of cases, e.g., M 31 (Bianchi et al. 1996), M 101 (Rosa & Benvenuti 1994), NGC 2076 (Sahu et al. 1998), NGC 2207 (Berlind et al. 1997), NGC 7070 (Brosch et al. 1985), NGC 7625 (Brosch & Loinger 1991) and some lensing galaxies like B1152+199 (Toft et al. 2000) and SBS 0909+532 (Motta et al. 2002). Gravitational lens systems can provide, under some hypotheses, a way to extract the extinction law of lensing galaxies by dividing the spectra of the lensed images of the background quasar between each other (see, e.g., Nadeau et al. 1991; Jean & Surdej 1998; Motta et al. 2002). If we assume a theoretical model for the extinction laws, i.e. the equations of Cardelli et al. (1989, hereafter CCM) which only depend on the  $R_V$  parameter, we can try to classify the extracted extinction law and estimate its redshift  $z_l$ . The common method used to reach this aim is to fit the parameters  $R_V$  and  $z_l$  by minimizing a  $\chi^2$  (see, e.g., Jean & Surdej 1998). Here, we propose another method, based on the Principal Component Analysis (PCA), used to classify and to estimate the redshift of a remote extinction law. Note that PCA was already used to determine and analyze an extinction law from multi-band photometric data of stars by Massa & Lillie (1978) and by Massa (1980a,b).

PCA is a mathematical technique which allows us to find patterns in data and to express the data in a more compact way that highlights their similarities and differences (see, e.g., Connolly et al. 1995 and Murtagh & Heck 1987, for a mathematical description). This technique is a very efficient way to summarize and classify data which may depend on a large number of parameters. Among many astronomical applications (see, for a list of examples, Cabanac et al. 2002), PCA is often used as a

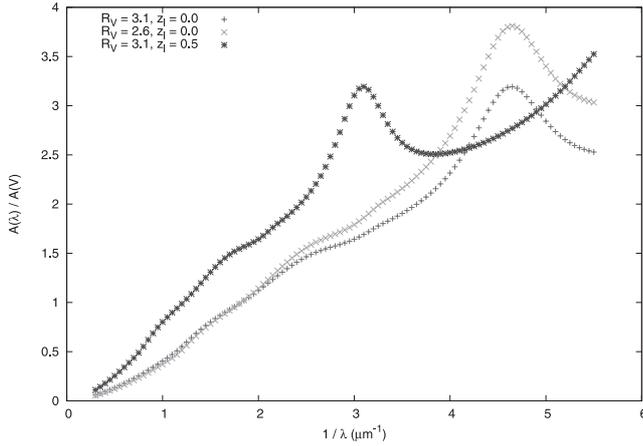
spectral classification tool for stars and galaxies (e.g., Scarlata et al. 2006; Cabanac et al. 2002; Connolly et al. 1995).

We propose to classify gravitational lens extinction laws and estimate their redshift by means of PCA. The first part of this article describes the method and the second part presents Monte-Carlo simulations to assess the reliability of the method and the accuracy of the estimated parameters. In the third part, we apply the method to a real case.

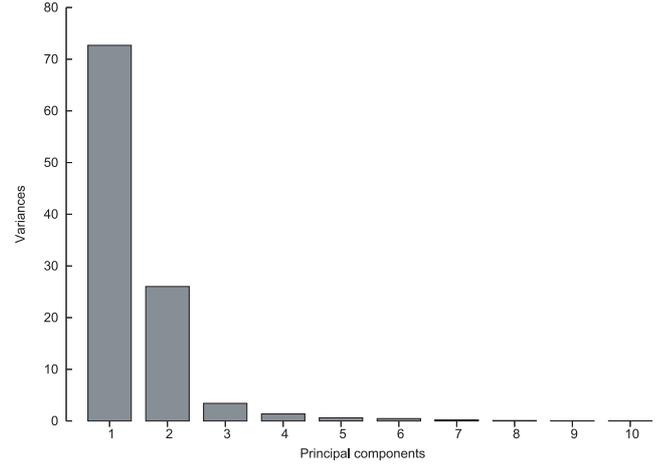
## 2. Description of the method

The principle of the method is the following. We first build a collection of theoretical extinction laws computed with the help of the relations of Cardelli et al. (1989) which only depend on the  $R_V$  parameter,  $R_V = \frac{A_V}{A_B - A_V}$ , the ratio of the total extinction in the  $V$  band to the selective extinction between the  $B$  and  $V$  bands. Thus, by computing extinction laws with different values of the  $R_V$  parameter and for redshifts  $z_l$  within a given range, we fill a collection of theoretical extinction laws, which has to be as exhaustive as possible. Figure 1 illustrates three examples of extinction laws from that collection: two at a redshift  $z_l$  of 0 with  $R_V$  values of 2.6 and 3.1 and one at  $z_l = 0.5$  with  $R_V = 3.1$ .

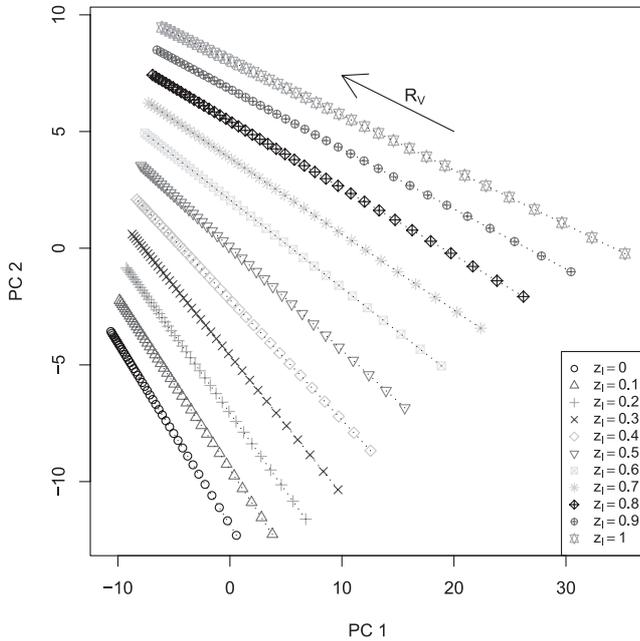
Next, we have to compute the principal axes of the collection. This is done with the *prcomp* command in the R environment (R Development Core Team 2005). Then, by projecting the extinction laws of the collection on the principal axes, we obtain a new expression of the data. For example, let us consider a collection made of 451 extinction laws which correspond to 41 values of the  $R_V$  parameter from 2 to 6 with a step of 0.1 and 11 redshifts from 0 to 1 with a step of 0.1. Each law has a spectral resolution of  $0.05 \mu\text{m}^{-1}$ , that is to say 105 wave numbers between 0.3 and  $5.5 \mu\text{m}^{-1}$ . The data are first normalized,



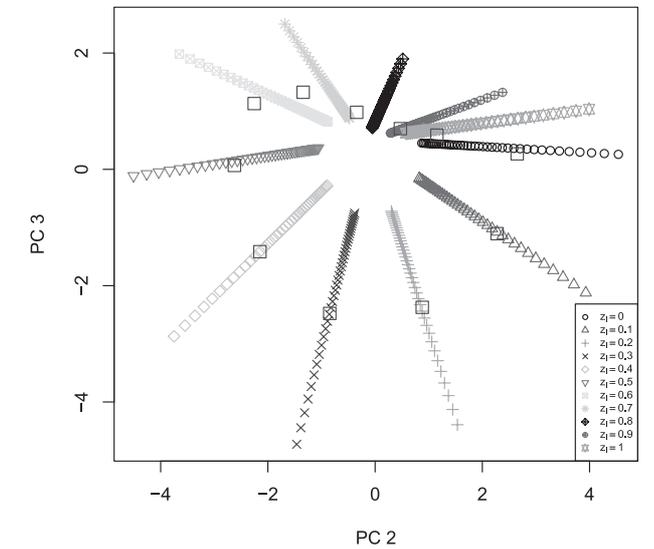
**Fig. 1.** Three extinction curves computed with the Cardelli et al. (1989) relations for different values of the  $R_V$  parameter and the redshift  $z_1$ .



**Fig. 3.** Variances of the first ten principal components.



**Fig. 2.** First two principal component diagram for the 451 extinction laws (see text). Each symbol corresponds to an extinction law. The arrow indicates the direction of increasing values for  $R_V$  ( $2 \leq R_V \leq 6$ ).



**Fig. 4.** Projections of Seaton's extinction law for redshifts between 0 and 1 (large open squares) over the second and third principal components of a collection of CCM extinction laws (small symbols). These moments were chosen for a better visibility.

i.e. their mean is subtracted and they are then divided by their standard deviation. After this operation, the extinction laws have a zero mean and a standard deviation equal to one. The normalization is done in order to avoid a possible preponderance of certain variables. In our case, the second step (the division by the standard deviation) is not strictly necessary because the variables are all of the same type (relative extinction  $A(\lambda)/A(V)$ ) and of the same order but this division improves the results.

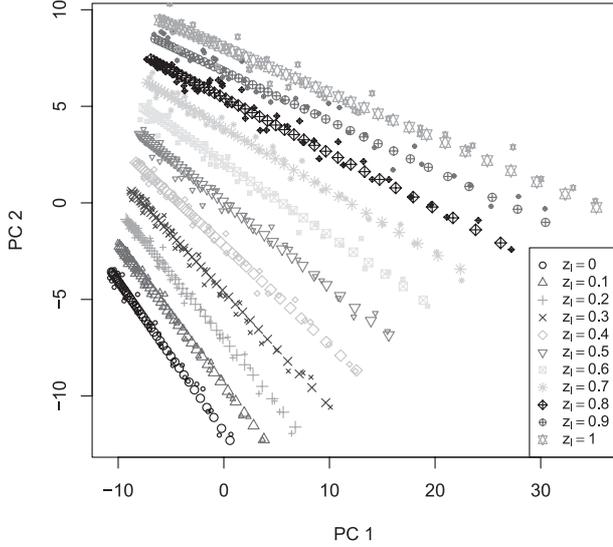
If we compute the principal axes and project the extinction laws on them, we obtain the principal components. Figure 2 shows the first two principal components for the 451 extinction laws.

The principal axes are computed in such a way that the projections of the extinction laws are the most spread, i.e. that their variances are maximum. Figure 3 represents the variances of the first ten principal components and shows that the first two components dominate. Consequently, these two components contain

most of the information necessary to describe the 451 extinction laws.

Finally, if we have an unknown extinction law that we wish to characterize, we project it on the principal axes of the collection and we have to search for its nearest neighbour among the 451 projections.

To assess the aptitude of the method for classifying an external extinction law, let us now test the method using an extinction law which does not come from the CCM relations: the extinction law published by Seaton (1979). The analytic expression of this extinction law is computed for  $R_V = 3.2$ . Figure 4 shows the projections of Seaton's law for different redshifts on the second and the third principal components (chosen for a better visibility) of a collection built with the CCM extinction laws. The result is that the projections of Seaton's law are close to the corresponding ones in the collection of CCM laws, at least for the lowest redshifts ( $0 \leq z_1 \leq 0.5$ ). For redshifts higher than 0.5, the discrepancy is more important because Seaton's law differs from the CCM law with  $R_V = 3.2$  for  $1/\lambda > 7 \mu\text{m}^{-1}$  (see Cardelli et al. 1989).



**Fig. 5.** Projections of two collections of extinction laws with a spectral resolution of  $0.05 \mu\text{m}^{-1}$  and a step in  $z_1$  and  $R_V$  of 0.1. The noisy collection (small symbols) was derived by adding a Gaussian noise to the extinction laws of the original collection (large symbols) from which the principal axes were computed. The signal-to-noise ratio of the noisy collection is 10 per resolution element.

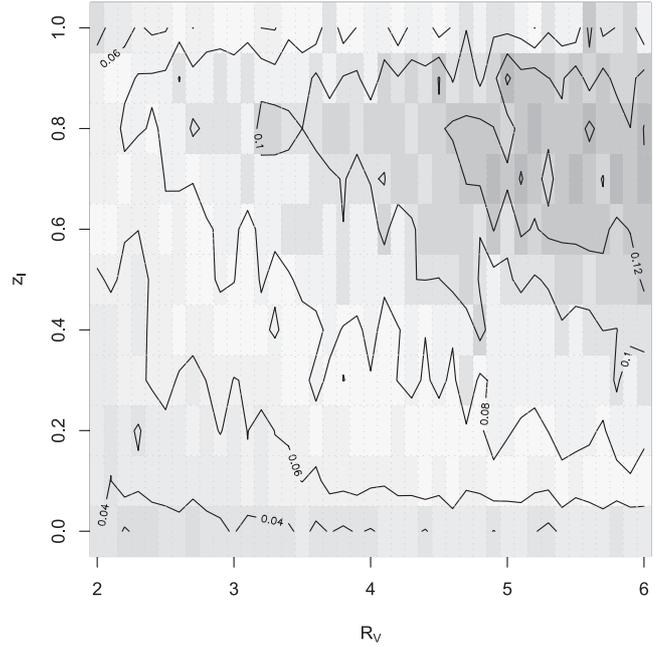
### 3. Monte Carlo simulations

#### 3.1. Presentation

To test the reliability and to estimate the accuracy of the method, we have carried out a large number of Monte Carlo simulations. We built 100 collections with the same values of parameters as the original collection but we added Gaussian noise to the extinction laws in such a way that they are characterized by a given signal-to-noise ratio. We then projected these noisy extinction laws on the principal axes of the original collection. Because of the noise added to the data, their projections do not necessarily match those of the original data but may get closer to projections of laws characterized by other parameter values. Therefore, by doing 100 simulations for each configuration of the parameters (spectral resolution, steps in  $z_1$  and  $R_V$  and signal-to-noise ratio) and by computing the standard deviation of the 100 projections around the projection of the original law, we can estimate the error on the determination of the parameters  $z_1$  and  $R_V$  for each projected law.

#### 3.2. Example

Let us consider two collections with the same spectral resolution ( $0.05 \mu\text{m}^{-1}$ ) and the same steps in  $z_1$  and  $R_V$  (0.1). We now add a Gaussian noise to the second collection in such a way that the final signal-to-noise ratio is 10 per resolution element. Figure 5 represents the projections of the two collections on the principal axes of the original collection (i.e. without noise). We can see in Fig. 5 that the projection of a noisy extinction law can be rather distant from the projection of the original law and, thus, its nearest neighbour may be the projection of an extinction law with different parameter values.



**Fig. 6.** Contour map of the uncertainty  $\sigma_{z_1}$  as a function of  $R_V$  and  $z_1$  for a signal-to-noise ratio of 3 per resolution element and a spectral resolution of  $0.1 \mu\text{m}^{-1}$ . The darker it is, the smaller the uncertainty is.

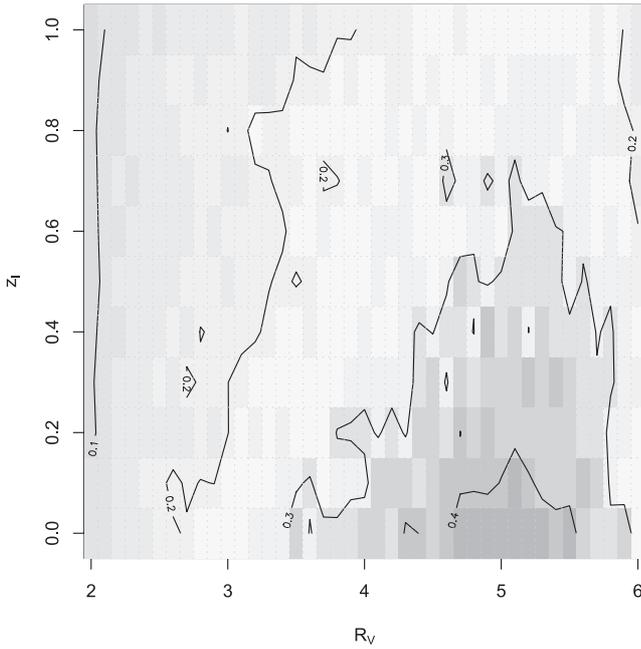
To find the nearest neighbour of a projection, we have first to compute the weighed distances  $d$  between all the projections (see Cabanac et al. 2002):

$$d = \sqrt{\sum_{i=1}^m \lambda_i (n_i - o_i)^2}$$

where  $m$  is the number of variables, i.e. the number of spectral resolution elements;  $\lambda_i$  is the eigenvalue associated with the eigenvector  $v_i$ ;  $n_i$  and  $o_i$  are the projections of the noisy extinction law and of the original extinction law on the eigenvector  $v_i$ , respectively. For a given projection of a noisy extinction law, the projection whose distance is minimal is its nearest neighbour, from which we extract the parameters  $z_1$  and  $R_V$  of the corresponding extinction law. By repeating this process 100 times, we obtain two sets of 100 values of  $z_1$  and  $R_V$  and, by computing their standard deviations, we can estimate the accuracy of the method.

#### 3.3. Results

The simulations show that the uncertainties on the estimated parameters ( $z_1$  and  $R_V$ ) depend on the parameters themselves, on the signal-to-noise ratio, on the spectral resolution and on the sampling rate in  $R_V$  and  $z_1$  of the collection. For example, let us consider a set of 100 collections of extinction laws with a spectral resolution of  $0.1 \mu\text{m}^{-1}$ , a step in  $R_V$  and  $z_1$  of 0.1 and a signal-to-noise ratio of 3. Figures 6 and 7 present, as contour maps, the uncertainties  $\sigma_{z_1}$  and  $\sigma_{R_V}$  affecting the estimates of the  $z_1$  and  $R_V$  parameters. Figure 6 shows that the error  $\sigma_{z_1}$  on the estimate of the  $z_1$  parameter varies from about 0.02 for the lowest values of  $z_1$  and  $R_V$  to 0.15 for the highest values of  $z_1$  and  $R_V$ . On the other side, as shown by Fig. 7, the uncertainty  $\sigma_{R_V}$  on the estimate of the  $R_V$  parameter varies from a bit less than 0.06 to about 0.46. For both  $z_1$  and  $R_V$  estimates, the relative uncertainties do not exceed 10% or so in this example. Furthermore, we



**Fig. 7.** Contour map of the uncertainty  $\sigma_{R_V}$  as a function of  $R_V$  and  $z_1$  for a signal-to-noise ratio of 3 per resolution element and a spectral resolution of  $0.1 \mu\text{m}^{-1}$ . The darker it is, the smaller the uncertainty is.

have to make sure that the collection is sufficiently exhaustive (i.e. with a high sampling rate in  $z_1$  and  $R_V$ ) in order to allow the projections to match with good nearby neighbours. Otherwise, the standard deviations will not be representative of the true error on the estimates. Of course, a much larger number of simulations have been carried out with different spectral resolutions, different sampling rates in  $z_1$  and  $R_V$  and various signal-to-noise ratios. Those simulations lead to consistent and encouraging results.

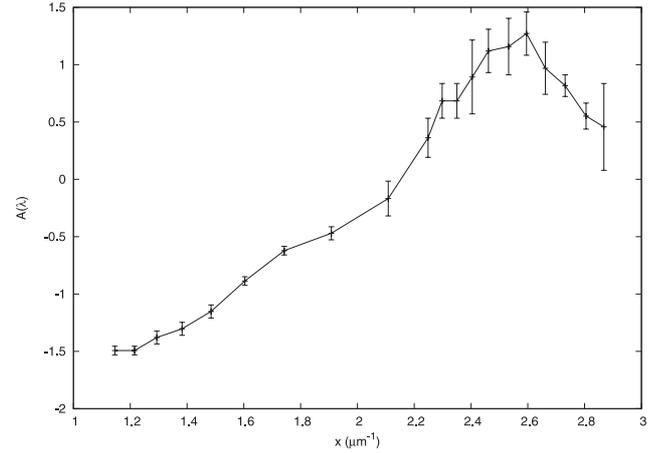
## 4. Application to a real case: SBS 0909+532

### 4.1. Introduction

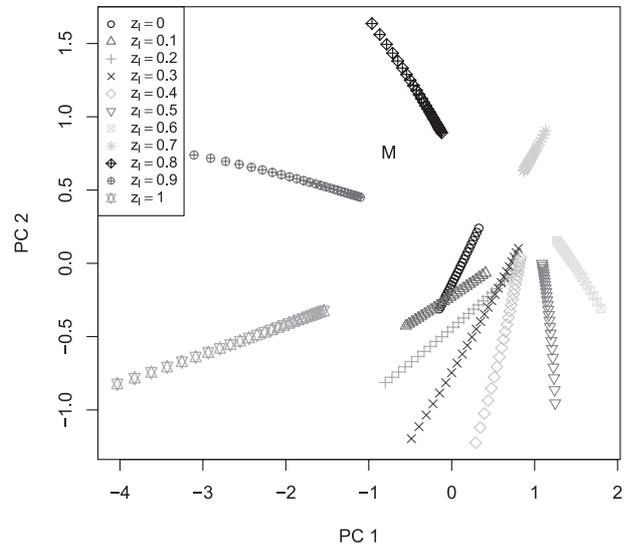
Motta et al. (2002) extracted the extinction law of the lensing galaxy of the gravitational lens system SBS 0909+532 from integral field spectroscopic observations and fitted the theoretical expressions from Cardelli et al. (1989) to it. After fixing the lens redshift to its known value ( $z_1 = 0.83$ ), they found by  $\chi^2$  minimization a value of 2.1 for the  $R_V$  parameter. We are now going to apply our PCA-based method to this extinction law in order to estimate the values of the  $z_1$  and  $R_V$  parameters.

### 4.2. Application of the method

The extinction law published by Motta et al. (2002) consists of the calculation of 20 differential extinctions between the two components of the gravitational lens system. On the other side, the relations of Cardelli et al. (1989) give a normalized expression of the extinction:  $A(\lambda)/A(V)$ . Since we are not supposed to know the lens redshift of SBS 0909+532, we cannot estimate the value of  $A(V)$  in the lens frame. Therefore, we have to normalize in the same way all the extinction laws, that of Motta et al. (2002) as well as those of Cardelli et al. (1989). We choose a classical normalization: their mean is first subtracted and the laws are then divided by their standard deviation. Figure 8 shows the extinction curve published by Motta et al. (2002), after normalization.



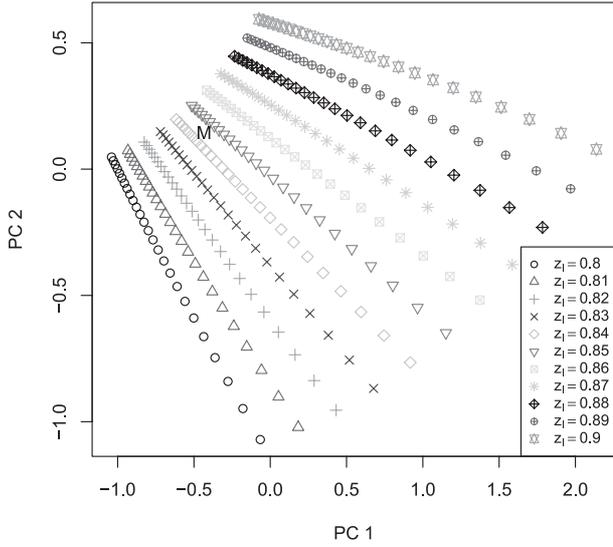
**Fig. 8.** Normalized extinction law of the lensing galaxy of SBS 0909+532, published by Motta et al. (2002)



**Fig. 9.** Projections of the extinction laws from the collection and of the one extracted by Motta et al. (2002), marked “M”, over the two first principal axes of the collection.

To classify the observed extinction law as a function of its position in the space defined by the principal axes of the original collection, the latter must be made of extinction laws with the same spectral sampling as the observed one. We therefore build a collection of normalized extinction laws with, as spectral sampling, the 20 wavelengths of the observed curve. Since we do not know a priori the values of  $z_1$  and  $R_V$ , we first choose large ranges for these parameters:  $z_1$  between 0 and 1 with a step of 0.1, and  $R_V$  between 1 and 6 with a step of 0.2. Then, we compute the principal axes of this collection and project on them the observed extinction law. The result for the first two principal components is depicted in Fig. 9. The apparent degeneracy of projections for  $z \leq 0.4$  results from the fact that the extinction laws are projected on a plane. In the full space, the points are indeed distinct.

We can already notice that the projection, marked “M”, of the law from Motta et al. (2002) is between the points representing the laws at  $z_1 = 0.8$  and  $z_1 = 0.9$ . As for the  $R_V$  parameter, we cannot estimate it yet. To yield a more accurate estimate of the parameters  $z_1$  and  $R_V$ , we repeat the same process with a smaller range of redshifts and a higher sampling rate. We therefore build a new collection of normalized extinction laws with



**Fig. 10.** “Zoom” of Fig. 9 for redshifts  $z_1$  between 0.8 and 0.9 and for  $R_V$  values between 1 and 6. The letter “M” still corresponds to the observed law.

a redshift between 0.8 and 0.9 and a step reduced to 0.01. We keep the same range and the same step for the  $R_V$  parameter ( $R_V$  between 1 and 6 with a step of 0.2). Again, we compute the principal components and we project the extinction laws on the principal axes. The results for the two principal components are presented in Fig. 10. The nearest projection of an extinction law from the collection is the one which corresponds to a redshift of 0.84 and a  $R_V$  value of 2.4. These results are to be compared with those of Motta et al. (2002):  $z_1 = 0.83$  (previously known) and  $R_V = 2.1 \pm 0.9$ .

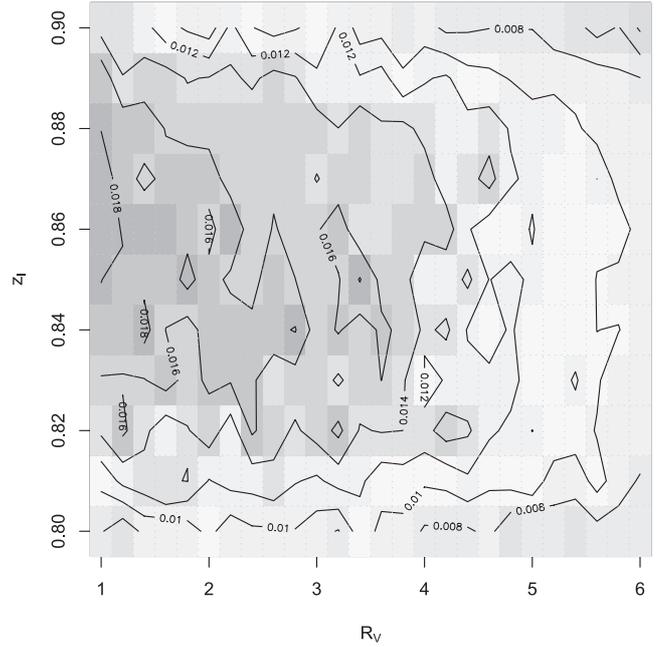
#### 4.3. Error estimate

Our results seem to be consistent with the published values but we still have to estimate their precision. For that purpose, we here also use Monte Carlo simulations. We simulate 100 collections with the same configuration as in the original collection but we add a Gaussian noise whose dispersion is, for each wavelength, the corresponding error stated in the extinction law of Motta et al. (2002). These simulated laws are then projected on the principal axes of the original collection and we calculate the dispersions  $\sigma_{z_1}$  and  $\sigma_{R_V}$  of the parameters  $z_1$  and  $R_V$  of the nearest projections. Figures 11 and 12 show these dispersions  $\sigma_{z_1}$  and  $\sigma_{R_V}$ .

Finally, for a redshift of  $z_1 = 0.84$  and a value of the parameter  $R_V = 2.4$ , the corresponding uncertainties deduced from the dispersion of the projections are  $\sigma_{z_1} \simeq 0.02$  and  $\sigma_{R_V} \simeq 1.2$ . These results are in very good agreement with those published by Motta et al. (2002). Of course, we can go further and retrieve better estimates of the parameters by repeating the process with a finer collection, i.e. smaller sampling steps in  $z_1$  and  $R_V$ , around the  $z_1$  and  $R_V$  values found so far.

#### 4.4. Conclusion

The result of the application of our method to this real case is very encouraging. Without knowing the redshift  $z_1$  and the value of the  $R_V$  parameter, we are able to estimate them with a good precision for  $z_1$  ( $\sigma_{z_1} \simeq 0.02$ ) and an uncertainty on  $R_V$  ( $\sigma_{R_V} \simeq 1.2$ ) comparable to those Motta et al. (2002) obtained by



## 5. Discussion and conclusion

We have presented a PCA-based method to classify remote extinction laws in lensing galaxies, in the same way as Cabanac et al. (2002) did with star and galaxy spectra.

By first computing the principal axes of an exhaustive collection of well-defined extinction laws and then by projecting on these axes another extinction law taken outside of the collection, we are able to classify it and estimate its redshift  $z_1$ .

Monte Carlo simulations have shown that, as long as the signal-to-noise ratio and the spectral resolution are reasonably good, the parameters  $R_V$  and  $z_1$  can be retrieved with a good precision. Numerous other simulations with different configurations of the spectral resolution and of the completeness of the collection confirm this encouraging result.

Application of this method to the extinction curve of the lensing galaxy of SBS 0909+532, published by Motta et al. (2002), yields a good estimate for the true redshift  $z_1$  and the  $R_V$  parameter. We could also have applied our method to the extinction curves of Falco et al. (1999) and, very recently, of Elíasdóttir et al. (2006) but these extinction curves do not have as good a spectral sampling as the one published by Motta et al. (2002), obtained with spectroscopic observations.

One advantage of this PCA-based method over the traditional  $\chi^2$  minimization is that, in case there is no clear estimate of the parameters, the position of the projection of the target law among the other ones can give some information about a “family resemblance” with other extinction laws, though limited to the ones included in the collection. This strengthens the necessity to have a very exhaustive collection of extinction laws. On the other hand, a high  $\chi^2$  does not yield such an information.

Finally, if we could retrieve several remote extinction laws with a good signal-to-noise ratio and a good spectral sampling, we could first estimate their redshift and the value of the  $R_V$  parameter and include them in the collection. This improved collection will then be used to classify further extinction laws and,

therefore, we could gradually explore the properties of the interstellar dust in distant galaxies.

*Acknowledgements.* The authors acknowledge support from the ESA PRODEX Programmes “HST Gravitational Lensing” and “XMM-LSS”, and from the Belgian Federal Science Policy Office for their support. This work is also supported by the European Community’s Sixth Framework Marie Curie Research Training Network Programme, Contract No. MRTN-CT-2003-505183 “ANGLES”. The authors also thank the referee, Rémi Cabanac, for his useful suggestions and comments.

## References

- Berlind, A. A., Quillen, A. C., Pogge, R. W., & Sellgren, K. 1997, *AJ*, 114, 107  
 Bianchi, L., Clayton, G. C., Bohlin, R. C., Hutchings, J. B., & Massey, P. 1996, *ApJ*, 471, 203  
 Brosch, N., & Loinger, F. 1991, *A&A*, 249, 327  
 Brosch, N., Greenberg, J. M., & Grosbol, P. J. 1985, *A&A*, 143, 399  
 Cabanac, R. A., de Lapparent, V., & Hickson, P. 2002, *A&A*, 389, 1090  
 Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245  
 Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., & Calzetti, D. 1995, *AJ*, 110, 1071  
 Elíasdóttir, Á., Hjorth, J., Toft, S., Burud, I., & Paraficz, D. 2006, *ApJS*, 166, 443  
 Falco, E. E., Impey, C. D., Kochanek, C. S., et al. 1999, *ApJ*, 523, 617  
 Jean, C., & Surdej, J. 1998, *A&A*, 339, 729  
 Massa, D. 1980a, *AJ*, 85, 1644  
 Massa, D. 1980b, *AJ*, 85, 1651  
 Massa, D., & Lillie, C. F. 1978, *ApJ*, 221, 833  
 Mediavilla, E., Muñoz, J. A., Kochanek, C. S., et al. 2005, *ApJ*, 619, 749  
 Motta, V., Mediavilla, E., Muñoz, J. A., et al. 2002, *ApJ*, 574, 719  
 Murtagh, F., & Heck, A. 1987, *Multivariate data analysis (Astrophysics and Space Science Library, Dordrecht: Reidel)*  
 Nadeau, D., Yee, H. K. C., Forrest, W. J., et al. 1991, *ApJ*, 376, 430  
 R Development Core Team 2005, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria  
 Rosa, M. R., & Benvenuti, P. 1994, *A&A*, 291, 1  
 Sahu, D. K., Pandey, S. K., & Kembhavi, A. 1998, *A&A*, 333, 803  
 Scarlata, C., Carollo, C. M., Lilly, S. J., et al. 2006, *ApJ* [arXiv:astro-ph/0611644]  
 Seaton, M. J. 1979, *MNRAS*, 187, 73P  
 Toft, S., Hjorth, J., & Burud, I. 2000, *A&A*, 357, 115