

Eigenvector Sky Subtraction

Michael J. Kurtz and Douglas J. Mink

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138

Accepted for publication in The Astrophysical Journal (Letters) 2000 March 7.

ABSTRACT

We develop a new method for estimating and removing the spectrum of the sky from deep spectroscopic observations; our method does not rely on simultaneous measurement of the sky spectrum with the object spectrum. The technique is based on the iterative subtraction of continuum estimates and Eigenvector sky models derived from Singular Value Decompositions (SVD) of sky spectra, and sky spectra residuals. Using simulated data derived from small telescope observations we demonstrate that the method is effective for faint objects on large telescopes. We discuss simple methods to combine our new technique with the simultaneous measurement of sky to obtain sky subtraction very near the Poisson limit.

Subject headings: methods: data analysis — techniques: spectroscopic — galaxies: redshifts

1. Introduction

Light from the night sky interferes with the analysis of the spectra of all but the brightest astronomical objects. For the determination of galaxy redshifts this limit was reached about forty years ago (Humason et al. 1956). With the advent of digital spectrographs (e.g. Westphal et al. 1975, Shectman and Hiltner 1976) simultaneous measurement of object and night sky spectra became routine, thus allowing sky subtraction. The basic principle of sky subtraction has been that the sky spectrum should be observed in as identical a manner as possible (in time, location on the sky, and optical path) to the sky spectrum observed in the direction of the object. As ever fainter objects have been observed the accuracy of the sky subtraction has become more critical, and increasingly sophisticated methods, mechanical, observational, and algorithmic, have been employed to match the observed sky to the sky in front of the object. For wide field, multi-object, fiber optic spectrographs this procedure has been especially problematic. Watson et al. 1998 review the problem of sky subtraction with fiber optic spectrographs in detail.

In this paper we present a new method for removing the influence of the night sky emission from the spectra of faint astronomical objects; this new method does not rely on the simultaneous measurement of the sky with the object.

2. Iteratively Estimating the Sky

Removing the sky from an object + sky observation is equivalent to estimating the sky spectrum in the direction of the object at the time of observation. Because the sky spectrum is composed of several components the first step in the estimation process must be to decide which of these components must be accurately estimated, and which can be.

Dark sky spectra may be characterized by a smooth continuum, plus weak absorption lines (from the zodiacal light) and a host of emission

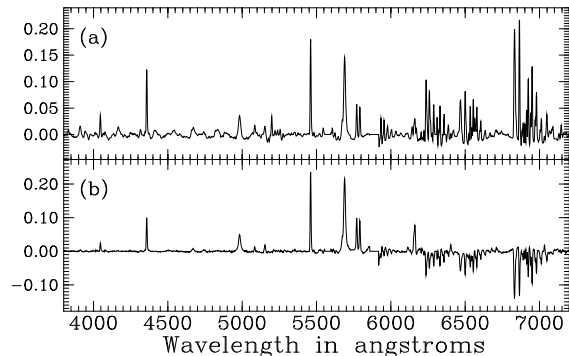


Fig. 1.— The first two Eigenvectors in the continuum subtracted sky model.

lines. For many projects involving multi-fiber spectroscopy of faint galaxies (e.g. redshift surveys) the continuum of the object is removed before the analysis (Kurtz and Mink 1998, hereafter KM98); in these instances it is not necessary to estimate the continuum of the sky separately; it suffices to estimate the continuum of the object + sky. Additionally, some lines in the sky spectrum are so bright that errors in their estimation, which are large compared with typical features in the object spectra, are unavoidable.

Our new method for sky subtraction begins by estimating and subtracting the sky + object continuum; this step removes typically more than 80% of the sky flux, as well as the object's continuum. All through the sky subtraction process, we ignore pixels which correspond to the bright sky lines we choose not to estimate.

Next we fit the residual sky + object spectrum with a two component Eigenvector model (section 3) of the continuum subtracted sky, and subtract this fit. Figure 1 shows the two component sky model used in section 4; to first approximation the first vector represents the mean sky, and the second adjusts for the fluctuations in the OH lines. Typically we remove more than 80% of the flux in the residual sky spectrum with this step.

The second order residual which results from

these steps typically shows large (wavelength) scale variations due to errors in the continuum estimation. With most of the sky lines removed a second order continuum fit may be made, resulting in a reflattened (Kurtz and Lasala 1991) residual sky spectrum.

Finally we fit a several component (in section 4 we use ten) Eigenvector model of the reflattened residual sky spectrum, and subtract it.

The residual from this last operation is the continuum subtracted object spectrum, along with any remaining noise and errors.

3. The Eigenvector Model

Eigenvector decomposition techniques (e.g. Courant and Hilbert 1924, Press, et al. 1986) have long been used to classify astronomical objects. Deeming 1964 used Principal Component Analysis (PCA) to classify stars on the basis of multicolor photometry; Brosche 1973 used PCA to classify galaxies on the basis of their integrated properties; and Kurtz 1982 used PCA to classify stars on the basis of their spectra. In these cases classification is exactly equivalent to fitting an Eigenvector model to the data, with the addition that the resulting coefficient space is partitioned into classes.

More recently several authors have used PCA and Singular Value Decomposition (SVD, a similar Eigenvector technique, see the discussion in Press, et al. 1986) to classify stellar (Bailer-Jones et al. 1998 and references therein) and galaxy (Bromley et al. 1998 and references therein) spectra.

In this work we use SVD to create two Eigenvector models, first for the continuum subtracted sky spectra, and second for the residual reflattened sky spectra. To accomplish this task we begin with a set of well-observed sky spectra (we typically use 500 – 1000), which does not contain spectra with cosmic ray hits, bad focus, or other anomalies. We process these spectra so that each contains exactly the same number of counts on

exactly the same wavelength scale. Additionally in all of these steps we ignore the pixels corresponding to the sky lines we choose not to estimate, (section 4) OII 5007Å, NaD, OII 6300Å, and OII 6363Å.

Next we subtract the continuum from each normalized sky spectrum using a high order (~ 40) spline fit. We decompose the continuum subtracted, normalized sky spectra into Eigenvectors using SVD; the first two Eigenvectors are our model for the continuum subtracted sky. Note that higher order Eigenvectors are contaminated by errors in the continuum fit.

Finally we fit the two component Eigenvector model for the continuum subtracted sky to our set of normalized, continuum subtracted sky spectra, and subtract the fit; then we reflatten the residual spectra using another high order spline fit. We then decompose the reflattened residual spectra using SVD; the resulting Eigenvectors are our final model for the reflattened residual sky spectra.

Because of the reflattening, the principal Eigenvectors in this final model do not show contamination by errors in the continuum estimation; thus there is not an obvious cut-off point for the number of Eigenvectors to use in the fits. This cut-off will vary from project to project and from observing protocol to observing protocol. In section 4 we use ten Eigenvectors; there is a clear change in slope in the log-singular-value vs. Eigenvector-number diagram at the tenth Eigenvector, indicating the transition from signal to noise.

4. Testing the Method

The purpose of our new sky subtraction method is to allow observations of fainter objects to be made with multi-object fiber-optic spectrographs, such as the 300 fiber Hectospec (Fabricant et al. 1994, Fabricant et al. 1998a) on the converted 6.5m MMT. A typical operating mode for this instrument is that each half hour

there would be three 10 minute exposures using the same fiber set-up; the sets of three 10 minute spectra would be combined to remove cosmic rays and other shot noise. In this operating mode, with 10% of the fibers devoted to sky observations, several hundred defect-free sky spectra would be observed per night.

This instrument is not yet in operation; here we therefore build sets of spectra for testing from the archive of spectra from the FAST spectrograph (Fabricant et al. 1998b). FAST is a long slit spectrograph residing on the 1.5m Tillinghast telescope. We choose two datasets, the first of well observed dark skies, both to create the Eigenvector models and to add to the object spectra, and the second a set of typical redshift survey spectra.

For the sky spectra we took all sky exposures (exactly 1000) made between 21 Jan 1999 and 27 Dec 1999 where the exposure time was greater than 500 seconds, the sun was at least 30° below the horizon, the moon was at least 20° below the horizon, and the instrumental set-up was close to that standardly used in redshift surveys. We removed spectra which showed signs of anomalies, including cosmic rays or residual noise from UV flooding, for a final list of 924 sky spectra. The mean number of counts per 1.7\AA pixel in these spectra is 569, with a factor of three variation about this value.

We obtained these sky spectra from regions along the slit adjacent to the objects being observed; they typically measured an area on the sky of $3'' \times 30''$, which, for constant observing time, produce a flux about three times what we expect from a $1.5''$ diameter fiber on the Hectospec/MMT, after taking into account the 18.2 times larger collecting area of the MMT.

The galaxy spectra are the set of all spectra from the 15R survey (Geller, et al. 2000) where the exposure time was greater than 500 seconds, the sun was at least 30° below the horizon, the moon was at least 20° below the horizon, and the redshifts were judged secure by S. Tokarz follow-

ing manual inspection. This dataset consists of 1648 spectra, of ~ 15 th mag (R) galaxies, taken between 1994 and 1997. We then put these spectra through a fully automatic redshift reduction (KM98), and remove the 15 spectra where the fully automated result did not match the result confirmed by manual inspection; typically this difference results from unremoved cosmic ray hits in the spectra. We further remove 110 spectra where the r statistic (a measure of quality of fit, Tonry and Davis 1979) is < 5 , the redshifts from absorption line spectra with $r > 5$ are nearly always correct (KM98); we routinely accept them without manual inspection. Our final test set is 1523 typical redshift survey spectra; each original spectrum is reducible fully automatically. The mean number of counts per 1.7\AA pixel in these spectra is 277, the fluctuations about this value are about a factor of 1.5.

We take each of the 1523 galaxy spectra, and combine it with ten different sky spectra, which are created by summing between 1 and 10 randomly selected sky spectra. This procedure yields a sample of 15230 test spectra with sky to object flux ratios which vary by more than two orders of magnitude.

Each of these combined spectra then has the sky removed by the procedure described in section 2, and has its redshift determined by the automated methods of KM98. We compare these redshifts with those derived from the original reduction of the galaxy spectra; a redshift is considered correct if it is within 300km/s of the original value.

Figure 2 shows the fraction of correct redshifts obtained in this test as a function of the flux ratio between the object and the sky, expressed in magnitudes. This flux ratio refers to light entering the fiber, and is thus primarily a measure of the seeing convolved central surface brightness of the object with respect to the brightness of the dark sky.

We divide the galaxy spectra into two subsamples: emission line dominated spectra, and

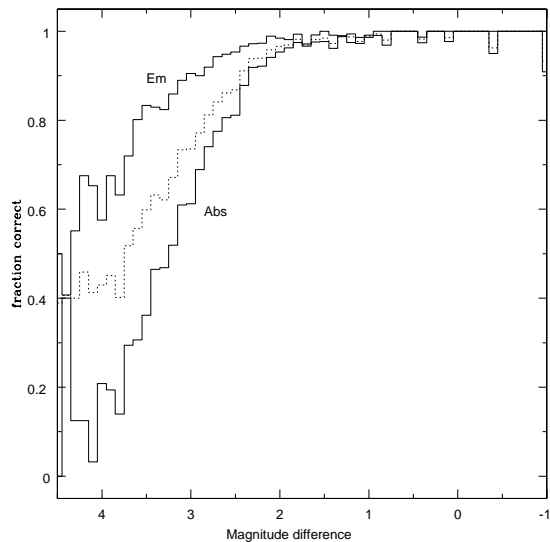


Fig. 2.— The fraction of correct redshifts as a function of the flux ratio of the object with the sky, expressed in magnitudes.

absorption line dominated spectra. The solid lines marked em/abs represent spectra where the emission/absorption line template (KM98) obtained the higher r value in the original reduction. Emission line objects still give redshifts for objects about a magnitude fainter than absorption line objects, everything else being held constant. The dotted line between the em/abs lines is the total, and represents a typical mix of spectra for a magnitude limited redshift survey at redshift 0.05 transformed to fainter apparent magnitude.

The absorption line redshifts are more than 95% correct until the objects are 2 magnitudes fainter than the sky and more than 50% correct until the objects are 3.25 magnitudes fainter than the sky. The emission line spectra are more than 95% correct until the objects are 2.5 magnitudes fainter than the sky, and 50% correct until they are 4.4 magnitudes fainter than the dark sky.

To simulate the Poisson limit we add noise to the original spectra corresponding to the Poisson noise expected from the number of counts

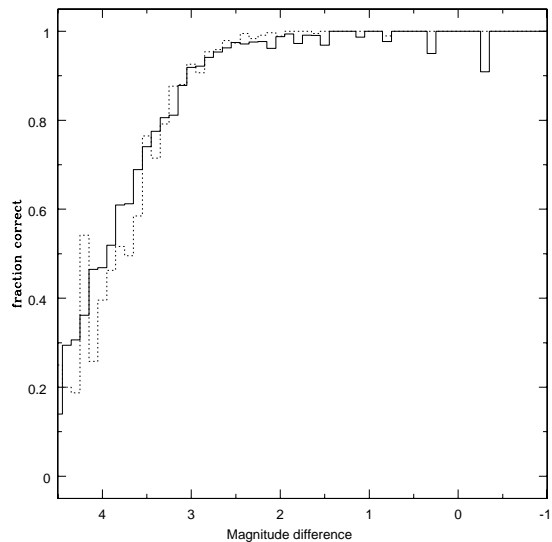


Fig. 3.— The fraction of correct redshifts obtained for Poisson noise limited spectra, compared with the absorption line objects in figure 2 shifted 0.7 magnitudes fainter

in the skies added to the spectra, using \sqrt{n} for each pixel and a Gaussian random deviate generator. This procedure simulates perfect sky subtraction. We then put these spectra through the automated processing of KM98 and obtain redshifts.

Figure 3 shows the fraction with correct redshifts as a function of brightness ratio with the sky. The dotted line shows the fraction correct for the absorption line objects at the simulated Poisson limit (i.e. with perfect sky subtraction); the solid line shows the absorption line result from figure 2 shifted 0.7 mag fainter; the sky subtraction in this test came within 0.7 mag of matching the limiting magnitude imposed by Poisson statistics. The plots for the emission line objects and the sum of all objects yield very similar results.

Figure 4 shows a typical absorption line spectrum at several stages in the test. On top (4a) is the original galaxy spectrum from the 15R survey, its correlation with our absorption line tem-

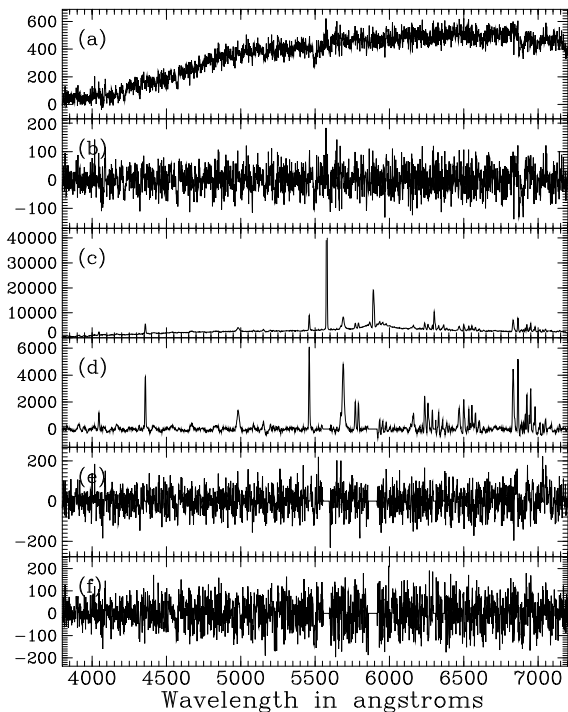


Fig. 4.— A typical absorption line spectrum at various stages of the test; see text.

plate gave an r value of 7.6; below it (4b) is that spectrum with the continuum subtracted. Below that (4c) is the sum of that spectrum and a sky spectrum which is the sum of seven original sky spectra, the object is 2.2 magnitudes fainter than the sky. The next figure (4d) is the sky + object with the bright sky lines and the continuum removed. Second from the bottom (4e) is the residual spectrum after the sky subtraction process, this spectrum yields the correct redshift with an r value of 4.2; the bottom plot (4f) shows the original spectrum in 4a with added random Poisson noise corresponding to the number of sky photons in 4c, and after the continuum has been removed. This spectrum yields the correct redshift with an r value of 6.1.

THIS PARAGRAPH WILL NOT BE PUBLISHED. Figure 5 is similar to figure 4, but shows a typical emission line spectrum. The initial correlation with the emission line template

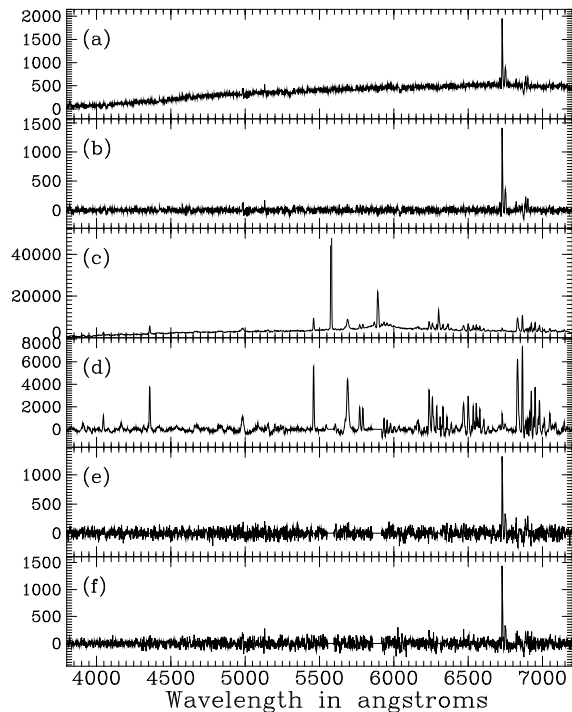


Fig. 5.— A typical emission line spectrum at various stages of the test; see text. THIS FIGURE WILL NOT BE PUBLISHED.

gave an r value of 24.6; the object is 2.5 magnitudes fainter than the sky, and the residual spectrum (5e) gave the correct redshift with an r value of 18.3, while the Poisson limited noise spectrum (5f) gave the correct redshift with an r value of 17.9.

5. Discussion

Our new method for sky subtraction can remove the spectrum of the sky from faint redshift survey spectra with an efficiency within a factor of two of the Poisson limit, and *without any simultaneous measurement of the sky spectra*.

The sky spectra in the test were observed over one year, and contain larger variations than would be found in a few-night Hectospec run. For example the principal Eigenvector in the re-flattened residual model has NI 5199Å as its

strongest feature; this line was substantially stronger during two months of the early summer than at any other time. Another Eigenvector, which shows a solar (zodiacal light) spectrum, has a clear yearly variation. Variations caused by changes in the instrument set-up, focus, or chip response (we periodically UV flood) will also affect the result. Even with these added sources of variance we obtain very good results for absorption line spectra two magnitudes fainter than the sky, which if the Mt. Hopkins sky is $V=21.46$ mag/arcsec (Massey and Foltz 2000), and assuming a $1.5''$ diameter fiber, corresponds to an object with a $1.5''$ aperture magnitude in V of 22.8; to measure this would take about two hours of integration on the converted MMT.

Clearly the new techniques we have described can be combined with simultaneous sky measurements to achieve improved results. One approach might be to fit a continuum subtracted mean spectrum derived from the sky spectra in a single pointing (either derived from the first Eigenvector, or some other method of obtaining a robust estimate of the mean sky), then a fit to a set of reflattened residual Eigenvectors, where the reflattened residual model is derived from a much larger set of reflattened residual sky spectra, perhaps those observed during an observing run. This procedure would, in the first pass, save for the continuum subtraction and the details of the fitting, mimic current methods for simultaneous sky estimation. The second-pass fit with the reflattened residual model would allow subtraction of correlated small amplitude sky flickering in the sky lines, and systematic effects such as due to bending of the fibers.

With this combination of methods we expect that sky subtraction, save for the unremovable Poisson noise component, will not be a dominant limitation in deep observations with multi-object fiber optic spectrographs.

We thank D. Fabricant, M. Geller, S. Tokarz, E. Falco, J. Roll, and W. Wyatt for discussions.

Software implementing these techniques has

been written in the IRAF (Tody 1986) environment, and is available from the authors.

REFERENCES

- Bailer-Jones, C. A. L., M. Irwin and T. von Hippel 1998. *MNRAS*, 298, 361
- Bromley, B. C., W. H. Press, H. Lin and R. P. Kirshner 1998. *ApJ*, 505, 25
- Brosche, P. 1973. *A&A*, 23, 259
- Courant, R. and Hilbert, D. 1924, *Methoden der Mathematischen Physik I*, Berlin:Springer Verlag
- Deeming, T. J. 1964. *MNRAS*, 127, 493
- Fabricant, D. G., E. H. Hertz and A. H. Szentgyorgyi 1994. *Proc. SPIE*, 2198, 251
- Fabricant, D. G., E. N. Hertz, A. H. Szentgyorgyi, R. G. Fata, J. B. Roll and J. M. Zajac 1998. *Proc. SPIE*, 3355, 285
- Fabricant, D. , P. Cheimets, N. Caldwell and J. Geary 1998. *PASP*, 110, 79
- Geller, M.J., et al 2000 in preparation
- Humason, M. L., N. U. Mayall and A. R. Sandage 1956. *AJ*, 61, 97
- Kurtz, M. J. 1982. Automatic spectral classification. Ph.D. Thesis, Dartmouth Coll., Hanover, NH.
- Kurtz, M. J. and D. J. Mink 1998. *PASP*, 110, 934
- Kurtz, M. J. and J. Lasala 1991. in *Objective Prism and Other Surveys*, ed. A.G.D. Phillip and A.R. Upgren, L. Davis Press, 133
- Massey, P. and C.B. Foltz 2000, *PASP*, April 2000, to appear.
- Press, W. H., Flannery, B. P. Teukolsky, S. A. & Vetterling, W.T. 1986, *Numerical Recipes*, Cambridge: University Press

Shectman, S. A. and W. A. Hiltner 1976. PASP, 88, 960

Tody, D. 1986. Proc. SPIE, 627, 733

Tonry, J. and M. Davis 1979. AJ, 84, 1511

Watson, F., A. R. Offer, I. J. Lewis, J. A. Bailey and K. Glazebrook 1998. ASP Conf. Ser. 37: Fiber Optics in Astronomy II, 50

Westphal, J. A., J. Kristian and A. Sandage 1975. ApJ, 197, L95-L98