ISSN 1330–0016
CODEN FIZBE7

SEPARATION TECHNIQUES FOR DISENTANGLING OF COMPOSITE SPECTRA

SAŠA ILIJIĆ[a], HERMAN HENSBERGE[b] and KREŠIMIR PAVLOVSKI[c,1]

[a]*Faculty of Geodesy, University of Zagreb, Kačićeva 26, HR-10000 Zagreb, Croatia*

[b]*Royal Observatory of Belgium, Ringlaan 3, B-1180 Brussel, Belgium*

[c]*Department of Physics, Faculty of Sciences, University of Zagreb, Bijenička cesta 32, HR-10000 Zagreb, Croatia*
E-mail address: kresimir@sirius.phy.hr

Dedicated to Professor Kseno Ilakovac on the occasion of his 70$^{\mathbf{th}}$ birthday

Received 20 June 2001; Accepted 21 January 2002
Online 6 April 2002

Disentangling of composite spectra is a promising technique for the analysis of double-lined spectroscopic binary systems. The technique makes use of a separation routine to extract model-component spectra out of a time series of observed composite spectra. Focusing on the differences between two possible approaches in the implementation of the separation routine, we compare the resulting limitations. We perform test runs on artificial data and conclude that separation in the wavelength domain is more versatile in several aspects, while the computational efficiency of separation in the Fourier domain allows working with larger data sets which is beneficial in a fully-blown disentangling process.

PACS numbers: 97.80.Fk, 39.30.+w UDC 531.755

Keywords: double-lined spectroscopic binary systems, composite spectra, disentangling, test runs, wavelength domain, Fourier domain

1. Introduction

Binary stars are the principal source of fundamental stellar quantities. In particular, stellar masses may be extracted only from observations of binary stars. To compare empirical stellar data to the theoretical stellar evolutionary models, high

[1]Corresponding author

accuracy is needed, with uncertainties of 1 – 2%. Moreover, the new generation of models which incorporate the effects of rotation, mixing and/or diffusion in stellar interiors, would require even more accurate data.

In practice, empirical data are coming from combined analysis of light and radial-velocity curves. While conceptionally simple, radial velocity measurements are in general hampered by a number of difficulties, e.g. line blending, large rotation, component confusion, signal-to-noise ratio (S/N), etc. Big advance has been achieved through the application of the cross-correlation technique (cf. Hill [1]). However, the method still suffers from several drawbacks, the need for template spectra being the most serious one. Confusion in the observed composite spectra commonly limits its analysis and has motivated the development of the spectral separation techniques. Simon and Sturm [2], and independently Hadrava [3], developed the *spectral disentangling technique* that self-consistently extracts component spectra and radial velocities without the use of template spectra. Initial applications have shown the power of the technique [4, 5]

Hensberge, Pavlovski and Verschueren [6] have made use of the disentangling technique in an elaborate study of the young, massive binary system V578 Mon in the Rosette Nebula cluster. Their study relies on high-quality observations secured at the European Southern Observatory. Special attention has been given to the homogenized and proper normalization of high-resolution echelle spectra since it was discovered that the disentangling process is very sensitive to the placement of the continuum. In a subsequent study, an error analysis of the disentangling technique has been initiated (Ilijić, Hensberge and Pavlovski [7]). The present work is a continuation of the effort focusing on the comparative study of the separation routines used for spectral disentangling.

2. *Separation techniques*

2.1. *Assumptions and goal*

The separation of the components of a composite spectrum is presently formulated in a mathematical context that implies simple physics. The observed spectrum is described as a sum of intrinsic spectra that do not vary with time except for wavelength shifts due to the orbital motion of stars. However, the light ratio between the components may vary with time. In practice, this allows to describe wide binaries (strictly speaking when out of eclipse or in mid-eclipse) and systems with weak tidal deformations. Applications to systems with spectrum variables or close binaries would require to analyze the consequences of the assumptions in more detail.

Writing $J_j(\lambda)$ for the intensity of radiation of the stellar system as observed at time labeled j and at wavelength λ, and $I_k(\lambda)$ for the intensity of radiation of component star labeled k (for a binary, $k = 1, 2$) as would be observed if it were not eclipsed and performed no orbital motion, we have

$$J_j(\lambda) = \sum_k \epsilon_{kj} \; I_k((1-\beta_{kj})\lambda)\,, \tag{1}$$

where ϵ_{kj} is the *eclipse factor* to the component star k at time j (unity if not eclipsed and less than that otherwise), and β_{kj} the *radial velocity* (RV) due to orbital motion of star k at time j (as a fraction of the speed of light)[2]. As observed spectra are most often available in the rectified (continuum-normalized) form, we write $J_j(\lambda) = J^*_j(\lambda)\; y_j(\lambda)$ for the composite spectra of the system and $I_k(\lambda) = I^*_k(\lambda)\; x_k(\lambda)$ for the spectra of component stars, where J^*_j and I^*_k are the continua and y_j and x_k are the rectified spectra. Equation (1) is now expressed in terms of rectified spectra

$$y_j(\lambda) = \sum_k \ell_{kj}(\lambda)\; x_k((1-\beta_{kj})\lambda)\,, \tag{2}$$

where

$$\ell_{kj}(\lambda) = \epsilon_{kj}\, \frac{I^*_k((1-\beta_{kj})\lambda)}{J^*_j(\lambda)} \approx \frac{\epsilon_{kj}\, I^*_k(\lambda)}{\sum_q \epsilon_{qj}\, I^*_q(\lambda)} \tag{3}$$

is the wavelength dependent *light factor* (LF) expressing the fractional contribution of continuum of component k at time j to the composite continuum around wavelength λ. The approximate equality in Eq. (3) comes from the assumption that the continua do not change significantly on the wavelength scale of the Doppler shifts due to orbital RVs.

Within this model, separation can be formulated as a linear least-squares problem where a time series of composite spectra (observed spectra), $y_j(\lambda)$, of an unresolved multiple stellar system together with β_{kj} and $\ell_{kj}(\lambda)$ (RVs and LFs) of component stars are used to extract their intrinsic spectra, $x_k(\lambda)$, (model spectra).

2.2. Separation in the wavelength domain

Supposing that the data set consists of a time series of rectified observed spectra labeled $j = 1, \ldots, N_{\rm obs}$, possibly sampled on different grids of wavelength points $\lambda_{ji}, i = 1, \ldots, N_j$, and denoting $y_{ji} = y_j(\lambda_{ji})$ and $\ell_{kji} = \ell_{kj}(\lambda_{ji})$, Eq. (2) may be written

$$y_{ji} = \sum_k \ell_{kji}\; \bar{x}_k((1-\beta_{kj})\lambda_{ji}, \{x_{k\alpha}\})\,, \tag{4}$$

where $\bar{x}_k(\lambda, \{x_{k\alpha}\})$ is the amplitude of the model spectrum to component k at wavelength λ, estimated from the set of free parameters $x_{k\alpha}$, $\alpha = 1, \ldots, M_k$. Note that the estimator must be able to provide the amplitude at all wavelengths and

[2] Note the astronomical convention where *radial velocity* (RV) of an object is the projection of its velocity onto the line of sight; positive RV is assigned to a receding object.

times (orbital phases) covered by the data, so due to orbital Doppler shifts it must be defined on a somewhat wider wavelength range than that of the data. If the free coefficients $x_{k\alpha}$ are chosen to be amplitudes of the rectified model spectrum sampled on a grid of wavelength points, then the estimator $\bar{x}_k$ is linear in $x_{k\alpha}$ and Eq. (4) represents a system of $\sum_j N_j$ coupled equations linear in $\sum_k M_k$ unknowns. To be useful in practice, the system must be over-determined and is solved by requiring the least-squares solution with each equation weighted by σ_{ji}^{-2} (the inverse square of the standard deviation of the data point obtained through data reduction).

Bagnuolo and Gies sampled both the observed and the model spectra on the same logarithmic grid of wavelength points[3] and used the estimator that picks the amplitude from the bin of the model spectrum closest to the required wavelength [8]. The system (4) was solved iteratively through normal equations. The technique, labeled *tomographic separation* by its authors, was successfully applied in the series of works on hot binaries, the most recent of which is Ref. [9].

Simon and Sturm used logarithmic binning as well, but applied a more advanced estimator (linear interpolation of amplitudes in two data bins of the model embracing the required wavelength, which is realistic for oversampled spectra) and used different resolutions for observed and for model spectra [2]. Due to the rank deficiency of the system of equations, they recommend the use of the *singular value decomposition* (SVD) to reach the least-squares solution of the system of equations; another reason to use SVD is the likely occurence of numerical instability due to the size of the matrix when solved through normal equations [10]. The separation technique of Simon and Sturm was the first to be used for self-consistent extraction of model spectra and determination of RVs (disentangling, see Sect. 2.4) [2, 4, 5].

The formulation of the separation in the wavelength domain leaves space for including weighting/masking at any wavelength in the minimization procedure, as well as continuously variable LFs, but at present these possibilities are not yet exploited in full generality.

2.3. Separation in the Fourier domain

Hadrava used the discrete Fourier transform (DFT) of spectra to uncouple the large system of equations (4), simplifying the numerical solution [3]. This may prove important when long stretches of observed spectra are considered. But the method offers less flexibility with regard to observational complications. Especially, when shorter pieces of spectrum are considered, it is not always possible to cut the observed spectra in such a way that sufficient continuum is enclosed at the edges and as a consequence end-of-range effects (ERE), well-known in many DFT applications, enter. Also, weighting is limited to an observed spectrum as an entity, so it is impossible to 'mask off' particular regions in the data.

From the viewpoint of Eq. (4), this approach requires all spectra (observed and model) be sampled on the same logarithmic wavelength grid, $N_j, M_k \rightarrow N_{\rm bin}$; stan-

[3]Use of binning equidistant in the logarithm of the wavelength is a common practice because the Doppler shift can be expressed in number of data bins independently of wavelength.

dard deviation of the amplitudes of observed spectra (in the wavelength domain) must be equal within each observed spectrum, $\sigma_{ji} \to \sigma_j$; model spectra are considered periodic in wavelength with the period being the used range in wavelength, and LFs are not allowed to depend on wavelength, $\ell_{kji} \to \ell_{kj}$. Taking the DFT of both sides of Eq. (4), and using the uppercase symbols to denote the DFTs of the spectra, one obtains

$$Y_{ji} = \sum_k \ell_{kj} \; X_{ki} \; \mathrm{e}^{\mathrm{i}\,\Omega_i \Delta_{kj}} \,, \tag{5}$$

where index $i = 0, \ldots, \frac{1}{2}N_{\mathrm{bin}}$ runs over the discrete set of frequencies $\Omega_i = 2\pi i/N_{\mathrm{bin}}$ (in radians per data bin)[4] RVs are specified through $\Delta_{kj} = \beta_{kj}/\beta_{\mathrm{grid}}$, where β_{grid} is the RV corresponding to one-bin shift on the logarithmic wavelength grid (RV-resolution of the grid). The estimator of the amplitude of the Doppler shifted model spectrum obtained this simple form by taking the DFT of the estimator expressed in wavelength domain as a convolution of the model spectrum with a shifted delta function. Equation (5) represents $\frac{1}{2}N_{\mathrm{bin}}{+}1$ systems of N_{obs} coupled equations, linear in as many unknowns as there are stars in the system. If there are more observed spectra than components in the multiple system, the subsystems of equations are over-determined and least-squares solutions are required for each subsystem. Each equation is weighted according to the weight assigned to the respective observed spectrum (with the possibility of distributing the weights among spectra differently at different Fourier frequencies). We recommended to use SVD to solve the systems of equations because the use of normal equations (as in Ref. [3]) may lead to singular or nearly singular matrices (see Sect. 4).

While the advantages in simplicity of implementation and the CPU-time requirements of (5) over (4) are obvious, the limitations of Fourier domain separation have to be considered carefully with each application. The separation routine based on (5) passed several tests on artificial data and behaved well in practice when used as part of the disentangling algorithm (see Ref. [7] and references therein).

2.4. A note on disentangling

The separation technique, as defined in the preceding sections, determines the model (component) spectra by minimizing the weighted sum of squares of residuals of the observed spectra and the composite spectra resulting from the model. Starting from Eq. (4), the minimized quantity is

$$s^2 = \sum_{ji} \frac{1}{\sigma_{ji}^2} \left| \, y_{ji} - \sum_k \ell_{kji} \; \bar{x}_k((1-\beta_{kj})\lambda_{ji}, \{x_{k\alpha}\}) \right|^2 . \tag{6}$$

Simon and Sturm assumed that s^2 obtained during separation with the 'true' β_{kj} is smaller than it would be obtained with β_{kj} deviating from the 'true' values. They

[4] It has been assumed that the sign of the exponent in the definition of DFT is positive and that N_{bin} is even. Only the positive frequency half of the true DFT is treated because the input data are real.

introduced *disentangling* as a non-linear least-squares problem where both β_{kj} and $x_{k\alpha}$ are free parameters [2] in order to determine RVs and component spectra self-consistently. While the assumption has not yet been proved mathematically, both practical experience (including the comparison with directly observed spectra taken in full eclipse) and error analysis support that it is valid [11, 7].

3. Tests on artificial data

The properties of the separation techniques can be demonstrated through numerical tests. We use artificial data and conduct Monte-Carlo simulations of repeated experiments. Rotationally broadened synthetic spectra of hot stars are combined into artificial spectra of a double-lined binary system. For the primary (brighter) star we take $T_{\rm eff} = 30\,000$ K and $v \sin i = 120$ km s^{-1}, and for the secondary (fainter) $T_{\rm eff} = 27\,500$ K and $v \sin i = 100$ km s^{-1}. The orbit is taken to be circular with RV semi-amplitudes $K_1 = 185$ km s^{-1} and $K_2 = 260$ km s^{-1}. Rectified spectra are combined with wavelength independent LFs $\ell_1 = \frac{3}{5}$ and $\ell_2 = \frac{2}{5}$ and Doppler shifts corresponding to orbital RVs at phases $\frac{1}{4}$, $\frac{3}{5}$ and $\frac{4}{5}$, and with $\ell_1 = \ell_2 = \frac{1}{2}$ at phase 0 (primary eclipse factor $\epsilon_1 = \frac{2}{3}$)[5]. Our numerical tests consist of creating 400 data sets differing only in their random part (Gaussian noise) and subjecting them to separation. We examine the average and the standard deviation of the model spectra that we obtain focussing on the systematic error that may be present. Tests were performed in different spectral regions, at different resolutions and S/N ratios. Our implementations of the separation routines operating in the wavelength-domain and the Fourier-domain were used. Here, we present two typical situations.

In our first example, a case that obeys the assumptions made in the Fourier-domain separation (ends of spectra in line-free regions with length of the order of K_1 and K_2; no blemishes; data with homogeneous S/N), the Fourier-domain separation and the wavelength-domain separation deliver almost identical results. As shown in Fig. 1, the $\pm 1\sigma$ regions obtained with Fourier-domain separation are symmetrically distributed around the synthetic spectrum.

In the second example, we show that significant systematic errors may enter the model if the Fourier-domain separation is applied on data that are *not* in accord with its validity assumptions. As can be seen in Fig. 2, the red end of the wavelength range was not placed in the continuum region and the model spectra obtained through Fourier-domain separation are consequently biased at both ends. We also use this example to show the effect that a blemish in the observed spectra may have on the model; a static absorption feature was introduced at 4096 Å (409.6 nm), and the model was evidently perturbed by that as well. In contrast to this, there was no systematic error in the model spectra obtained from the same data set through the wavelength-domain separation what is due to two reasons: (i) wavelength-domain

[5] The physical parameters of the binary are similar to those determined for V578 Mon [6], with the exception that the rotational velocity adapted in our model is higher (what complicated the disentangling process.

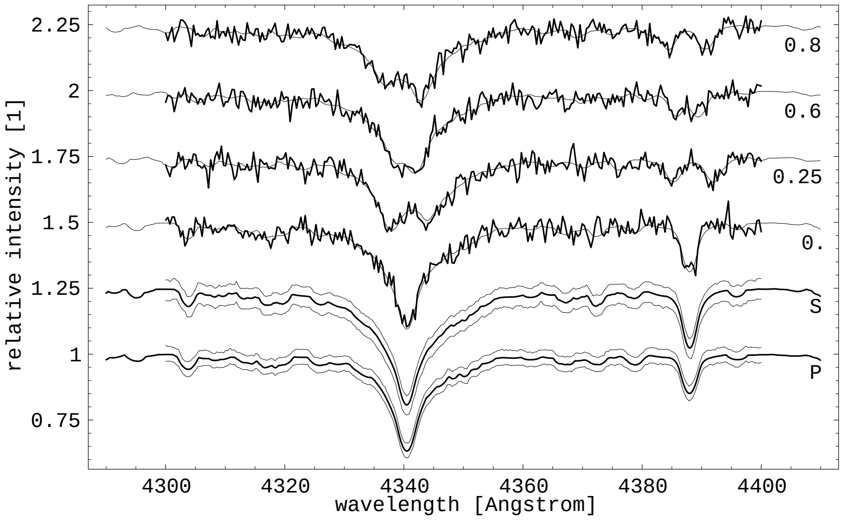

Fig. 1. Fourier domain separation: Hγ *and* He I *lines,* $\beta_{\rm grid} = 25$ km s^{-1}, $S/N = 40$. *Synthetic component spectra (**h**eavy lines labeled P and S) are combined into composite spectra (**t**hin lines labeled with orbital phase). 400 artificial data sets are created (**o**ne shown in heavy lines) and subjected to separation. The resulting* $\pm 1\sigma$ *region is shown (**t**hin lines). All spectra are plotted on the same scale; shifts of 0.25 in relative intensity are applied to the spectra.*

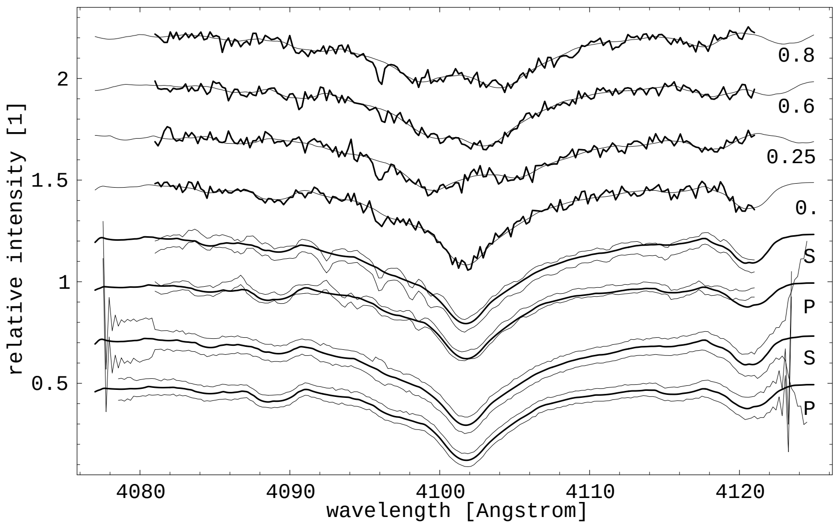

Fig. 2. Wavelength domain vs. Fourier domain separation: Hδ *line,* $\beta_{\rm grid} = 15$ km s^{-1}, $S/N = 50$. *The* $\pm 1\sigma$ *regions obtained with wavelength domain separation (**t**he two bottom spectra) and with Fourier domain separation (**t**he next two spectra) are shown (see also caption to Fig. 1).*

separation is not sensitive to a particular placement of ends of spectral range (it extends the model spectra as far as necessary at both ends instead of limiting them to the length of the data and treating them as periodic) and (ii) the region affected by the static feature was 'masked off' during separation (zero-weight, $\sigma_{ji}^{-1} = 0$, was assigned to the region around 4096 Å (409.6 nm) which resulted in a somewhat higher σ in the model, but no systematic error).

4. *Some practical remarks*

In practice, it is likely that all spectra of a double-lined binary system are observed out of eclipse and that both time and wavelength independent LFs are used for the separation; $\ell_{1j}(\lambda) \to \ell_1$ and $\ell_{2j}(\lambda) \to \ell_2$. In such a situation, irrespective of the choice of the separation technique, model spectra can be determined only up to a (coupled) constant. This is most clearly seen from Eq. (5) if we take the frequency-zero ($i = 0$) subsystem of equations that determines X_{10} and X_{20} (the constant terms in model spectra); through the least-squares solution of the subsystem, only the weighted sum $\ell_1 X_{10} + \ell_2 X_{20}$ can be determined (normal matrix is singular [3]). Therefore, c/ℓ_1 may be added to $x_1(\lambda)$ if c/ℓ_2 is correspondingly subtracted from $x_2(\lambda)$, where c is an arbitrary constant.

Interstellar absorption can easily be included into the model as a wavelength dependent *absorption factor* in front of the sums in (1) or (2), and to the extent this absorption factor is known, data can be corrected for it prior to separation. However, as absorption enters the right hand sides of (1) and (2) in a multiplicative way it is not possible to model it in terms of free coefficients together with the spectra of model stars and still end up with a linear least-squares problem. Telluric lines and other spectral features that do not belong to (move with) component stars are even more difficult to deal with because their intensities change with time, and thus from exposure to exposure. Separation in wavelength domain can take advantage of the possibility of masking out telluric lines assigning relatively higher standard deviation (lower weight) to affected spectral regions.

5. *Conclusion*

Disentangling techniques have lead to the most precise determination of mass, radius and other fundamental stellar parameters in well-detached binaries. In addition, studies of the chemical composition of individual components in binaries have been shown to be possible on disentangled spectra if the composite spectra were carefully reduced in a differential way [12]. Indeed, systematic reduction errors are to our present knowledge the main limiting factor for a successful disentangling. While the different disentangling techniques have proven useful, and several tests on artificial data have been made, a larger experience on a variety of cases is needed. Claims that non-additive components as terrestrial atmospheric absorption lines or interstellar diffuse or line absorptions can be disentangled can, in our opinion, only be true approximatively. This claim is evidently important for projects planning to apply the method on extragalactic data. A combination of DFTs while iterating to the best solution and a final application in the log-wavelength space, exploiting its flexibility to take into account observational blemishes, appears presently to us to be the best choice.

References

[1] G. Hill, in *New Frontiers in Binary Star Research*, eds. K. C. Leung and I.-S. Nha, ASP Conf. Ser. **38** (1993) p. 127.

[2] K. P. Simon and E. Sturm, Astron. Astrophys. **281** (1994) 286.

[3] P. Hadrava, Astron. Astrophys. Suppl. Ser. **114** (1995) 393.

[4] E. Sturm and K. P. Simon, Astron. Astrophys. **282** (1994) 93.

[5] K. P. Simon, E. Sturm and A. Fiedler, Astron. Astrophys. **292** (1994) 507.

[6] H. Hensberge, K. Pavlovski and W. Verschueren, Astron. Astrophys. **358** (2000) 553.

[7] S. Ilijić, H. Hensberge and K. Pavlovski, in *Astrotomography, Indirect Imaging Methods in Observational Astronomy*, eds. H. M. J. Boffin, D. Steegs and J. Cuypers, Lecture Notes in Physics Vol. **573**, Springer (2001) p. 269.

[8] W. G. Bagnuolo and D.R. Gies, Astrophys. J. **376** (1991) 266.

[9] L. R. Penny, D. Seyle, D. R. Gies, J. A. Harvin, W. G. Bagnuolo, M. L. Thaller, A. W. Fullerton and L. Kaper, Astrophys. J. **548** (2001) 889.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN*, 2nd ed., Cambridge University Press, Cambridge (1992).

[11] R. I. Hynes and P. F. L. Maxted, Astron. Astrophys. **331** (1998) 167.

[12] K. Pavlovski and H. Hensberge, in *Poster Proc. IAU Symp. 200 'Birth and Evolution of Binary Stars'*, eds. B. Reipurth and H. Zinnecker, IAP (2000) p. 109.

TEHNIKE RAZDVAJANJA ZA RASPETLJAVANJE SLOŽENIH SPEKTARA

Raspetljavanje složenih spektara je obećavajuća metoda za proučavanje dvojnih zvjezdanih sustava. Ta metoda primjenjuje rutinu za razdvajanje kojom se izlučuje spektre zvijezda komponenata modela iz vremenskog niza opaženih složenih spektara. Usredotočujući se na razlike u pristupu pri izvedbi rutine za rastavljanje, uspoređujemo ograničenja koja iz njih proizlaze. Proveli smo ispitivanje s umjetnim podacima i zaključili da je rastavljanje u području valnih duljina svestranije u više pogleda, dok rastavljanje primjenom Fourierovog transformata dozvoljava rad s većim skupovima podataka, što je povoljno u potpunom procesu raspetljavanja složenih spektara.