Searching double-peaked emission line profiles in the spectra of galaxies through the symmetry of the cross-correlation function

B. García-Lorenzo1,2⋆

1 Instituto de Astrofísica de Canarias, C/Vía Lactea S/N, 38350-La Laguna, Tenerife, Spain
2 Dept. Astrofísica, Universidad de La Laguna, C/Astrofísico Francisco Sánchez, E-38205 Tenerife, Spain

Abstract

The presence of double-peaked/multicomponent emission line profiles in spectra of galaxies is commonly done by visually inspection. However, the identification of complex emission line profiles by eye is unapproachable for large databases such as the Sloan Digital Sky Survey (SDSS) or the integral field spectroscopy surveys of galaxies (e.g. CALIFA or MaNGA). We describe a quick method involving the cross-correlation technique for detecting the presence of complex (double-peaked or multiple components) profiles in the spectra of galaxies, deriving simultaneously a first estimation of the velocity dispersions and radial velocities of the dominant gaseous component. We illustrate the proposed procedure with the well-known complex \([\text{[OII]} \lambda \lambda 4959,5007]\) profiles of the central region of NGC 1068.

Key words: cross-correlation — radial velocities — line: profiles — galaxies: surveys

1 Introduction

Early spectroscopic studies of galaxies showed that profiles of emission lines could present asymmetries, shoulders, or double peaks (Heckman et al. 1981; Pelat & Alloin 1980; Glaspey et al. 1976; Seyfert 1943, e.g.). These features come from the light of several gaseous systems with different kinematics that is integrated along the observer's light of sight.

They have been interpreted as due to rotating gaseous disks, outflows/inflows or dual active galactic nuclei (Shen et al. 2011; Greene & Ho 2005; Zhou et al. 2004; Arribas et al. 1996, e.g.).

Systematic search of double-peaked emission line profiles in large spectra databases (e.g. Sloan Digital Sky Survey, SDSS (York et al. 2000)) have been performed using different selection criteria, but the final confirmation is usually done by visual inspection (Ge et al. 2012; Pilugin et al. 2014; Smith et al. 2014; Liu et al. 2010). In this work, we present an automatic procedure to detect multi-component emission line profiles in large databases of spectra of galaxies based on the symmetry of the cross-correlation function.

The cross-correlation (C-C hereafter) technique has been extensively used in astronomy to infer radial velocities (Anglada-Escudé & Butler 2012; Westfall et al. 2011; Allende Prieto 2007; Frerichter & Melia 2004; Gunn et al. 1996; Storm et al. 1992; Dalle Ore et al. 1991; Tonry & Davis 1979). The C-C technique globally compares a problem spectrum with a reference spectrum or template. The C-C of two spectra analyses the similarity between one spectrum and a wavelength (or velocity) shifted version of the other, as a function of this wavelength (or velocity) shift. The C-C function only contains the frequencies that are common to both spectra. Therefore, the C-C function provides a clear indication of the shift at which the two spectra are most similar and also a quantitative measure of that similarity. As long as the largest peak of the C-C function is symmetric, it might be used to derive the shift of the problem spectrum as well as its velocity dispersion in conjunction with the width of the template (Tonry & Davis 1979). When the two spectra are the same, the C-C function becomes the auto-correlation.

We propose a methodology for searching multi-component/double-peaked emission line profiles in the spectra of galaxies based on the deviation from symmetry of the peak of the C-C function. Details on the C-C technique can be found in Furenlid & Furenlid (1990); Tonry & Davis (1979) and references therein. The C-C algorithm might be used in the search for binary active galactic nuclei in large spectra databases (e.g. SDSS). It may be also used to locate spaxels with spectra showing multicompo-

⋆ E-mail: bgarcia@iac.es
nent emission line profiles in Integral Field Spectroscopic (IFS) surveys of galaxies (e.g. the on-going Calar Alto Legacy Integral Field Area Survey, CALIFA (Sánchez et al. 2012) or The Mapping Nearby Galaxies at APO, MANGA (http://dunlap.utoronto.ca/research/surveys/)).

2 CROSS-CORRELATION TECHNIQUE FOR SEARCHING DOUBLE-PEAKED LINE PROFILES

2.1 Cross-correlation function shape traced by bisectors

The procedure developed in this work is based on the symmetry of the C-C function nearby its main peak. The estimation of the shift and velocity dispersion of the spectra of galaxies might be obtained by fitting a smooth symmetric function to the peak of the C-C function (Tonry & Davis 1971). However, C-C is not an even function and therefore, this approach is only valid when the shapes of the spectral features in the problem spectrum are similar to those in the reference spectrum. Emission lines in the spectra of galaxies are commonly assumed to be fitted by Gaussian profiles. If single Gaussian profiles are assumed for emission lines in the template and problem spectra, the velocity shift and velocity dispersion of the problem spectrum can be derived by fitting a single Gaussian to the peak of the C-C function (see Fig. 1a,b). However, if the emission line profiles in the reference and in the problem spectrum are significantly different, the peak of the C-C function will be asymmetric (see Fig. 1c,d).

The reference spectrum must have a large signal-to-noise ratio and a spectral resolution similar to or better than the problem spectrum. As the C-C of two spectra is similar to their convolution, the spectral resolution of the C-C function might be significantly reduced during the C-C operations. This results in a significant reduction in the capability of detecting double-components (see Fig. 1b). However, the template can be selected to have a negligible velocity dispersion compared to the problem spectrum. Indeed, a variation of the C-C technique using delta functions has been successfully used for detection of binary stars (Furenlid & Furenlid 1990). In this case, the C-C spectral resolution, and hence the capability of detecting double-components, strongly depends on the problem spectrum spectral resolution. Moreover, the shape of the C-C peak will correspond to the average shape of the line profiles in the problem spectrum (figures 1b and 1d). Therefore, the presence of double or multiple components in emission line profiles is reduced to study the shape of the C-C peak profile.

The C-C peak profile can be characterized by tracing its bisector. Bisector shapes are commonly used in the analysis of the mechanisms that cause asymmetries and variations in stellar spectra (Bastürk et al. 2011, e.g.). The bisector of a symmetric profile should remain at constant wavelength (or velocity) for all parts of the profile, dividing it into two equal parts; any existing asymmetry between the base and the peak of the line will remain reflected in the shape of the bisector. Bisectors for any profile can be constructed by connecting the midpoints of horizontal line segments spanning the width of the profile at a number of intensity positions inside the profile. The comparison of the C-C peak profile bisector \( (V_b) \) with the C-C peak velocity \( (V_p) \), the remained value as if it were symmetric) gives a reference value to identify the presence of various gaseous components in the profiles. This procedure does not infer the actual number of gaseous components forming the observed profile, but an evidence of the presence of various gaseous components (at least two). The sign of \( V_r - V_p \) also gives information about the equivalent wavelength (velocity) of the double/multiple components relative to the dominant component (wavelength or velocity at the C-C peak); if the sign is positive, the secondary (or equivalent) component is redshifted respect to the dominant component. A positive detection of double/multiple components in a galaxy spectrum will depend on the signal-to-noise of the C-C function and hence on the signal-to-noise of the problem spectrum. Figure 2 shows the bisectors of the C-C peak functions for the examples in figure 1. Tracing the bisector on a single emission line profile would also provide evidences of the presence of double/multiple components forming that profile. However, any observational or instrumental signature (e.g. cosmic-ray) not properly removed during the data reduction process and affecting the selected single emission line could result in an asymmetric line bisector and hence in an spurious double/multiple component detection. Including several emission lines in the selected cross-correlation spectral range will smoothed any undesirable feature affecting a single-line since the shape of C-C peak function will correspond to their average shape. Moreover, as the C-C function also provides a quantitative measure of the similarity between the problem and template, other advantage of tracing the bisector on the C-C peak function instead of on single emission line profiles is that any shape of the line profiles can be selected to generate the template. Here, we have assumed gaussian profiles for the emission lines to create the template but any other profile shape could be assumed (e.g. Lorentz profiles).

2.2 Implementation of the procedure

The C-C technique uses the information contained in all the lines of the selected spectral range for calculating the C-C function. The resulting shape of the C-C peak corresponds to an average shape of the line profiles in the problem spectrum.

In order to calculate the C-C function, it is necessary to prepare the problem spectrum following these steps:

1. The ends of the C-C spectral range are selected at the continuum to avoid discontinuities at the edges.
2. The data are divided by the continuum to remove any residual curvature in the continuum.
3. In order to minimize edge effects, the continuum is subtracted (just subtract the unity when step 2 was performed).

When a stellar fit had been carried out previously and properly subtracted from the original spectra, steps 2 and 3 can be stripped.

The reference spectrum is generated including as many delta functions as single emission lines are expected in the spectral range selected to apply the C-C technique. The wavelength (or velocity) position of these delta functions corresponds to the center of the emission lines. Different weights (in flux) can be given to the delta functions to
Figure 1. Example of the cross-correlation technique applied in the [OIII]λλ4959, 5007 spectral range, assuming Gaussian emission line profiles and a redshift of about 6610 km/s. Spectral resolution for this model is 1650. (a) Reference spectrum (blue-dashed lines) shifted at the assumed redshift, and problem spectrum (black line) at redshift-400 km/s. Velocity dispersions for problem and template spectrum are 115 km/s. (b) The cross-correlation function of the spectra in (a). The green-dashed lines correspond to the Gaussian profile fit to the peak of the C-C function. The Gaussian centroid (at -390 km/s) and full-width-half-maximum (384.17 km/s) provides the velocity shift between problem and template and the estimation of the problem velocity dispersion, respectively. (c) Problem spectrum (black line) showing double components. The brightest component is described by the problem spectrum (V\textsubscript{sys}=6210 km/s and σ=115 km/s) in (a), and secondary component is at V\textsubscript{sys}=6710 km/s with a σ=115 km/s. Both components have been selected to be clearly resolved. Blue-dashed line is the template (the same template than in (a)). (d) The C-C function of the problem and template spectra in (c). The C-C peak clearly has an asymmetric shape although a double-peak is not resolved due to the degradation of the spectral resolution in the calculation of the C-C function. The Gaussian fit to the asymmetric C-C peak provides a velocity shift between template and problem spectra of -289 km/s, and the estimation of the velocity dispersion for the problem is 235 km/s. (e) The velocity dispersion of the template (blue line) is ~ 8 km/s, negligible compared to the velocity dispersion of the emission lines (σ=115 km/s) in the problem spectrum (black line). (f) The C-C function of the problem and template spectra in (e). The asymmetric C-C peak profile is now resolved in two peaks. Gaussian fit to the peak of the C-C function provides and estimation of redshift and velocity dispersion of -311 km/s and 225 km/s, respectively, for the problem spectrum.
account for fix intensity ratio between emission lines according to atomic parameters. Noise contribution is not desirable to be included in the template to avoid degradation of the signal-to-noise when the C-C function is computed. Both template and problem spectra are normalized to the maximum of the brightest emission line in the selected wavelength. When the reference has been generated and the problem spectrum has been rectified, the cross-correlation function can be computed. The maximum of the cross-correlation function is located and a Gaussian is fitted.

When asymmetries due to double components are presented in the problem spectrum, the bisector of the C-C peak function is twisted and differs from the velocity at the peak (see (b) and (c)). The difference between the velocity at the peak of the C-C function and the bisector at a determined intensity level indicates the presence of asymmetries (double or multiple components) in the problem spectrum profile. The sign of this difference indicates if these asymmetries are blue or red-shifted respect to the dominant component (peak of the C-C function).

![Figure 2. Bisectors (blue dashed-lines) of the C-C peak functions in Fig. 4d. Black dotted line corresponds to the peak velocity dividing the profile in two parts.](image)

When the emission lines profiles in the template and problem spectrum are similar (Gaussian profiles are assumed) black dotted line divides the C-C peak function profile in two symmetric parts. In this case, bisector (blue dashed-lines) and peak velocity extension to different intensities (black dotted-lines) are similar (see (a)). When asymmetries due to double components are presented in the problem spectrum, the bisector of the C-C peak function is twisted and differs from the velocity at the peak (see (b) and (c)). The difference between the velocity at the peak of the C-C function and the bisector at a determined intensity level indicates the presence of asymmetries (double or multiple components) in the problem spectrum profile. The sign of this difference indicates if these asymmetries are blue or red-shifted respect to the dominant component (peak of the C-C function).

Obviously, the level of noise in the problem spectrum affects the detection of asymmetries. In this sense, a signal-to-noise threshold is defined to reduce the number of false positive detections of double/multiple components on emission line profiles in the spectra of galaxies. Removing bisector lower levels might also help to avoid false positive detections due to poor signal-to-noise of the spectra. In practice, we consider a positive asymmetry detection (blue, red, or multi-asymmetries) when \( N_{\text{asym}} \) is equal or larger than 2, that is, at least two intensity levels tracing the bisector of the C-C peak function must satisfied that \( |V_p-V_p(i)| > \Delta V \).

2.3 Application to integral field spectroscopy

In this section, we present some examples to illustrate the results from the proposed procedure for searching double or multiple gaseous components in the spectra of galaxies applied on IFS data.

2.3.1 The data

We have used the IFS data of the central 24′′ × 20′′ of the Seyfert 2 galaxy NGC 1068 presented in García-Lorenzo et al. (1999). NGC 1068 is the nearest and brightest example of a barred galaxy with an active galactic nucleus. The presence of several kinematically distinct gaseous components in the central region of NGC 1068 is evident from the large amount of emission-line profiles obtained using different instruments and techniques: Ozaki 2004; Emsellem et al. 2004; Gerssen et al. 2004; Ishigaki et al. 2004; Groves et al. 2004; Cecil et al. 2002; García-Lorenzo et al. 1999; Arribas et al. 1998; Cecil et al. 2002.
Careful visual examination (see Fig. 3 and description of the spectra [Emsellem et al. 2006]) revealed the presence of a minimum of three different gaseous systems in NGC 1068. When studying IFS data, the visual inspection of the spectra is a tedious task but accessible for examining a several hundred of spectra. This is not the case for the large amount of spectra provided by IFS surveys (e.g. CALIFA or MaNGA surveys) of galaxies.

Figure 3 shows the [OIII]λλ4959, 5007 line profiles at each observed position of NGC 1068 as a spectra diagram [García-Lorenzo et al. 1999]). These profiles offer a large variety of examples to test the developed algorithm.

2.3.2 Results for NGC 1068

Applying the proposed procedure of studying the symmetry of the C-C peak function explained in section §2 to the spectra from the central region of NGC 1068, we can point out the following:

- Only 12% of the emission lines profiles in the central 24″×30″ of the Seyfert 2 galaxy NGC 1068 (Fig. 3) seems to present symmetric profiles. The percentage of symmetric [OIII]λλ4959, 5007 profiles is 24.5% for those spectra with a signal-to-noise larger than 200.
- 6% of the total number of spectra presents pure red-asymmetries. Red asymmetries are detected in the 26% of spectra with a signal-to-noise larger than 200.
- 51% of the total number of spectra presents pure blue-asymmetries. This percentage is 31.5% for those spectra with a signal-to-noise ratio > 200.
- 31% of the spectra in the central region of NGC 1068 present asymmetries, to the blue or red depending on the bisector intensity level. In the case of spectra with a signal-to-noise larger than 200, the percentage of multi-asymmetric emission line profiles is 18%.

Examples of spectra (the same spectra than in Fig. 7 in García-Lorenzo et al. [1999]) displaying asymmetries and the bisectors shape of the C-C peak function are shown in Fig. 4. The [OIII]λλ4959, 5007 profiles in Fig. 4a) show a blue shoulder (corresponding to component 4b in García-Lorenzo et al. [1999]) and the additional component in Emsellem et al. (2006)). The shape of the C-C function for this spectrum (Fig. 4a1) and (a2)) also presents a clear blue asymmetry that is traced by the shape of its bisector. The peak of the C-C function indicates a radial velocity of 1145 km/s for the dominant component in this profile, which is in agreement to the systemic velocity, 1144 km/s (Emsellem et al. 2006)). The difference in velocity between the bisector at a 10% peak intensity level and the velocity at the peak of the C-C function (V_p(10%)-V_p), is -302 km/s, indicating a blue asymmetry. The [OIII]λλ4959, 5007 emission lines in Fig. 4b) presents double-peaked [OIII] profiles that it is also reproduced in the C-C peak function. The radial velocity of the dominant component in this profile (velocity at the peak of the C-C function) is 1618 km/s (474 km/s larger than the systemic velocity), which was identified as component 4r in García-Lorenzo et al. (1999) and as the additional component in Emsellem et al (2006). The bisector of the C-C peak function clearly indicates the presence of a secondary bluer component in this profile. The difference V_p(10%)-V_p is -485 km/s. This secondary component was identified as component 1+3 in García-Lorenzo et al. (1999) and as the narrow component in Emsellem et al. (2006). Emission lines in Fig. 4c) show a peaked profile with a clear red shoulder that is reproduced by the shape of the C-C peak function bisector. In this case, V_p=1124.52 km/s (component 1/narrow in García-Lorenzo et al. (1999)/Emsellem et al. (2006), respectively), while V_p(10%)-V_p= 209 km/s, indicating a red component identified as component 3 in García-Lorenzo et al. (1999) and broad component in Emsellem et al. (2006). For this profile, the lower level of the bisector (Fig. 4c2)), turns to the blue, V_p(5%)-V_p(10%) = -104 km/s, although still indicating a global red asymmetry (V_p(5%)-V_p = 105 km/s). This twist in the bisector indicates the presence of a faint additional component, that was identified as component 2 in García-Lorenzo et al. (1999). The visual inspection of emission line profiles in Fig. 4d) showed single profiles with not signs of double-components. The C-C function is symmetric respect to its peak, being V_p = 1071 km/s. Only at a 5% level of intensity of the C-C peak function, a blue asymmetry is detected by the proposed procedure (with a V_p(5%)-V_p = -188 km/s), while for the rest of the levels, the difference V_p(i)-V_p (for i=10%,100% in steps of 5%) is always smaller than 10 km/s. This false positive asymmetry detection is due to the low signal-to-noise at the level in which the asymmetry has been detected. In practice, this case will be labeled as symmetric in the implemented algorithm due to the condition of having at least two intensity levels with positive detection for asymmetries. Figure 3 also indicates the identified asymmetries for the observed spectra with a signal-to-noise ratio larger than 200 in the central region of NGC 1068.

The proposed procedure only indicates those spectra showing asymmetries due to double or multiple components, but the deprojection of the different gaseous systems needs any other procedure such as Gaussian fit [Arribas et al. 1997(e.g.)].

3 CONCLUSIONS

The identification of double/multiple component emission line profiles in spectra of galaxies is the first step for the sample selection of many research topics (e.g. binary black holes, ionized gas kinematic deprojection). This paper deals with a quick estimation of velocity dispersions, radial velocities of the dominant component and the detection of the presence of double or multiple components among a large set of spectra of a galaxy of galaxies using the cross-correlation technique and the shape of the C-C peak function. The proposed procedure allows processing a large amount of problem spectra in a short period of time (in the order of minutes) using the peak, full-width-half-maximum and symmetry of the C-C peak function.

ACKNOWLEDGMENTS

The author thanks A. Elf-Darwich, and A. Bongiovanni for their help and useful discussions. This work was partially funded by the Instituto de Astrofísica de Canarias.
Figure 3. Spectra diagram of the [OIII]λλ4959,5007 emission lines at 463 positions on the central region of NGC 1068 (García-Lorenzo et al. 1999). The spectra at each location are auto scaled to better show the profile shape. Emission lines nearer to the optical nucleus (taken to be the origin) are brighter than those farther out. The plotted spectral range is 4925-5075 Å, the same selected for the application of the cross-correlation technique. Different colors indicate the asymmetry detected in those profiles with a signal-to-noise larger than 200 (red: red-asymmetry; blue: blue-asymmetry; magenta: multi-asymmetries; green: symmetric profiles).

by the Spanish Ministerio de Economía y Competitividad (MINECO; grant AYA2009-12903) and by the Spanish Agencia Canaria de Investigación, Innovación y Sociedad de la Información (proID20100121).

REFERENCES
Allende Prieto, C. 2007, AJ, 134, 1843
Anglada-Escudé, G., & Butler, R. P. 2012, ApJSS, 200, 15
Arribas, S., Mediavilla, E., & García-Lorenzo, B. 1996, ApJ, 463, 509
Baştürk, Ö., Dall, T. H., Collet, R., Lo Curto, G., & Selam, S. O. 2011, A&A, 535, A17
Cecil, G., Dopita, M. A., Groves, B., et al. 2002, ApJ, 568, 627
Cecil, G., Bland, J., & Tully, R. B. 1990, ApJ, 355, 70
Dullea, C., Faber, S. M., Jesus, J., Stoughton, R., &
Figure 4. (Left-column) Examples of emission line profiles of NGC 1068. Spectra are located at positions (a) [+0.02,+3.49]; (b) [-1.02,-0.68] arcsec; (c) [+2.97,+4.08] arcsec; and (d) [-8.86,6.24] arcsec from the optical nucleus. (Central-column) (a1),(b1),(c1), and (d1) The cross-correlation functions of the spectra in the left-column ((a),(b),(c), and (d)) using two delta function as template spectrum in the spectral range including the [OIII]$\lambda\lambda 4959, 5007$ emission lines (4925-5075 Å). (Righ-column) The cross-correlation peak function (zoom of function in central column). Dashed line corresponds to the traced bisector.
Burstein, D. 1991, ApJ, 366, 38
Emsellem, E., Fathi, K., Wozniak, H., et al. 2006, MNRAS, 365, 367
Fromerth, M. J., & Melia, F. 2000, ApJ, 533, 172
Furenlid, I., & Furenlid, L. 1990, PASP, 102, 592
García-Lorenzo, B., Mediavilla, E., & Arribas, S. 1999, ApJ, 518, 190
Ge, J.-Q., Hu, C., Wang, J.-M., Bai, J.-M., & Zhang, S. 2012, ApJSS, 201, 31
Gerssen, J., Allington-Smith, J., Miller, B. W., Turner, J. E. H., & Walker, A. 2006, MNRAS, 365, 29
Glaspey, J. W., Eilek, J. A., Fahlman, G. G., & Auman, J. R. 1976, ApJ, 203, 335
Greene, J. E., & Ho, L. C. 2005, ApJ, 627, 721
Groves, B. A., Cecil, G., Ferruit, P., & Dopita, M. A. 2004, ApJ, 611, 786
Gunn, A. G., Hall, J. C., Lockwood, G. W., & Doyle, J. G. 1996, A&A, 305, 146
Heckman, T. M., Miley, G. K., van Breugel, W. J. M., & Butcher, H. R. 1981, ApJ, 247, 403
Ishigaki, T., Hayashi, T., Ohtani, H., et al. 2004, PASJ, 56, 723
Liu, X., Shen, Y., Strauss, M. A., & Greene, J. E. 2010, ApJ, 708, 427
Ozaki, S. 2009, PASJ, 61, 259
Pelat, D., & Alloin, D. 1980, A&A, 81, 172
Pilyugin, L. S., Zinchenko, I. A., Cedrés, B., et al. 2012, MNRAS, 419, 490
Sanchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, A&A, 538, A8
Seyfert, C. K. 1943, ApJ, 97, 28
Shen, Y., Liu, X., Greene, J. E., & Strauss, M. A. 2011, ApJ, 735, 48
Smith, K. L., Shields, G. A., Bonning, E. W., et al. 2010, ApJ, 716, 866
Storm, J., Carney, B. W., Latham, D. W., Davis, R. J., & Laird, J. B. 1992, PASP, 104, 168
Tonry, J., & Davis, M. 1979, AJ, 84, 1511
Veilleux, S. 1991, ApJSS, 75, 357
Westfall, K. B., Bershady, M. A., & Verheijen, M. A. W. 2011, ApJSS, 193, 21
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zhou, H., Wang, T., Zhang, X., Dong, X., & Li, C. 2004, ApJL, 604, L33