Spectroscopic binary processing within Gaia DPAC

Yassine Damerdji, Thierry Morel and Eric Gosset
Institut d'Astrophysique et de Géophysique, Université de Liège, Allée du 6 Août, Bât. B5c, 4000 Liège, Belgium yassine.damerdji@ulg.ac.be

Introduction

The Gaia satellite, which is due to be launched in 2013, will be equipped with the Radial Velocity Spectrometer (hereafter RVS), a medium resolution ($R \sim 11\,500$) spectrograph covering the spectral range 846–874 nm. The main objective of this instrument is to complement the data from the astrometric channel by providing the radial velocity, which is needed to fully describe the space motion of the source (in addition, the spectral properties of the brightest objects will also aid in characterizing them and in determining their physical parameters). The stars observed will transit in the RVS field of view about 50 times on the average during the mission with values ranging from 15 to 161 observations [2]. Therefore, Gaia will provide an unbiased, and unprecedented in terms of quantity, database of time series measurements of radial velocities for binary stars, which can in turn be used to dramatically improve our knowledge of this important population of our Galaxy [8].
However, the exploitation of this unique database will pose a number of challenges that are mainly related to the sheer amount of data collected and the faintness of most of the sources observed. As usable radial velocities are expected to be collected for single stars down to $V \sim 14$ mag [1], data for hundred of thousands of binary systems will have to be treated. This inevitably calls for an automatic processing without human intervention and having developed a set of procedures that have been fully tested based on extensive simulations before launch (although additional validation tests on data subsets will also take place during operations). The other difficulty lies in the fact that the precision steeply degrades with decreasing magnitude (e.g., reaching for a single transit a 1σ error bar of about 15 km s^{-1} for a G5 dwarf with V=14 [14]). The situation could still be worse than simulated due to the details of the CCD charge transfer problem. The techniques used to derive the orbital elements should therefore be robust against the presence of noise.
We are in charge of the Gaia Development Unit DU434 (within Coordination Unit CU4, itself part of the Data Processing and Analysis Consortium [DPAC]), which will derive Keplerian solutions for the candidate spectroscopic variables. This paper briefly presents the methodology and algorithms developed in this context to bypass the difficulties mentioned above and to best exploit the enormous potential of the RVS data.

1. Processing chain

The candidate spectroscopic variables are identified either by CU6/CU7 (variability detection using radial velocities obtained within CU6) or by CU8 (stellar classification identifying binary stars using photometric data), and are selected by CU4 for processing. After further preselection (see Pourbaix, these proceedings), DU434 gathers the radial velocities and RVS spectra from CU6, and determines the orbital elements (period, center of mass velocity, semi amplitude, eccentricity, longitude of periastron and time of periastron passage) of SB1 and SB2 systems (with one or two sets of radial velocity measurements, respectively). For each input time series of radial velocities and observed spectra, the following steps are performed:

- All faulty measurements are removed to ensure that the remaining data are suitable for further processing.
- Search for the orbital period using the time series of radial velocities (but also considering the photometric periodicities provided by CU7). After extensive tests using different techniques, a

combination of two methods proved the most efficient for this specific case. First, the best local maxima in the periodogram are preselected using the HMM method [5, 6]. This consists in the fit of a simple sine function to the phase diagram (with arbitrary T_0). This is applied to V_1 and ($|V_2 - V_1|$, V_1) for SB1 and SB2 systems, respectively. The best period is then selected and the approached solution derived using the fit of a simple sine function to the true anomaly ($V = f[e, T_p]$; see [15]). Eccentric solutions with small orbital periods may be considered spurious and hence rejected based on their location in the $\log P - e$ plane according to the SB9 data [10].

- Search for an approached orbital solution using the time series of radial velocities following [15] (see above). Various methods have been considered and/or are being developed and might eventually be implemented: canonical ones [9, 11], genetic algorithms coupled with periodograms computed following [15], as well as pattern recognition algorithms (see Section 3).
- Refine the orbital solution using the time series of radial velocities. The derivation of the final orbital solution is performed using a Levenberg-Marquardt minimization algorithm. Two different minimization schemes are used for low- and high-eccentricity solutions ([12] and [13] for e above and below 0.03, respectively).
- Test the significance of the found eccentricity by comparing it with a circular orbit.
- Compare the spectroscopic orbital solution with a mere temporal trend (polynomial) solution.
- Refine the spectroscopic binary solution using time series of observed spectra.
- Examine the residuals for an additional periodic signal.

2. Performance

The performance of the algorithms has been tested on artificial data internally simulated within DU434 (using the Gaia satellite scanning law to generate the RVS transit dates and thus the radial-velocity curves), but also on a large population of stars (currently a total of 10 millions) simulated by the Gaia Object Generator [7]. Among those, about 80 000 and 10 000 objects are SB1 and SB2 binary systems with a probability exceeding the threshold for a significant spectroscopic variability, respectively ($\mathscr{P} >$ 99.865 %). A number of systems are also SB3 or SB4's. In addition, the individual components may be intrinsically variable. These simulated data are expected to be representative of the RVS data that will be acquired during operations and are therefore particularly well suited for performing quality control tests. Figure 1 shows the recovery rate (with the condition that the found and real periods should be associated to the same periodogram peak) of the orbital period as a function of the spectroscopic variability probability for simulated SB1 and SB2 systems with components that are not intrinsically variable. For $\mathscr{P}$ $>$ 99.865 %, the success rate is above about 70 and 90% for SB1 and SB2 systems, respectively. These figures will be hard to improve. The difference between the found and input eccentricities for the systems for which the orbital period was successfully recovered is shown in Fig.2. This difference amounts to less than 0.1 in about 65 and 75% of the cases for SB1 and SB2 systems, respectively.

3. Current developments

The computation of the significance level of the highest peak in a Fourier periodogram is a well-known topic as long as the sampling is perfectly regular. However, in the case of an odd sampling, the distribution of the height of such a dominant peak is generally unknown. Unfortunately, we usually need to compute such a probability to estimate the likelihood of the possible presence of a periodicity in the data, and certainly to test the residuals of the orbital-solution fit for additional periodicities. On the basis of

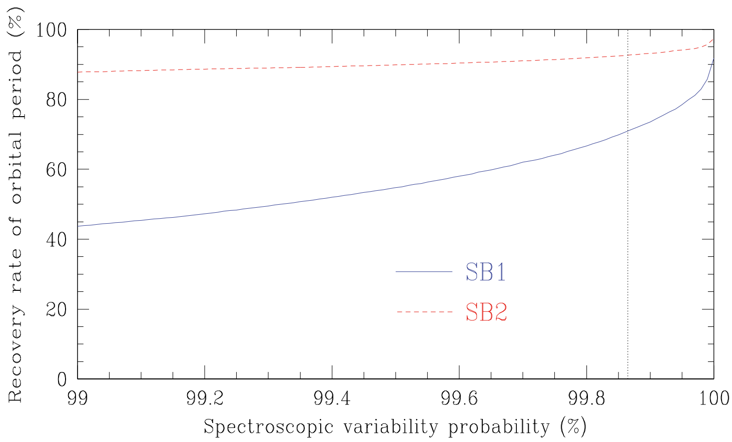

Figure 1: Recovery rate of the orbital period as a function of the probability for a spectroscopic variability for simulated SB1 (*solid line*) and SB2 (*dashed line*) systems. Only binaries with two components that are not intrinsically variable are considered. The threshold for a significant spectroscopic variability ($\mathscr{P} > 99.865$ %) is indicated by a vertical, dotted line.

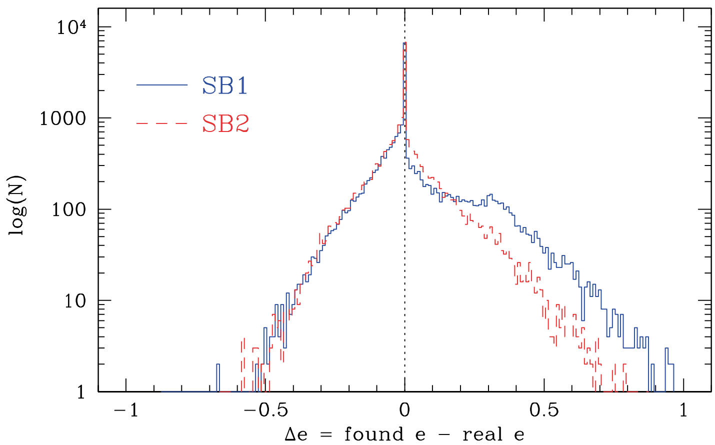

Figure 2: Difference between the found and input eccentricities for simulated SB1 (*solid line*) and SB2 (*dashed line*) systems.

numerous simulations, we investigated the distribution of the highest peak for typical Gaia RVS sampling patterns. The latter are dependent on the observed position of the object on the sky, and this has an impact on the distribution; this is clearly a fundamental problem that we try to solve at least approximately. On the other hand, we investigated the statistical behaviour of the χ^2 statistic related to the fit of an orbital solution; we noticed that, if the period is considered as a free parameter, the fit is highly non-linear and the estimation of the actual number of degrees of freedom of the χ^2 statistic turned out to be an awkward problem which is not analytically solvable. Again, careful simulations can help.

The fit of an eccentric orbital solution is also a non-linear problem. Therefore, before applying a least-square local minimization, we have to ensure that we are already close to the main minimum of the χ^2 function. Therefore, we need a good a priori guess of the parameter values prior to the minimization process. The non-linearity prevents the use of global minimization methods. We investigated several tools. One possibility is to be sure to cover the whole domain of the parameter ranges. Therefore, we investigated the use of Monte-Carlo or genetic algorithm approaches. The genetic algorithm methods included the possibility for the individuals of the populations (trial solutions) characterized by sets of genes (values of the parameters), that have been evaluated (goodness of fit) to be selected, to survive and participate to the reproduction by generating the next generation (alternative trial solutions) from their own combined genes. Spontaneous mutations were also permitted. Although leading to good results, these methods were rather time consuming and were not affordable in the Gaia context [2]. The alternative is to derive good estimates of the parameters from a simple empirical analysis. Since the eccentricity (which is the main source of non-linearity if the period is known) can be guessed from a mere look at the radial-velocity curve, we decided to apply pattern recognition methods. We tested two approaches, the first one consists in the interpolation of the radial velocity curve at regular phases (for a given tested period). Many interpolation techniques were tested: linear, Fourier series and spline interpolations. Catmull-Rom spline interpolation (a variant of Cubic Hermite spline) seems to be robust against noise. The interpolated points are the learning process inputs. Alternatively, we use the Fourier series fit coefficients as inputs. It seems that this approach performs better than the first one, since the fit coefficients contain more information than interpolated points.

For the learning process, we tested many algorithms: nearest neighbour techniques, artificial neural networks, support vector machine and regression trees [3]. The Extremely Randomized Tree algorithm [4] gives very encouraging results in addition to being very fast and easy to interpret. It is at the basis of further developments. The detailed results of the pattern recognition techniques applied to radial velocity curves will be reported in a future paper.

Acknowledgements

This research is supported in part by ESA-Belspo PRODEX funds, in particular via Gaia-DPAC contract ESA C90289.

References

[1] Cropper M., Katz D. 2011. *EAS Publications Series*, 45, 181

[2] Damerdji Y., Delchambre L., Morel T., Gosset E., Rauw G. 2011. *Bulletin de la Société Royale des Sciences de Liège*, 80, 480

[3] Delchambre L. 2010. TFE Master en Sciences Informatiques, *University of Liège*, Recherche d'une évaluation robuste de l'excentricité dans des systèmes binaires d'étoiles

[4] Geurts P., Ernst D., Wehenkel L. 2006. *Machine Learning*, 63, 3

[5] Gosset E., Royer P., Rauw G., Manfroid J., Vreux J.-M. 2001. *MNRAS*, 327, 435

[6] Heck A., Manfroid J., Mersch G. 1985. *A&AS*, 59, 63

[7] Isasi Y., Borrachero R., Martinez O., et al. 2011. The Gaia Object Generator GOG 10.0 - User Guide, GAIA-C2-UG-UB-YI-003-10 (DPAC internal document).

[8] Katz D., et al. 2011. *EAS Publications Series*, 45, 189

[9] Lehmann-Filhès R. 1894. *AN*, 136, 17

[10] Pourbaix D., Tokovinin A. A., Batten A. H., Fekel F. C., Hartkopf W. I., Levato H., Morrell N. I., Torres G., Udry S. 2004. *A&A*, 424, 727

[11] Russell H. N. 1902. *ApJ*, 15, 252

[12] Schlesinger F. 1910. *Pub. Allegheny Obs.* (Pittsburgh), 1, 33

[13] Sterne T. E. 1941. *Proc. Nat. Acad. Sc. USA*, 27, 175

[14] Viala Y., Blomme R., Frémat Y., et al. 2010. WP650 Software Test Plan and Verification Report - Single Transit Analysis (Cycle 6), GAIA-C6-SP-OPM-YV-007-1 (DPAC internal document).

[15] Zechmeister M., Kürster M. 2009. *A&A*, 496, 577