MODELING RADIAL VELOCITY SIGNALS FOR EXOPLANET SEARCH APPLICATIONS

Prabhu Babu*, Petre Stoica
Division of Systems and Control, Department of Information Technology
Uppsala University, P.O. Box 337, SE-75 105, Uppsala, Sweden

Jian Li
Department of Electrical and Computer Engineering, University of Florida, Gainesville, 32611, FL, U.S.A.

Keywords: Radial velocity method, Exoplanet search, Kepler model, IAA, Periodogram, RELAX, GLRT.

Abstract: In this paper, we introduce an estimation technique for analyzing radial velocity data commonly encountered in extrasolar planet detection. We discuss the Keplerian model for radial velocity data measurements and estimate the 3D spectrum (power vs. eccentricity, orbital period and periastron passage time) of the radial velocity data by using a relaxation maximum likelihood algorithm (RELAX). We then establish the significance of the spectral peaks by using a generalized likelihood ratio test (GLRT). Numerical experiments are carried out on a real life data set to evaluate the performance of our method.

1 INTRODUCTION

Extrasolar planet (or shortly exoplanet) detection is a fascinating and challenging area of research in the field of astrophysics. Till mid 2009, 353 exoplanets have been discovered. Some of the techniques available in the astrophysics literature to detect exoplanets are astrometry, the radial velocity method, pulsar timing, the transit method and gravitational microlensing. Among these methods, the radial velocity analysis is the most commonly used technique, in which the Doppler shift in the spectral lines and hence the radial velocity of the parent star is measured. The spectrum of the measured Doppler shifts is then analyzed to detect the exoplanet(s) revolving around the star.

Most often the radial velocity measurements are obtained at nonuniformly spaced time intervals due to hardware and practical constraints, which limits the application of commonly used spectral analysis methods. The most straightforward way to deal with this problem is to use the standard periodogram by ignoring the nonuniformity of data samples, which results in an inaccurate spectrum. In (Roberts et al., 1987), a method named CLEAN was proposed, which is based on iterative deconvolution in the frequency domain to obtain a clean spectrum from an initial dirty one. A periodogram related method is the least squares periodogram (also called the Lomb-Scargle periodogram) (Lomb, 1976; Scargle, 1982) which estimates the sinusoidal components by fitting them to the observed data. Most recently, (Yardibi et al., 2010; Stoica et al., 2009) introduced a new method called the Iterative Adaptive Approach (IAA), which relies on solving an iterative weighted least squares problem.

In this paper, we analyze the radial velocity data by using a relaxation maximum likelihood algorithm (RELAX) initialized with IAA estimates. The significance of the spectral peaks is then established via a generalized likelihood ratio test (GLRT). Numerical experiments are carried out on a real life radial velocity data set.

In Section 2, we describe the model used in this paper for the radial velocity data. Section 3 presents the RELAX and GLRT methods, and Section 4 contains the results for a real life example. Finally, the paper is concluded in Section 5.

*Corresponding author. This work was supported in part by the Swedish Research Council (VR).

Babu P., Stoica P. and Li J. (2010).
MODELING RADIAL VELOCITY SIGNALS FOR EXOPLANET SEARCH APPLICATIONS.
In *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics*, pages 155-159
DOI: 10.5220/0002770601550159

2 DATA MODEL

Let $\{y(t_n)\}_{n=1}^N$ denote the radial velocity of a star measured at a set of possibly nonuniform time instants $\{t_n\}_{n=1}^N$. Based on the Keplerian model of planetary motion (Zechmeister and Kurster, 2009) (Cumming, 2004), the radial velocity data are modeled as follows:

$$y(t_n) = C_0 + \sum_{m=1}^{M} \beta_m[\cos(\omega_m + \nu_m(t_n)) + e_m\cos(\omega_m)], \quad n = 1, \cdots, N \tag{1}$$

where C_0 is the constant radial velocity, and

$$\begin{aligned} \tan(\nu_m(t_n)) &= \sqrt{\tfrac{1+e_m}{1-e_m}}\tan(E_m(t_n)) \\ E_m(t_n) - e_m\sin(E_m(t_n)) &= \tfrac{2\pi(t_n - T_m)}{P_m}, \end{aligned} \tag{2}$$

M: Number of exoplanets revolving around the star. The number of planets M is usually unknown.

e_m: Eccentricity of the orbit of the m^{th} planet.

ω_m: Longitude of the periastron for the m^{th} planet.

P_m: Orbital period of the m^{th} planet; P_m is related to orbital frequency f_m by $f_m = \frac{1}{P_m}$.

T_m: Periastron passage time of the m^{th} planet.

β_m: Radial velocity amplitude of the m^{th} planet.

$\nu_m(t_n)$, $E_m(t_n)$: True and eccentric anomaly of the m^{th} planet, with t_n denoting their time dependence.

We divide the entire 2D space $\mathcal{G}$, defined as $\mathcal{G} = \{(e,f),\, 0 \le e < e_{\max},\, -\frac{f_{\max}}{2} < f < \frac{f_{\max}}{2}\}$, into a grid of prespecified size K. We point out here that the grid $\mathcal{G}$ does not include the parameter T, and that T is taken to be zero for the time being but will be estimated as described later on. The choices of $e_{\max}$ and $f_{\max}$ in $\mathcal{G}$ depend on the sampling pattern: to determine them, we calculate the spectral window defined as:

$$W(e,f) = \left|\frac{1}{N}\sum_{n=1}^{N}\exp(j\nu(t_n))\right|^2, \quad 0 \le e < 1, -\infty \le f \le \infty. \tag{3}$$

For any choice of (e,f) and t_n, there exists a $\nu(t_n)$ obtained via (2). Following (Eyer and Bartholdi, 1999), the parameters $e_{\max}$ and $f_{\max}$ are chosen such that the region $\mathcal{G}$ leads to an unambiguous $W(e,f)$ (see (Babu et al., 2010) for more details).

3 PARAMETER ESTIMATION AND STATISTICAL SIGNIFICANCE TESTING: RELAX AND GLRT

The estimates obtained from IAA are usually fairly accurate. However, if the grid ($\mathcal{G}$) is not chosen fine enough (to reduce the computation time), then IAA might miss some true peaks (see (Babu et al., 2010) for an elaborate discussion on IAA for radial velocity data). In that case, applying RELAX (Li and Stoica, 1996), a parametric iterative estimation algorithm, can refine the IAA estimates. Algorithm 1 briefly describes the steps involved in RELAX. The P largest peaks picked from IAA are used as initial estimates for RELAX, which has a beneficial effect on the convergence of RELAX compared with using other more arbitrary initial estimates. In the case of radial velocity data, the choice of $P = 5$ peaks appears to be reasonable for most applications. RELAX generally converges within a few iterations (typically in less than 10 iterations).

Next we note that, under the assumption the noise in the data is Gaussian distributed, the RELAX estimates are optimal in the maximum likelihood sense (Li and Stoica, 1996). We can then use the generalized likelihood ratio test (GLRT) to establish the statistical significance of the estimated planet parameters. We first apply RELAX to the largest IAA peak and use GLRT to test the null hypothesis that there are no planets (or, in other words, that the data is made only of white noise) against the hypothesis that there is at least one exoplanet. If the test rejects the null hypothesis then we will proceed and apply RELAX to the two largest peaks and subsequently test the hypothesis that there is one exoplanet in the data against the hypothesis that there are at least two exoplanets; and so on. As an example, for the following hypotheses

H_0: There are no planets.

H_1: There is at least one exoplanet with eccentricity $\hat{e}_1$, orbital frequency $\hat{f}_1$ and periastron passage time $\hat{T}_1$.

the log-likelihood (LL) functions are given by:

$$\begin{aligned} &\mathrm{LL}(H_0) = \\ &\quad -\tfrac{N}{2}\ln\left(\sum_{n=1}^{N}|y(t_n)|^2\right) + C, \\ &\mathrm{LL}(H_1) = \\ &-\tfrac{N}{2}\ln\left(\sum_{n=1}^{N}|y(t_n) - \hat{r}_1\cos(\nu(t_n)) - \hat{q}_1\sin(\nu(t_n))|^2\right) + C \end{aligned} \tag{4}$$

where C is an additive constant, $\nu(t_n)$ is calculated from the RELAX estimates ($\hat{e}_1$, $\hat{f}_1$, $\hat{T}_1$), and $\hat{r}_1, \hat{q}_1$ are the least square estimates of r, q corresponding to ($\hat{e}_1$, $\hat{f}_1$, $\hat{T}_1$), see Algorithm 1. Under the assumption that hypothesis H_0 is true, the log-likelihood-ratio, defined as $2\,(\mathrm{LL}(H_1) - \mathrm{LL}(H_0))$, is asymptotically a random variable with a chi-square distribution. Then the GLRT is given by

$$2(\mathrm{LL}(H_1) - \mathrm{LL}(H_0)) \underset{H_0}{\overset{H_1}{\gtrless}} \Lambda \tag{5}$$

where Λ denotes a fixed threshold. The threshold is usually chosen such that $\text{prob}(X \leq \Lambda) = \xi$, where $X \sim \chi^2_5$ denotes a chi-square distributed random variable with 5 degrees of freedom (because of the 5 unknowns per planet in the data model, namely e, f, T, r and q), and ξ determines the significance level of the test. Choosing $\xi = 0.99$ gives a false alarm probability of 0.01 and the corresponding threshold is $\Lambda = 15$.

4 A REAL LIFE EXAMPLE: HD 208487

In this section, we consider the application of the algorithm introduced in the previous section to a real life radial velocity data set. Our goal is to detect the exoplanets present in a star system and estimate their eccentricities, frequencies and periastron passage times. We will show the following plots:

- Amplitude vs. orbital frequency for IAA and RELAX (eccentricity and periastron passage time values for the peaks in the amplitude spectrum are indicated in the plots).
- Likelihood ratio vs. the planet number.
- Observed and fitted data sequences.

The data set used here consists of 31 samples of radial velocity measurements of the star HD 208487. The parameters $e_{\max}$ and $f_{\max}$ are determined from the spectral window to be 0.5 and 1 cycles/day. The spectrum obtained using IAA is shown in Fig.1(a), which indicates the presence of more than one planet. The 5 largest peaks in the IAA spectrum are picked up and are used to initialize RELAX. The GLRT plot shown in Fig.1(c) suggests the existence of three planets in the HD 208487 star system with the following parameters (see Fig.1(d) and also Table 1): ($e_1 = 0.326$, $f_1 = 0.0078$ cycles/day, $T_1 = 130.9$ days), ($e_2 = 0.315$, $f_2 = 0.069$ cycles/day, $T_2 = 14.2$ days) and ($e_3 = 0$, $f_3 = 0.0408$ cycles/day, $T_3 = 2.9$ days). However (Tinney et al., 2005) reported that the star has only one planet with an orbital frequency of 0.0077 cycles/day. Fig. 1(e) and 1(f) show the plots of measured data and the fitted data obtained assuming the existence of one and, respectively, three planets. It is seen clearly from these figures that the three planet model fits the measured data much better than a single planet model.

5 CONCLUSIONS

The real life example discussed in the paper suggests that our algorithm successfully detects the presence of spectral peaks (planets) in radial velocity data and accurately identifies both their frequencies and eccentricities as well as their periastron passage times. The example used here is typical of cases usually encountered in exoplanet search and hence the proposed algorithm is believed to be an effective and useful tool.

REFERENCES

Babu, P., Stoica, P., Li, J., Chen, Z., and Ge, J. (2010). Analysis of radial velocity data by a novel adaptive approach. *The Astronomical Journal*, 139:783–793.

Cumming, A. (2004). Detectability of extrasolar planets in radial velocity surveys. *Monthly Notices of the Royal Astronomical Society*, 354(4):1165–1176.

Eyer, L. and Bartholdi, P. (1999). Variable stars: Which Nyquist frequency? *Astronomy and Astrophysics, Supplement Series*, 135:1–3.

Li, J. and Stoica, P. (1996). Efficient mixed-spectrum estimation with applications to target feature extraction. *IEEE Transactions on Signal Processing*, 44(2):281–295.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39(1):10–33.

Roberts, D. H., Lehar, J., and Dreher, J. W. (1987). Time series analysis with CLEAN. I. Derivation of a spectrum. *The Astronomical Journal*, 93(4):968–989.

Scargle, J. D. (1982). Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal*, 263:835–853.

Stoica, P., Li, J., and He, H. (2009). Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram. *IEEE Transactions on Signal Processing*, 57(3):843–858.

Stoica, P. and Moses, R. (2005). *Spectral Analysis of Signals*. Prentice Hall, Upper Saddle River, N.J.

Tinney, C., Butler, R., Marcy, G., Jones, H., Penny, A., McCarthy, C., Carter, B., and Fischer, D. (2005). Three Low-Mass Planets from the Anglo-Australian Planet Search 1. *The Astrophysical Journal*, 623(2):1171–1179.

Yardibi, T., Li, J., Stoica, P., Xue, M., and Baggeroer, A. B. (2010). Iterative adaptive approach for sparse signal representation with sensing applications. *IEEE Transactions on Aerospace and Electronic Systems*, 46:425–443.

Zechmeister, M. and Kurster, M. (2009). The generalised Lomb-Scargle periodogram. *Astronomy and Astrophysics*, 496(2):577–584.

APENDIX

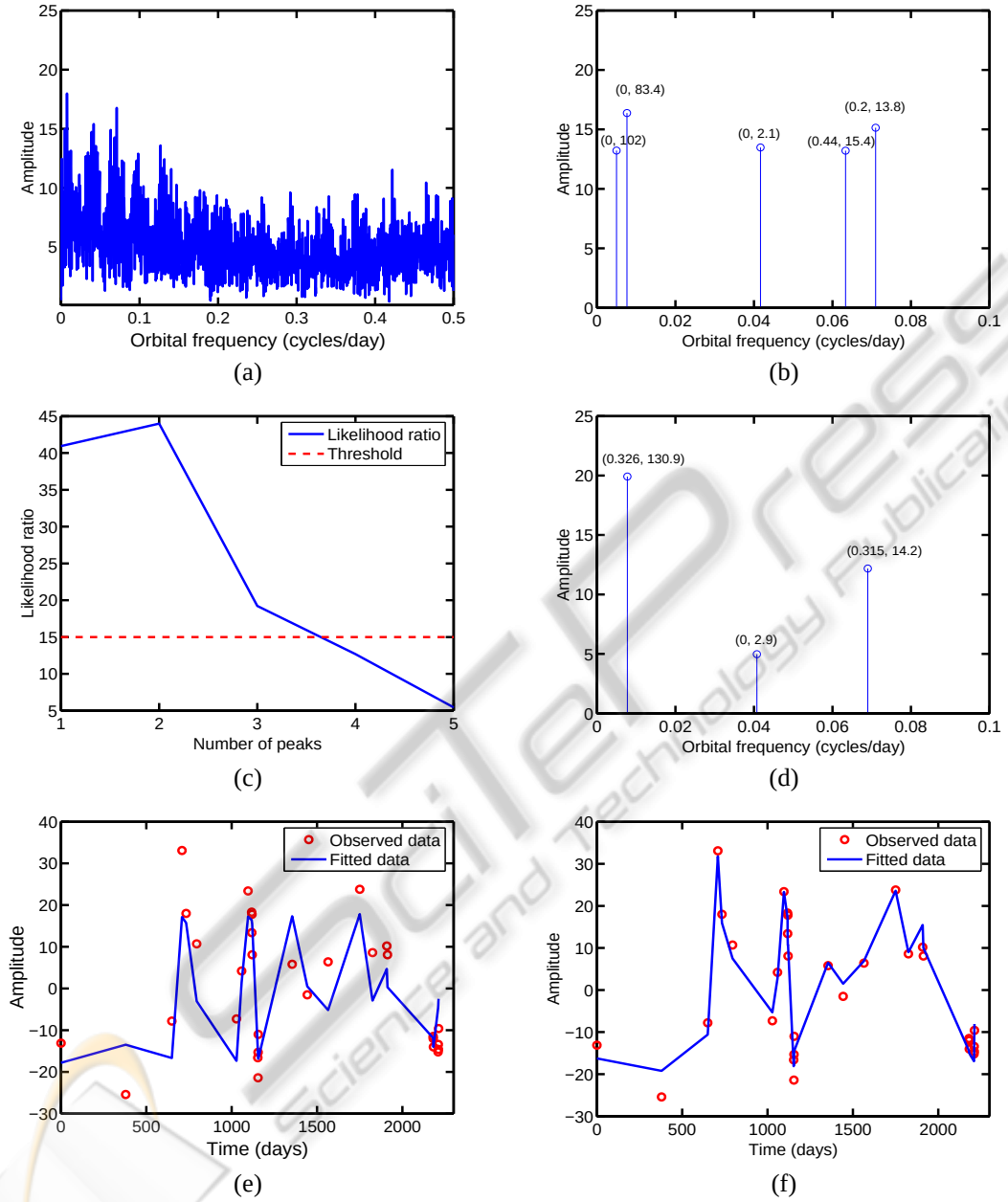

Figure 1: HD 208487: (a) the IAA spectrum, (b) the 5 largest peaks in the IAA spectrum, (c) the likelihood ratio, (d) the RELAX spectrum, (e) comparison of the observed data and fitted data obtained from the parameters of the planet reported in (Tinney et al., 2005) and (f) comparison of the observed data and fitted data obtained from the three detected planets.

Table 1: Parameters of the planets of HD 208487 star system. *) The third planet becomes statistically insignificant if the false alarm probability is decreased from 10^{-2} to 10^{-4}, in which case the threshold becomes $\Lambda = 25$.

Planet No.	**Previous work**				**This work**			
	e	f(cycles/day)	T(days)	β	e	f(cycles/day)	T(days)	β
1	0.24	0.0077	92	20	0.3260	0.0078	130.9	19.9
2	-	-	-	-	0.3150	0.0690	14.2	12.18
3*)	-	-	-	-	0	0.0408	2.9	4.96

Algorithm 1: RELAX.

Let $\{(e_p^0, f_p^0, T_p^0)\}_{p=1}^P$ denote the parameters of the P most dominant planets in the IAA spectrum (e.g. $P = 5$). The initial values of their radial velocity amplitudes $\{(r_p^0, q_p^0)\}_{p=1}^P$ are taken to be zero.

for $p = 1$ to P **do**
  for $i = 1$ to I (e.g. $I = 10$) **do**
    for $u = 1$ to p **do**
      $y_u^i(t_n) = y(t_n) -$
      $\sum_{k=1,k\neq u}^{p} \left(r_k^{i-1} \cos(\nu_k^{i-1}(t_n)) + q_k^{i-1} \sin(\nu_k^{i-1}(t_n)) \right)$
      where ν_k^{i-1} is obtained from
      $(e_k^{i-1}, f_k^{i-1}, T_k^{i-1})$. Then
      (e_u^i, f_u^i, T_u^i)
      $= \arg\min_{\{e,f,T\}} \min_{\{r,q\}} \sum_{n=1}^{N} \left| y_u^i(t_n) - r\cos(\nu(t_n)) - q\sin(\nu(t_n)) \right|^2$
      The inner minimization with respect to $\{r, q\}$ in the above equation is a least squares problem and the estimates $\{r_u^i, q_u^i\}$ can be obtained analytically (see (Stoica and Moses, 2005)). The minimization with respect to $\{e, f, T\}$ is carried out via a 3D grid search performed around $(e_u^{i-1}, f_u^{i-1}, T_u^{i-1})$.
    end for
  end for
  for $u = 1$ to p **do**
    $(e_u^0, f_u^0, T_u^0) \leftarrow (e_u^I, f_u^I, T_u^I)$ and $(r_u^0, q_u^0) \leftarrow (r_u^I, q_u^I)$.
  end for
end for
$\{(\hat{e}_p, \hat{f}_p, \hat{T}_p) \leftarrow (e_p^I, f_p^I, T_p^I)\}_{p=1}^P$ and $\{(\hat{r}_p, \hat{q}_p) \leftarrow (r_p^I, q_p^I)\}_{p=1}^P$