Fitting Rectangular Signals to Time Series Data by Metaheuristic Algorithms

Andreas M. Chwatal and Günther R. Raidl

Vienna University of Technology, Vienna, Austria
{chwatal|raidl}@ads.tuwien.ac.at

Abstract. In this work we consider the application of two metaheuristics, namely evolution strategies and scatter search, to the problem of fitting rectangular signals to time-data series. The application background is to search for exoplanet-transit signals in stellar photometric observation data.

1 Introduction

Fitting parameterized models to data series is a frequently performed task in scientific computing. Nevertheless, finding (near-)optimal fits of superposed periodical signals to time-series data becomes a non-trivial problem when non-sinusoidal models are considered. In this case it is not always possible to derive good model parameters from the Fourier spectrum. Noisy data may further complicate this task. Finding good fits, which is in fact a continuous parameter optimization problem, is a computationally challenging task under these circumstances. In this work, we consider the problem of fitting rectangular signals to time-data series, and present metaheuristic algorithms to solve the problem.

2 Problem Description

Our particular application background comes from the field of astronomy, in particular the problem of finding signals from exoplanet-transits in stellar photometric light-curves. For a comprehensive overview on exoplanets and detection methods see [1]. A transiting planet periodically shadows some of the light from its host star for a short time when it moves into our line of sight to the star. During the transit the luminosity of the star is marginally reduced. By neglecting the in- and egress phases, the transit-lightcurve can be described by a periodic rectangular signal. The corresponding parameters are the period P the transit occurs with, a phase offset τ, the length l of the transit, and finally the transit depth d. The latter parameter corresponds to the percentage of light from the star beeing shadowed by the transiting planet. Figure 1 depicts the situation for a single planet. Assuming M planets, the signal of the model at time t is given

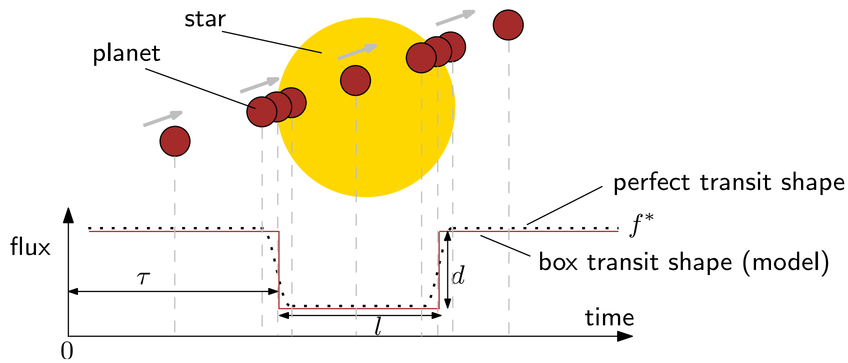

Fig. 1. Transiting planet and corresponding lightcurve

by

$$\phi(t) = f^* - \sum_{j=1}^{M} \delta_j(t), \tag{1}$$

where f^* denotes a further parameter describing the regular flux (luminosity) of the host star; $\delta_j(t)$ denotes the reduction of f^* resulting from planet j and is given by

$$\delta_j(t) = \begin{cases} d_j & \text{if } \tau_j < t \bmod P_j \leq \tau_j + l_j \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The observed data series is given by a list of $\{(t_i, f_i)\}, 1 \leq i \leq N$, where t_i denotes a particular observation time and f_i the observed flux (i.e. luminosity) at that given time. The standard deviation of the input values will be denoted by σ_f. Let further $m_j = (P_j, l_j, d_j, \tau_j)$ and hence $\boldsymbol{m}$ the vector of all model parameters (except f^*). The overall quality of the fit can be characterized by

$$f(\boldsymbol{m}, f^*, \boldsymbol{t}, \boldsymbol{f}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\phi(t_i) - f_i)^2}. \tag{3}$$

The overall objective is to find a parameter setup for $\boldsymbol{m}$ and f^* minimizing Eq. (3), i.e. to find a model with minimal deviation from the observations.

Due to stellar fluctuations and measurement errors real-world instances contain noisy signals. The signal-to-noise ratios can be expected to be very low, i.e. the respective values of d_j will be in the same magnitude as σ_f.

3 Previous Work

Several applications of genetic algorithms in astronomy are outlined in [2], and have since then been successfully applied for many purposes. In particular for the detection of exoplanets, evolutionary algorithms have been used with some success. For instance, an evolution strategy for fitting Keplerian models to radial velocity data is described in [3].

The development of efficient transit detection algorithms has recently gained more interest in the scientific community, as space-based missions like CoRoT[1] provide a great amount of observational data. One of the most popular approaches is the *box fitting least-square algorithm* [4]. This approach, as well as *phase dispersion minimization* [5] have the main drawback, that they are only directly applicable for finding single planet transits.

So far no multi-planet system could be discovered by the transit method, which is likely due to the difficulty of detecting their signals in (existing) observational data. More efficient techniques to tackle this continuous parameter optimization problem will thus be a valuable contribution to exoplanet research.

4 Fitness-Landscape Analysis

In order to evaluate the applicability of metaheuristics for solving this problem, we performed a comprehensive fitness-landscape analysis. For this purpose we created numerous test instances containing signals from up to five planets. For each configuration we created multiple instances with different signal-to-noise ratios. The results confirm the expected low fitness-distance correlation. Strong mutual dependencies of the parameters make the problem a difficult one. Nevertheless, there are some structures in the solution landscape of the objective function that indicate the chance to develop efficient heuristic algorithms.

5 Metaheuristic Algorithms

There exists a variety of algorithms for heuristically solving difficult continuous parameter optimization problems, amongst which *evolution strategies* (ES) and *scatter search* (SS) are very popular [6, 7]. Both algorithms are population-based approaches, iteratively modifying and evaluating a set of candidate solutions. In contrast to the frequently used binary encoding in genetic algorithms, we directly encode the parameters as real values in both approaches. We do not need to use the model parameters d_j (for each planet) and f^* as optimization parameters, as their optimal values w.r.t. the other model parameters can efficiently be derived.

The ES can be classified as a (μ, λ)-ES with self-adaptation of strategy parameters [8] where μ denotes the size of the population and λ the number of offsprings created in each generation. Mutation is considered to be the primary operator and selection is traditionally done in a deterministic way by choosing the μ best offsprings. Mutation is performed by adding Gaussian random values to the parameters x_k (see Eq. (4)), where the standard deviation is given by a strategy parameter σ_k, associated with each parameter.

$$x'_k = x_k + N_k(0, \sigma'_k) \tag{4}$$

[1] CoRoT: **Co**nvection **Ro**tation und planetary **T**ransits; European space telescope

These strategy parameters are also modified by the evolutionary operators, which facilitates self-adaption of the search process.

Contrary to evolutionary algorithms, SS combines solutions in a more deterministic way based on various decision and combination rules [7]. The modification operators work on a pool of candidate solutions, called the *reference set*. It is ensured that high-quality solutions as well as diverse solutions are maintained in the reference set in order to be able to primarily search in promising regions of the search space, but not to get stuck to local optima.

In both algorithms we optionally use additional local optimization methods including gradient descent and the Nelder-Mead (downhill simplex) method [9].

6 Results

For an extensive evaluation of our algorithms we created artificial test-instances. For this purpose we used five configurations for a particular number of signals, ranging from zero to five. An important aspect is, that a real stellar signal typically does not only contain the rough rectangular signal from the transiting planet, but also portions of stellar jitter and measurement errors. We take this into account by adding Gaussian random variables to each data point in the artifical signal. We thus create three instances for each configuration: one strictly rectangular signal and two further noisy ones with different standard deviations.

In our computational experiments we evaluate the strengths and drawbacks of each algorithm. Our experiments indicate, that the presented heuristic methods are fast and powerful tools to search for multi-planet transit candidates in large data sets. More details will be presented in the full paper.

References

1. Deeg, H., Belmonte, J.A., Aparicio, A., eds.: Extrasolar Planets. Cambridge University Press (2008)
2. Charbonneau, P.: Genetic Algorithms in Astronomy and Astrophysics. Astrophysical Journal Supplement **101** (1995) 309–334
3. Chwatal, A.M., Raidl, G.R.: Determining orbital elements of extrasolar planets by evolution strategies. In Moreno-Díaz, R., et al., eds.: Computer Aided Systems Theory – EUROCAST 2007. Volume LNCS 4739 of LNCS. (2007) 870–877
4. Kovács, G., Zucker, S., Mazeh, T.: A box-fitting algorithm in the search for periodic transits. Astronomy and Astrophysics **391** (2002) 369–377
5. Stellingwerf, R.F.: Period determination using phase dispersion minimization. Astrophysical Journal **224** (1978) 953–960
6. Bäck, T.: Evolutionary Algorithms in Theory and Practice. Oxford University Press, New York (1996)
7. Glover, F., Laguna, M., Marti, R.: Fundamentals of scatter search and path relinking. Control and Cybernetics **29**(3) (2000) 653–684
8. Schwefel, H.P.: Numerical Optimization of Computer Models. Wiley, Chichester (1981)
9. Nelder, J., Mead, R.: A simplex method for function minimization. The Computer Journal (7) (1964) 308–313