Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

AGCD: a robust periodicity analysis method based on approximate greatest common divisor*

Juan YU‡, Pei-zhong LU

(*School of Computer Science and Technology, Fudan University, Shanghai 200433, China*)

E-mail: juanyu10@fudan.edu.cn; pzlu@fudan.edu.cn

Received Oct. 9, 2014; Revision accepted Mar. 17, 2015; Crosschecked May 15, 2015

Abstract: Periodicity is one of the most common phenomena in the physical world. The problem of periodicity analysis (or period detection) is a research topic in several areas, such as signal processing and data mining. However, period detection is a very challenging problem, due to the sparsity and noisiness of observational datasets of periodic events. This paper focuses on the problem of period detection from sparse and noisy observational datasets. To solve the problem, a novel method based on the approximate greatest common divisor (AGCD) is proposed. The proposed method is robust to sparseness and noise, and is efficient. Moreover, unlike most existing methods, it does not need prior knowledge of the rough range of the period. To evaluate the accuracy and efficiency of the proposed method, comprehensive experiments on synthetic data are conducted. Experimental results show that our method can yield highly accurate results with small datasets, is more robust to sparseness and noise, and is less sensitive to the magnitude of period than compared methods.

Key words: Periodicity analysis, Period detection, Sparsity, Noise, Approximate greatest common divisor (AGCD)
doi:10.1631/FITEE.1400345 **Document code:** A **CLC number:** TP301.6

1 Introduction

Periodicity is one of the most common phenomena in the physical world. Many natural phenomena, such as tidal patterns, sunspots, temperature changes, manifest strong or weak periodicity. Animals migrate annually, and people commute between their working places and home daily. Period and periodic patterns are important features, which can be used to detect anomalies, and predict the trend of periodic events. Thus, the problem of periodicity analysis is of research interest, such as bit synchronization in communications, pulse repetition interval (PRI) analysis of radar, hop rate estimation of frequency-hopping spread spectrum (FHSS) signals, period detection of light curves of variable stars in astronomy (Huijse *et al.*, 2011; 2012), periodic behavior analysis of animals or humans in data mining (Li *et al.*, 2010; 2012), periodicity analysis of DNA sequences or gene expression data in bioinformatics (Junier *et al.*, 2010).

Period detection is a challenging problem in practical application. First, the acquired datasets or sample sets of periodic events are often incomplete, due to the limitations of data collection. In other words, datasets used for periodicity analysis are usually sparse and unevenly sampled. For example, receivers for passive radar surveillance may not be able to observe each consecutive pulse emitted by the radar from the third party. Second, datasets of periodic events are often noisy, due to the intrinsic oscillation of periodic processes, and extrinsic factors such as environmental noise or poor time resolution.

The period detection problem has been investigated in signal processing for decades. Fourier

‡ Corresponding author
* Project supported by the National Natural Science Foundation of China (No. 60673082)
ORCID: Juan YU, http://orcid.org/0000-0001-9562-1325

transform and autocorrelations are typical methods for analyzing the periodicity of signals. However, these traditional methods are sensitive to noise and sparsity. These methods are based on a typical assumption that datasets are abundant and evenly sampled in time. Thus, these traditional methods cannot be directly applied to sparse and noisy datasets.

To cope with sparse and noisy datasets, several period detection methods have been proposed. Fogel and Gavish (1988) considered the problem of period estimation from noisy datasets. They modeled the observations as a periodic point process, and proposed a periodogram-based method to discover the unknown period. They returned the value maximizing the periodogram as the estimated period. Gray *et al.* (1994) devised a maximum likelihood period estimator for the independent and identically distributed (i.i.d.) Gaussian noise case. But they did not consider missing measurements or outliers. Sidiropoulos *et al.* (2005) formulated the period detection problem as a maximum likelihood estimation problem, and proposed a family of estimators called separable least squares line search (SLS2) estimators. Following Sidiropoulos *et al.*'s work, Clarkson and McKilliam showed that the maximum likelihood (ML) estimator for the sparse and noisy period estimation problem can be understood as a lattice problem, and proposed the lattice line search (LLS) estimators to solve the problem (Clarkson, 2008; McKilliam and Clarkson, 2008).

The above mentioned methods operate by sampling an objective function over an interval derived from the range of the true period. That is to say, these methods require the prior knowledge about the rough range of the period. Thus, they are not suitable for period detection applications in which the prior knowledge is impossible to acquire. In addition, the determination of the required sampling interval is also a challenging problem. If the chosen sampling interval is too small, then the optimized period value will never be selected. Otherwise, the computational cost will be high.

Unlike above mentioned methods that need prior knowledge about the range of the period, Casey and Sandler presented a family of methods based on the modified Euclidean algorithm (MEA) (Casey and Sadler, 1996; Sadler and Casey, 1998). The MEA methods are motivated by the fact that the period is an approximate greatest common difference of differences between consecutive observations, and directly calculate the period from datasets using the MEA. The time complexity of these methods outperforms that of most of existing methods, which is $O(N\log N)$ in the best case, where N is the number of observations. Although the MEA methods are efficient, they are sensitive to noise and sparsity.

Most recently, Li *et al.* (2012) proposed a new method to detect periods from sparse and noisy datasets. This method exhaustively tries all possible periods combined with a process of segmenting and overlapping, and finally chooses the trial period maximizing the periodicity score as the result period. The datasets they considered are binary sequences, which denote the periodic events happening or not at the corresponding relative timestamps. Compared with other existing methods, the method they proposed makes use of both positive observations (events that happened) and negative observations (events that did not happen) in the course of period estimation, which makes it more accurate and more robust to noise and sparseness than other methods. However, the high accuracy of this method depends on large datasets. Besides, both the accuracy and efficiency of this method are very sensitive to the magnitude of the true period. Time complexity of the method in Li *et al.* (2012) is $O(n^2)$, where n is the length of the given binary sequence. It is obvious that n is closely related to the true period p, and we have $n \approx N/(1-\alpha) \cdot p$, where N is the number of observed observations, and α is the proportion of missed observations. Therefore, its time complexity can be represented as $O((N/(1-\alpha))^2 \cdot p^2)$.

In this paper, we propose a novel period detection method to effectively solve the problem of period detection from sparse and noisy datasets. Our method is motivated by the fact that the true period is actually the approximate greatest common divisor (AGCD) of the given observational dataset. The main idea of the proposed method is that it calculates and counts the occurrence times of all possible AGCDs by exhaustively searching noise space, and then selects the AGCDs with the highest occurrence times as the estimate. That is to say, our method is to find the most frequently occurring AGCD. Both our method and the MEA-based methods consider the period detection problem as finding the greatest common factor of given observations. The difference is that the MEA-based methods calculate the period through repeated subtractions, whereas our method is based on the noise exhaustive search procedure.

Compared with existing period estimation methods, our proposed method has the following advantages. First, our method does not require the prior knowledge of the rough range of the hidden period, which makes it more general in applications than most existing methods. Second, it is more robust to sparseness and yields higher accuracy than other methods under the same circumstances. Third, it can achieve high accuracy with smaller datasets than what other methods need. According to the theorem introduced in Casey and Sadler (1996), given N randomly chosen positive integers $\{k_1, k_2, ..., k_N\}$, and $P\{\gcd(k_1, k_2, ..., k_N) = 1\} = [\zeta(N)]^{-1}$, where $\zeta(N)$ is the Riemann Zeta function, $P\{\gcd(k_1, k_2, ..., k_N) = 1\}$ converges to 1 very quickly as N increases. Fourth, the performance of our method is less sensitive to the magnitude of period p than the other methods, which makes it applicable to various periodic events. Finally, the efficiency of our method outperforms those of most existing methods. The high efficiency of our proposed method results from the quick convergence of the AGCD algorithm, as the search space of the AGCD algorithm is the noise space rather than the period space and the noise space is generally much smaller than the period space.

2 Data model and approximate greatest common divisor

2.1 Data model

The observations on a periodic event or behavior can be modeled as a chronological sequence of timestamps, i.e., $T = \{t_i\}_{i=1}^N$, where each element t_i records the time at which the periodic event was observed. Following the definition of Fogel and Gavish (1988) and Casey and Sadler (1996), each element t_i in T is defined as

$$t_i = k_i \times p + \varphi + r_i, \tag{1}$$

where p is the period ($p > 0$), φ is a uniformly distributed phase ($0 \leq \varphi < p$), $\{k_i\}_{i=1}^N$ are indices representing which occurrences have been observed, and $\{r_i\}_{i=1}^N$ are zero-mean i.i.d. error terms.

Without loss of generality, we assume that all the samples and parameters are integers. For the noise terms, we assume that they have i.i.d. symmetric probability density distributions, such as uniform distribution $\mathcal{U}[-r, r]$ and white Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Furthermore, we assume that $|r_i| < p/4$ for all i. This assumption makes sure that the difference of two error terms is less than $p/2$.

For missing observations, they are often modeled through the distribution of the $\{k_i\}_{i=1}^N$. The Bernoulli process $B(\lambda)$ is a commonly used model, determining whether a measurement is missing or not, i.e., $P(k_{i+1} = k_i + 1) = 1 - \lambda$, where λ is the probability that the $(i+1)$-th observation was not observed. In addition, missing observations can be modeled by taking the jumps in the k_i as uniformly distributed on the discrete interval $[1, M]$, where M is a positive integer. That is to say, $k_{i+1} = k_i + x_i, i = 1, 2, \ldots, N-1$, where $x_i \in [1, M]$ is a uniformly distributed random integer variable.

In the data model, the period p, indices $\{k_i\}_{i=1}^N$, phase φ, and noise terms $\{r_i\}_{i=1}^N$ are all unknown parameters. The problem we consider is to estimate the latent period p from the given sparse and noisy observations. In other words, given a sequence of observations $T = \{t_i\}_{i=1}^N$ satisfying Eq. (1), which is in ascending order as each observation $\{t_i\}_{i=1}^N$ is a relative timestamp, we concentrate on the estimation of period p. This is due to the fact that, once p has been estimated, the estimation of other parameters is straightforward. Note that although our method focuses on integer period estimation, it can be easily extended to real number period scenarios. We can scale up the real number observations by using finer time granularity, thereby turning into an integer period estimation problem.

2.2 Approximate greatest common divisor

The problem of AGCD was first proposed by Howgrave-Graham (2001), which refers to recovering an unknown integer p from given two near multiples x_1 and x_2 of an unknown integer p, where $x_1 = q_1p + r_1$ and $x_2 = q_2p + r_2$, r_1 and r_2 are unknown error terms, and q_1 and q_2 are unknown positive integers smaller than x_1 and x_2 respectively (q_1 and q_2 are used to factorize x_1 and x_2 with p, and we are not interested in them in this study). It has been proved that p can be accurately recovered from x_1 and x_2 when the error terms $|r_i| < \sqrt{p}$ $(i = 1, 2)$.

The typical solution to the AGCD problem for two near multiples is the GCD exhaustive search.

The algorithm tries each noise (r_1, r_2) in the error space and checks whether the greatest common divisor $\gcd((x_1 - r_1), (x_2 - r_2))$ is large enough. The whole procedure requires r^2 Euclidean algorithm calculations, where r is the noise threshold and $r < \sqrt{p}$. Thus, the complexity is $O\left(r^2 \log t\right)$, where $\log t$ is the time complexity of the Euclidean algorithm. We can further measure the time complexity of the AGCD algorithm in terms of period p. According to the definitions, t is a near multiple of period p, and $r < \sqrt{p}$. Finally, the complexity of the algorithm is $O\left(p \log p\right)$ from the perspective of period p.

3 Period detection using the approximate greatest common divisor algorithm

The problem we consider is to estimate the latent period p of a periodic event from its observational dataset, which is not fully equivalent to the approximate integer common divisor problem introduced in Section 2.2. One significant difference is that the period detection problem does not make such assumptions that k_i and p are very large integers and approximate to $\sqrt{t_i}$. That is to say, we do not have the prior knowledge about the range of the unknown period p to help us determine the right answer among a lot of common divisors resulting from the GCD exhaustive search procedure. Without prior knowledge about the range of latent period p, it is hard for us to correctly estimate p with only two observations.

Fortunately, we have a large set of observations available, not just two, and those extra observations can also give information about the latent period p. We execute the AGCD exhaustive search for each pair of observations, and record the number of occurrences of each common divisor. Finally, we find that the number of occurrences will peak at the real period p or multiples of the real period p, i.e., $2p, 3p, \cdots$. In other words, the true period p and its multiples appear much more frequently than other non-period values. Fig. 1 gives an example about the occurrence times of part of the common divisors resulting from one of the experiments in Section 4, where the true period of the testing data is 100. From Fig. 1 we can see that the period value and its multiples appear more frequently among all the estimated values.

This observation motivates us to propose a novel period detection method based on the AGCD algorithm, which executes the AGCD algorithm on all pairs of observations and records the occurrence of each common divisor, and then estimates the divisor with the highest occurrence as the period. The proposed method is presented in detail in the following subsections.

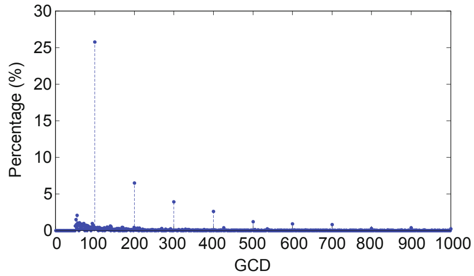

Fig. 1 Frequency of candidate greatest common divisors (GCDs) ($p = 100$)

3.1 The approximate greatest common divisor algorithm

The core algorithm of the proposed period detection method is depicted in detail in Algorithm 1. It consists of three steps.

The first step is to adaptively obtain the noise threshold. There are two common ways for determining the threshold. One way is to set the threshold according to some prior knowledge. In that case, the first step can actually be omitted. The other way is to adaptively choose the threshold according to the given sample set. In this paper, the data-adaptive method is adopted.

The second step simplifies the problem by eliminating the unknown parameter φ. There are several methods to eliminate φ, such as taking pairwise differences and taking adjacent-sample differences. We tried these eliminating methods, and found that the method used in this paper is most suitable for our AGCD based method.

The third step executes the AGCD exhaustive search procedure over the dataset D to detect the hidden period. Specifically, we calculate AGCDs for every pair of observations in D, and determine the most frequently occurring AGCD as the result period.

Algorithm 1 AGCD period estimation

Require: a set of observations $T = \{t_1, t_2, \dots, t_N\}$.
Ensure: period p.

1: Set noisy threshold r:
  (1) Take adjacent-element differences, given by $t'_i = t_{i+1} - t_i$;
  (2) $t = \min\{t'_i\}_{i=1}^{N-1}$;
  (3) Set noisy threshold $r = \sqrt{t}$;
2: Eliminate phase φ, and form a new set $D = \{d_1, d_2, \cdots, d_{N-1}\}$, where $d_j = t_j - t_1$, $1 \le j \le N-1$
3: Estimate period according to frequencies of occurrence of all possible GCDs:
  (1) $G = \varnothing$;
  (2) Compute the pairwise AGCDs among $\{d_1, d_2, \cdots, d_{N-1}\}$
    for all possible $r_i \in (-r, r)$ and $r_j \in (-r, r)$
      $g = \gcd(d_i + r_i, d_j + r_j)$;
      if $g \ge t$
        if $g \in G$
          occur[g] = occur[g] + 1;
        else
          occur[g] = 1;
          $G = G \cup g$;
        end if
      end if
    end for
  (3) Find the GCD g with the highest frequency as the period, i.e., $p = g$.

3.2 Problem simplification

Since we focus on the estimation of period p, Eq. (1) can be further simplified by eliminating the unknown parameter φ. To this end, we let each sample subtract the minimal sample of T, i.e., $t_i - t_1$, and obtain a new sample set $D = \{d_i\}_{i=1}^{N-1}$, where

$$d_i = k'_i \times p + r'_i, \tag{2}$$

$k'_i = k_{i+1} - k_1$, and $r'_i = r_{i+1} - r_1$. According to the assumptions on the original observations, it can be easily derived that samples in D are also in ascending order, and that the error terms r'_i also have i.i.d. symmetric probability density distributions.

Now, the problem becomes estimating period p from sample set D. When noise terms r'_i are zero, period p will be the GCD of all the samples in D with a high probability. Casey and Sadler (1996) have proved that $\gcd(k'_1 p, k'_2 p, \cdots, k'_{N-1} p) \to p$ which converges quickly as $N \to \infty$. In the noise case, the simplified problem can be considered as a general version of the approximate common divisor problem in Howgrave-Graham (2001), and we can name it an AGCD problem for multiple samples.

3.3 Computational complexity

In our AGCD period detection algorithm, the first step requires $O(N)$ calculations to identify the noise threshold r, the second step takes $O(N)$ calculations to eliminate the unknown phase φ, and the third step invokes the AGCD algorithm approximately $N(N-1)/2$ times to calculate the period in the worst case, where N is the number of observations. Of the three steps, the third one is the core and the most time-consuming. Finally, the whole complexity of the proposed AGCD algorithm is approximately $O\left(N^2 p \log p\right)$.

In practice, the complexity of our proposed method could be further refined through the following tricks. First, the execution time of the AGCD algorithm of the third step can be further reduced by grouping the samples into several clusters according to a given threshold on the gaps between two consecutive samples, and then choosing samples from different clusters to calculate their common divisors. Therefore, the required number of AGCD calculations could be remarkably reduced from N^2 to N. Therefore, the complexity of our method could be reduced to $O\left(N \cdot p \log p\right)$.

Furthermore, the dataset size N can be considered as a constant, and the time complexity can be further reduced to $O(p \log p)$. One reason is that the proposed algorithm could detect the accurate period with only a small set of observations, such as $N = 30$. According to the theorem introduced in Casey and Sadler (1996), given N randomly chosen positive integers $\{k_1, k_2, ..., k_N\}$, $P\{\gcd(k_1, k_2, \dots, k_N) = 1\} = [\zeta(N)]^{-1}$, where $\zeta(N)$ is the Riemann Zeta function, and thus $[\zeta(N)]^{-1}$ converges to 1 from below very quickly as N increases. Thus, there is no need to use large or infinite datasets to detect the period. Another reason is that the observational datasets that we can collect are seldom very large under many circumstances, let alone approaching infinite.

4 Evaluation

4.1 Experimental settings

In this section, we compare our method with the single iteration MEA method (Casey and Sadler,

1996) and the exhaustive trial of period method proposed in Li *et al.* (2012). We name these methods AGCD, MEA, and ETP, respectively. These algorithms were implemented in C++. All the experiments were conducted on an Intel Core i5 3.0 GHz machine with 4 GB RAM. As we could not find real datasets to evaluate our method, we designed a data generator according to the data model given in Section 2. The designed generator can generate different synthetic datasets through the settings of period p, sample size N, noise threshold r, and sparsity of observational datasets λ (i.e., the probability of a missing observation). In our generator, the default values of these parameters were set as follows: $p = 100$, $N = 40$, $r = 10$, and $\lambda = 0.8$.

We compare the performance of the three methods in terms of accuracy and efficiency. For each experiment, we report the results with one of these four parameters varying while the others are fixed. For each setting of parameters, we independently ran the experiment 100 times. The accuracy of the methods is quantified as the success rate and the mean error. The success rate is the percentage of accurately estimated experiments over the 100 trials. The mean error is the average of the estimated errors over 100 trials, i.e., $\text{ME} = \frac{1}{n}\sum_{i=1}^{N}|\hat{p}_i - p|$.

4.2 Accuracy

In this set of experiments, we examine the accuracy performance of the three methods under various parameter settings.

4.2.1 Accuracy with regard to the noise threshold

Figs. 2a and 2b depict the accuracy performance curves of the three methods with regard to different noise threshold r. The accuracy of the AGCD method is remarkably higher than the other methods when the noise threshold $r \le \sqrt{p}$. The AGCD method is less accurate than the ETP method in some cases when the size of dataset is $N = 80$. The reason is that the ETP method takes advantage of both positive and negative information during the period detection process, whereas our method uses only positive information. However, the high accuracy of ETP relies on large datasets, whereas our method needs fewer observations and can achieve similar accuracy with smaller datasets than the ETP method.

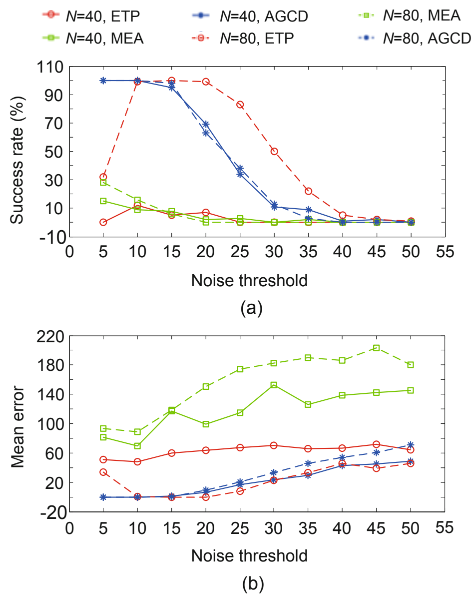

Fig. 2 Success rate (a) and mean error (b) with regard to noise threshold r ($p = 100$, $\lambda = 0.8$)

4.2.2 Accuracy with regard to sparseness

Figs. 3a and 3b show the accuracy performance of the three methods in terms of their robustness to sparsity. The proposed method always outperforms the other methods, and keeps a steadily high accuracy when the sparsity of datasets increases. Thus, our method is more robust to sparsity than the other two methods.

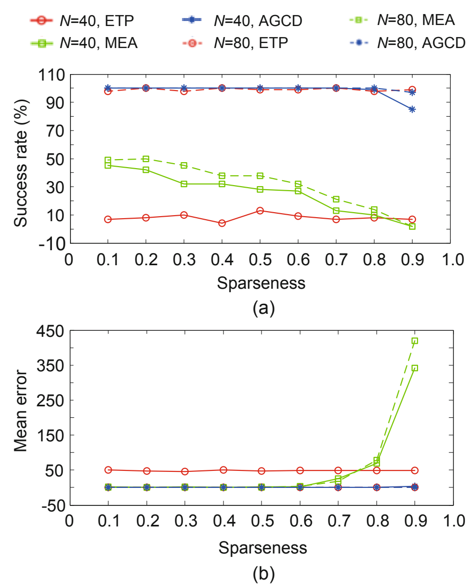

Fig. 3 Success rate (a) and mean error (b) with regard to sparseness λ ($p = 100$, $r = 10$)

4.2.3 Accuracy with regard to the number of observations

Figs. 4a and 4b show the accuracy performance of the three methods as the size of datasets N varies. Obviously, the more observations that can be used for the period estimation, the higher the accuracy that can be achieved. Our method can achieve high accuracy with smaller datasets than the other two methods. Our method accurately estimates the period with only 40 observations, while the ETP method needs at least 80 observations to achieve the same accuracy. This means that our method can be used in the situation where large datasets are hard to acquire.

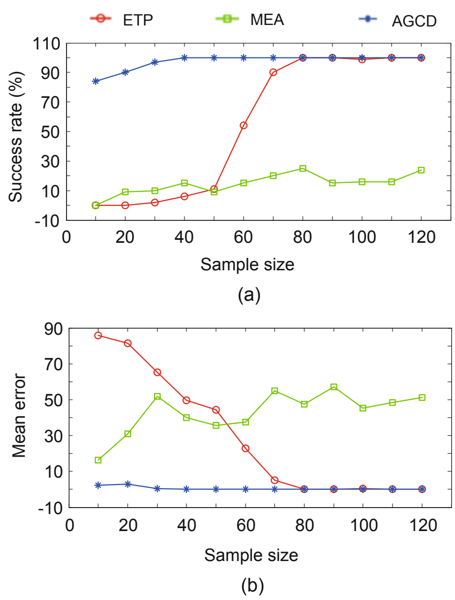

Fig. 4 Success rate (a) and mean error (b) with regard to sample size ($p = 100$, $r = 10$, $\lambda = 0.8$)

4.2.4 Accuracy with regard to different periodic events

Figs. 5a and 5b show the accuracy of the three methods over various periodic events with different periods. The proposed method keeps a steadily high accuracy as the magnitude of period p becomes larger, while the accuracy of the other two methods significantly degrades, especially the ETP method. This indicates that the accuracy of our method is independent of the magnitude of the period, which makes it applicable for various periodic events with different periods.

4.3 Efficiency

The time performance of the three methods is shown in Fig. 6. The running time of the proposed method is close to that of the MEA method, which is currently the fastest periodic estimation method to the best of our knowledge. And the efficiency of our method is much less sensitive to the size of

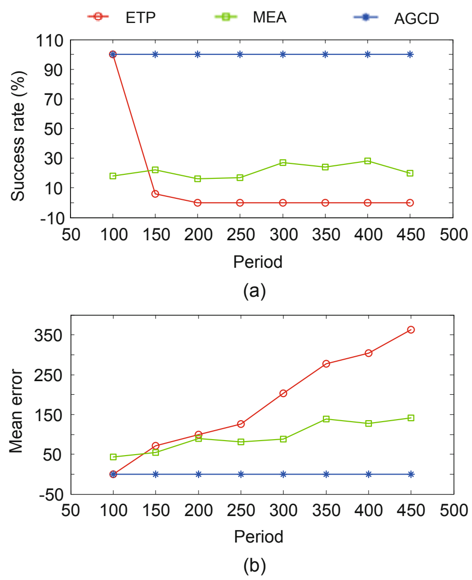

Fig. 5 Success rate (a) and mean error (b) with regard to different period values ($N = 40$, $r = 10$, $\lambda = 0.8$)

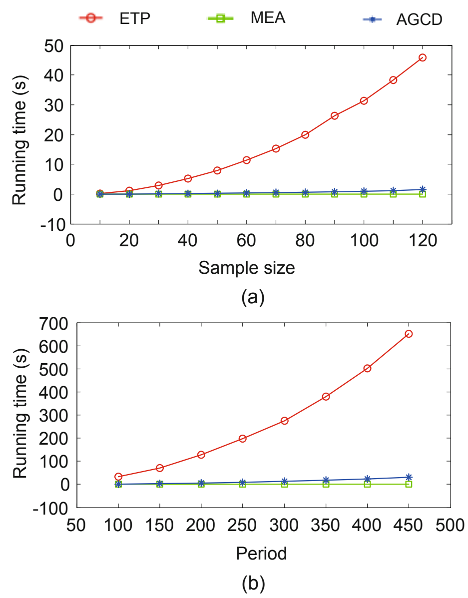

Fig. 6 Efficiency with regard to sample size N with $p = 100$ (a) and period p with $N = 40$ (b) ($r = 10$, $\lambda = 0.8$)

datasets and the magnitude of the period. Among the three methods, the ETP method is the most time-consuming one, whose running time grows dramatically as the size of datasets or the magnitude of period increases. Although our method costs a little more time than the MEA method, the accuracy of our method is much higher.

5 Conclusions

In this paper, we propose an efficient and effective method for detecting latent periods from finite, sparse, and noisy observational datasets of periodic events. The good performance of the proposed method is demonstrated via comprehensive experiments. Through the experiments we can see that the AGCD method outperforms the other two methods in most situations. The AGCD method is less sensitive to the size of datasets and the magnitude of period than the compared methods, and it has high efficiency which is close to the fastest period detection method, MEA. All these benefits of the proposed method make it more general in application than the other methods.

In our future work, we will consider extending our method to more complicated scenarios, such as periodic events with multiple periods or with multiple phases, which are also common situations in the real world.

References

Casey, S.D., Sadler, B.M., 1996. Modifications of the Euclidean algorithm for isolating periodicities from a sparse set of noisy measurements. *IEEE Trans. Signal Process.*, **44**(9):2260-2272. [doi:10.1109/78.536682]

Clarkson, I.V.L., 2008. Approximate maximum-likelihood period estimation from sparse, noisy timing data. *IEEE Trans. Signal Process.*, **56**(5):1779-1787. [doi:10.1109/TSP.2007.912268]

Fogel, E., Gavish, M., 1988. Parameter estimation of quasi-periodic sequences. Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, p.2348-2351. [doi:10.1109/ICASSP.1988.197110]

Gray, D.A., Slocumb, J., Elton, S.D., 1994. Parameter estimation for periodic discrete event processes. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, p.93-96. [doi:10.1109/ICASSP.1994.389868]

Howgrave-Graham, N., 2001. Approximate integer common divisors. Proc. Int. Conf. on Cryptography and Lattices, p.51-66. [doi:10.1007/3-540-44670-2_6]

Huijse, P., Estevez, P.A., Zegers, P., *et al.*, 2011. Period estimation in astronomical time series using slotted correntropy. *IEEE Signal Process. Lett.*, **18**(6):371-374. [doi:10.1109/LSP.2011.2141987]

Huijse, P., Estevez, P.A., Protopapas, P., *et al.*, 2012. An information theoretic algorithm for finding periodicities in stellar light curves. *IEEE Trans. Signal Process.*, **60**(10):5135-5145. [doi:10.1109/TSP.2012.2204260]

Junier, I., Hérisson, J., Képès, F., 2010. Periodic pattern detection in sparse Boolean sequences. *Algor. Mol. Biol.*, **5**:31.1-31.11. [doi:10.1186/1748-7188-5-31]

Li, Z., Ding, B., Han, J., *et al.*, 2010. Mining periodic behaviors for moving objects. Proc. 16th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, p.1099-1108. [doi:10.1145/1835804.1835942]

Li, Z., Wang, J., Han, J., 2012. Mining event periodicity from incomplete observations. Proc. 18th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, p.444-452. [doi:10.1145/2339530.2339604]

McKilliam, R., Clarkson, I.V.L., 2008. Maximum-likelihood period estimation from sparse, noisy timing data. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p.3697-3700. [doi:10.1109/ICASSP.2008.4518455]

Sadler, B.M., Casey, S.D., 1998. On periodic pulse interval analysis with outliers and missing observations. *IEEE Trans. Signal Process.*, **46**(11):2990-3002. [doi:10.1109/78.726812]

Sidiropoulos, N.D., Swami, A., Sadler, B.M., 2005. Quasi-ML period estimation from incomplete timing data. *IEEE Trans. Signal Process.*, **53**(2):733-739. [doi:10.1109/TSP.2004.840761]