Juan Yu and Pei-zhong Lu, 2015. AGCD: a robust periodicity analysis method based on approximate greatest common divisor. *Frontiers of Information Technology & Electronic Engineering*, **16**(6):466-473. [doi:10.1631/FITEE.1400345]

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

Key words: Periodicity analysis, Period detection, Sparsity, Noise, Approximate greatest common divisor (AGCD)

Contact: Juan Yu
Email: juanyu10@fudan.edu.cn
ORCID: http://orcid.org/0000-0001-9562-1325

Introduction

- In this paper, we propose a novel period detection method, AGCD, 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.

The AGCD algorithm

Algorithm 1 AGCD period estimation

Require: a set of observations $T = \{t_1, t_2, \ldots, t_N\}$.

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 ∈ (−r, r) and r_j ∈ (−r, r)
  g = gcd(d_i + r_i, d_j + r_j);
  if g ≥ t
    if g ∈ G
      occur[g] = occur[g] + 1;
    else
      occur[g] = 1;
      G = G ∪ g;
    end if
  end if
end for
```

(3) Find the GCD g with the highest frequency as the period, i.e., $p = g$.

Accuracy

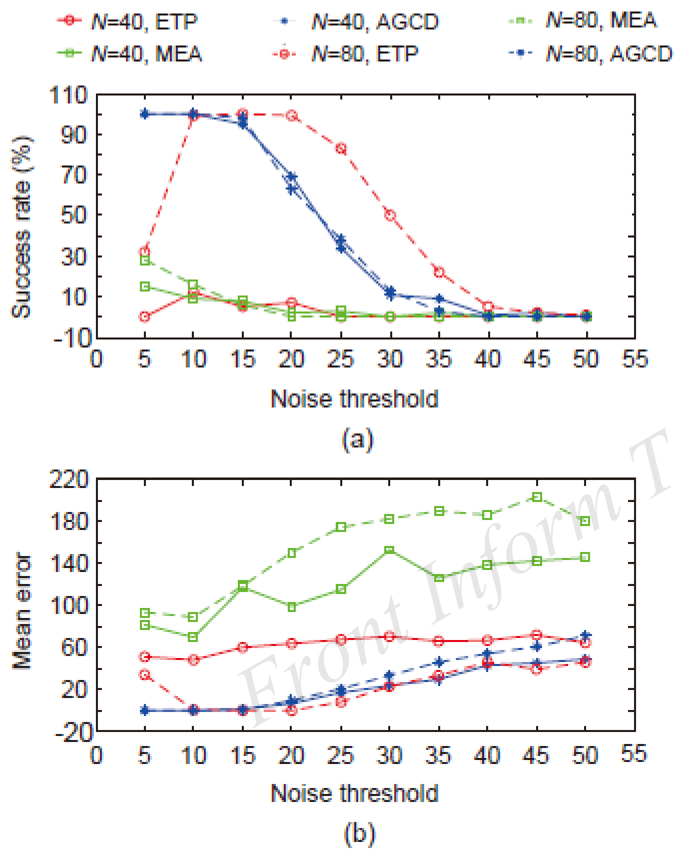

Fig. 2 Success rate (a) and mean error (b) with regard to noise threshold r ($p = 100$, $\lambda = 0.8$)

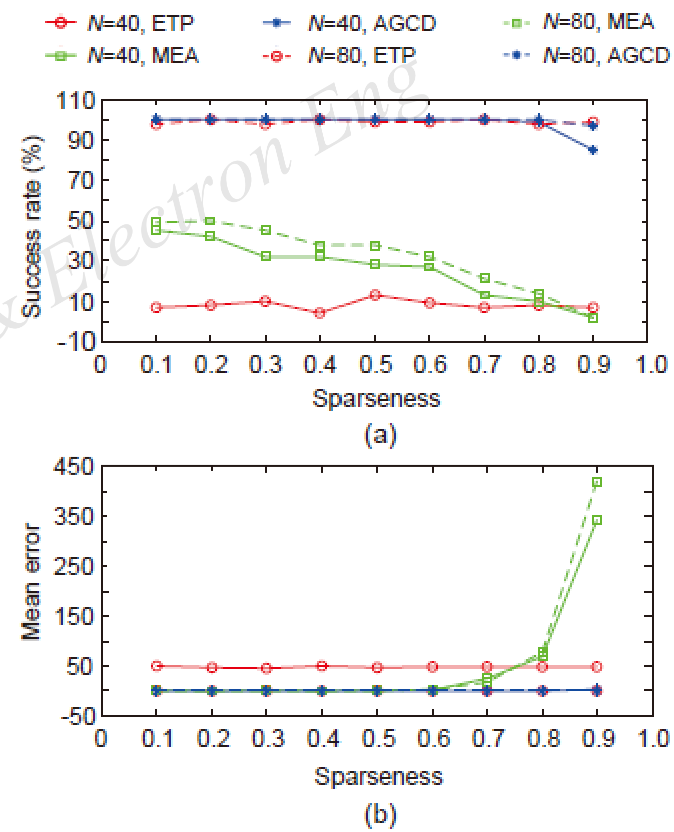

Fig. 3 Success rate (a) and mean error (b) with regard to sparseness λ ($p = 100$, $r = 10$)

Accuracy

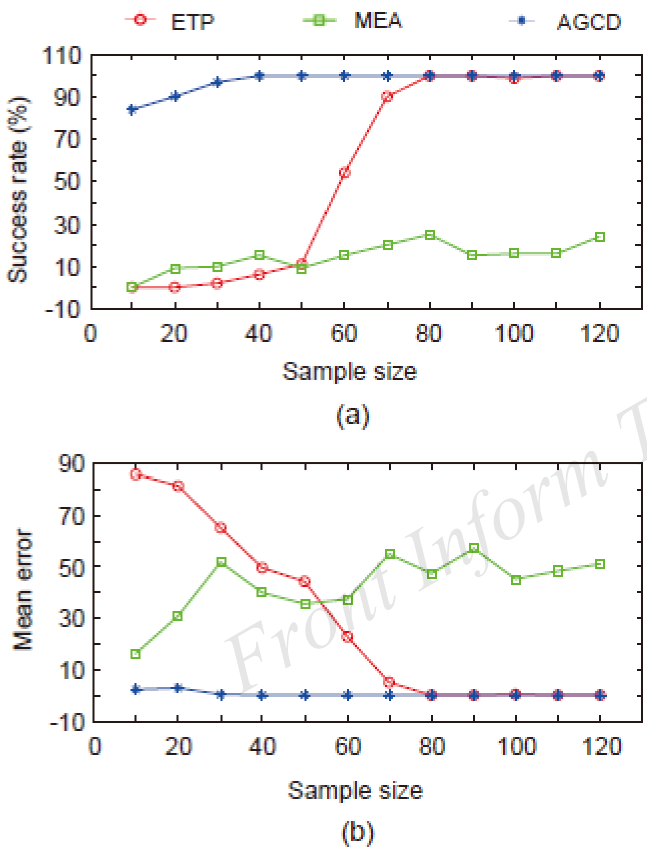

Fig. 4 Success rate (a) and mean error (b) with regard to sample size ($p = 100$, $r = 10$, $\lambda = 0.8$)

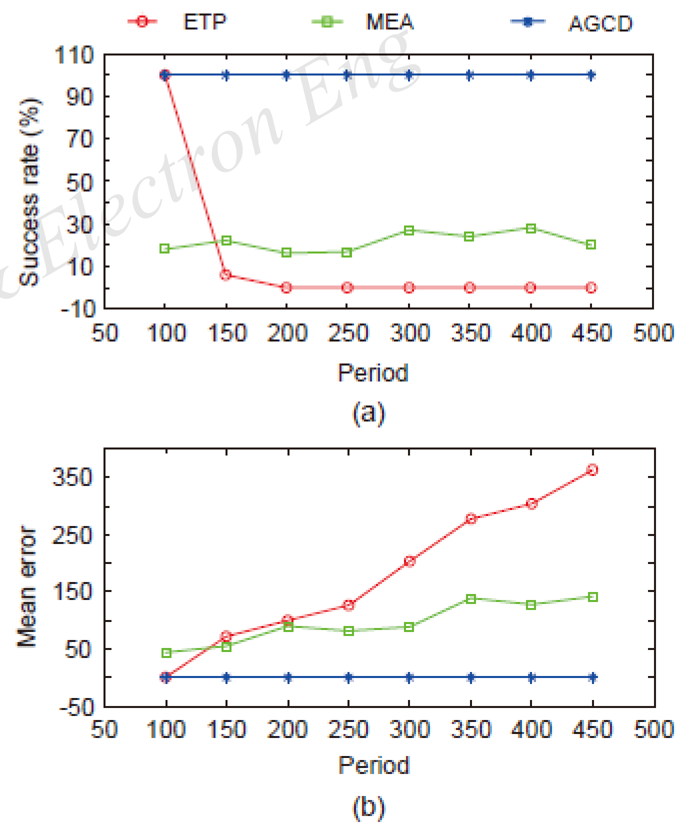

Fig. 5 Success rate (a) and mean error (b) with regard to different period values ($N = 40$, $r = 10$, $\lambda = 0.8$)

Efficiency

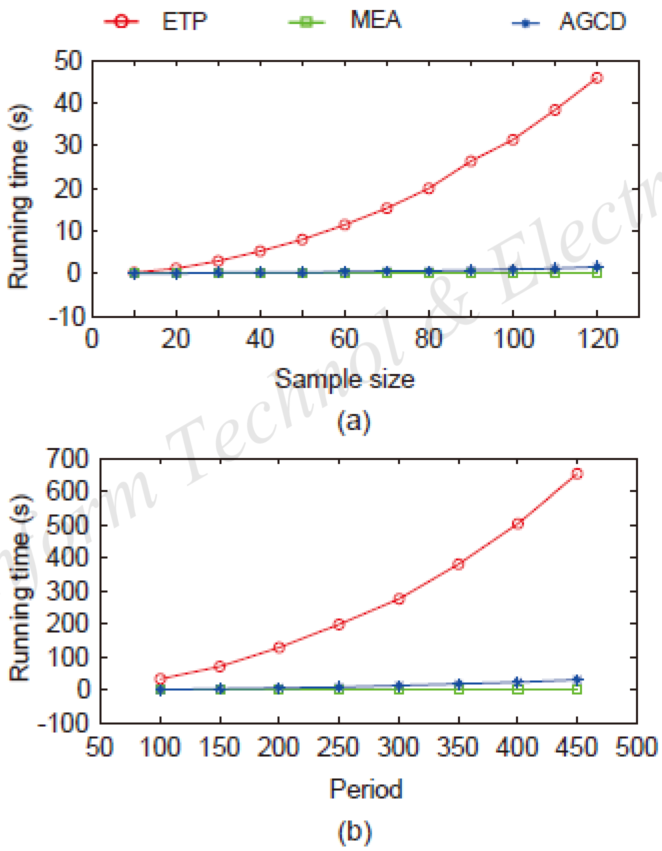

Fig. 6 Efficiency with regard to sample size N with $p = 100$ (a) and period p with $N = 40$ (b) ($r = 10$, $\lambda = 0.8$)

Conclusions

- 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.
  - 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.