

An Improved Correlation Identification Algorithm in Frequency Domain

Qiu Lishan, Hao Kaixue and Li Mei ⁺

Information Engineering, China University of Geosciences, Beijing, China

Abstract. An improved correlation identification algorithm in frequency domain is proposed by us. Using this method to calculate the impulse of system, we added a small constant to the result of the Fourier-transform autocorrelation of the input signal to eliminate the bad effects and to increase the identification accuracy of the system, because there is a zero in divisors. When the sampling rate is greater than the clock frequency (we called more-times speed), there will be a zero problem, and the zero will be aggrandized. When the zeros exist in divisors, this will produce a larger error. Formula deduction and simulations shows that this method can be used to restrain noise and improve the accuracy of identification. Compared with the traditional identification, the identification gets a lot of improvement. The degree of exactitude by using proposed method is about 100 times as much as the traditional method.

Keywords: correlation identification, more times speed, zero problem, improvement of accuracy.

1. Introduction

In the process of geophysical exploration of underground medium, there is a large random noise which has great influence on the result of exploration. Using correlation identification, we improved the quality of the exploration underground, because the method itself could suppress the noise. When the current of the medium is stable and the density is small, the system of the earth could be viewed as Linear Time-invariant System. Actually we can regard the process of detecting the medium as the process of solving function of tested system (Li Mei, 2013) [1]. The traditional solution of getting the function of the system is in time domain. P.M.Duncan formulated the basic principle of correlation identification and used it in geophysical prospecting at the earliest (1980) [2]. The scholars of later such as Luo Bao counted the rectangular model of m sequence autocorrelation instead of the real (2014) [3]. Based on the rectangular model, some simulations and result of exploration are published by Zhang Jianguo (2012) [4]. Zhu Changshan proposed the triangular of m sequence autocorrelation and reduced the deviation of identification, but the method is accomplished in time domain (1989) [5]. Remarkably having used FFT to calculate the cross-correlations, Shi Qingzhou referred to frequency domain, but the system is calculated in domain approximately (1984) [6]; Luo Weibin formulated and solved the function in the frequency, but no details about the result in frequency domain (2007) (2009) [7] [8]. The autocorrelation of m sequence that is approximate is the major way of solving the system function with correlation in time domain. A big error will occur in time domain. If we identify the system in frequency domain in more-times speed, the divisors in the algorithm will contain zero. Therefore the zero points lead to jitter so that the identification a failure. Hence we changed the algorithm in more-times-speed in the frequency domain. Adding a small constant to the Fourier transform of m sequence, we solved the zero problem.

2. Method And Simulation

We view the earth as a Linear Time Invariant system, the identification process as illustrated in Fig. 1.

⁺ Corresponding author. Tel.: +8613439876640.
E-mail address: maggieli@cugb.edu.cn.

The system input is pseudo random sequence (m sequence) $x(t)$. Through the system $h(t)$, we get the signal of output which contains additive noise $n(t)$. Finally the signal of output is $z(t)$.

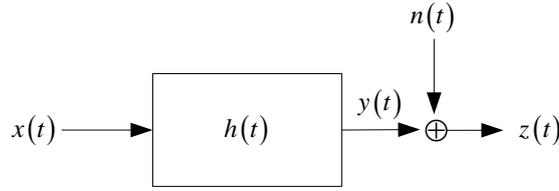


Fig. 1: The model of identification.

The whole process can be described by (1).

$$z(t) = x(t) * h(t) + n(t) \quad (1)$$

2.1. The Tradition Identification in Frequency Domain

The tradition identification in domain is (2).

$$Z(w) = X(w)H(w) + N(w) \quad (2)$$

The spectrum of noise $N(w)$ cannot be subtracted, it is unknown. The response in frequency domain can be calculated by (3).

$$H(w) = Z(w) / X(w) \quad (3)$$

2.2. The Correlation Identification in Frequency Domain

The correlation identification is a kind of common method system identification field, it used to identify the uncharted function and parameters of system. Through correlation calculation to (1), we get Wiener-Hopf equation (4).

$$R_{xz}(t) = R_{xx}(t) * h(t) + R_{xn}(t) \quad (4)$$

The signal of input is m sequence that is random. Therefore the input is not relative to noise.

$$R_{xn}(t) = 0 \quad (5)$$

The noise can be restrained in the Wiener-Hopf equation. The system $h(t)$ is that we need to get. There are two ways to calculate $h(t)$ in frequency domain and in time domain. The pointed part of one cycle of m sequence autocorrelation is replaced by an impulse function, the part of the other cycle instead of constant. The impulse response is (6).

$$h(k) = \frac{N}{a^2 \Delta (N+1)} \left\{ \frac{1}{N-p} \sum_{i=0}^p R_{xz}(i) + R_{xz}(k) \right\} \quad k = 0, 1, 2, \dots, p \quad (6)$$

It (7) is the identification calculation in the time domain. Δ is the code width of the m sequence. a is the amplitude of the m sequence. N is the length of the m sequence. R_{xz} is the cross-correlation of input $x(t)$ and output $z(t)$. Given being the approximate process, the autocorrelation would lead to much error. Hence we calculate the system $h(t)$ in the frequency domain. After Fourier transform, we change (4) to (7).

$$\begin{aligned} \int_{-\infty}^{\infty} R_{xz}(t) e^{-j\omega t} dt &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} R_{xx}(\tau) h(t-\tau) d\tau \right] e^{-j\omega t} dt \\ &= \int_{-\infty}^{\infty} R_{xx}(\tau) \left[\int_{-\infty}^{\infty} h(t-\tau) e^{-j\omega t} d\tau \right] d\tau \\ &= R_{xx}(\omega) H(\omega) \end{aligned} \quad (7)$$

We replace the left of (7) with $R_{xz}(\omega)$.

$$H(\omega) = R_{xz}(\omega) / R_{xx}(\omega) \quad (8)$$

It (8) is the identification calculation in frequency domain. Then we transform $H(\omega)$ with the inverse Fourier transform. We get the final result.

2.3. The Improved Correlation Identification in Frequency Domain

When the sampling rate is higher than the clock frequency of m sequence (more-times speed), 0 points exist in the spectrum of m sequence autocorrelation. When divisors are the spectrum $R_{xx}(w)$ (Fig. 2), the 0 points or approximate 0 points make enormous joggle in the same place of the frequency. These disturb severely that identification system is distinguished.

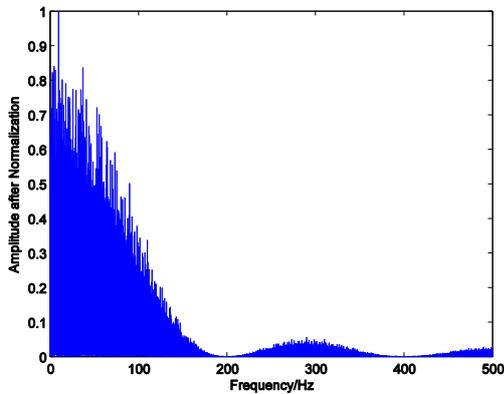


Fig. 2: The spectrum of m sequence autocorrelation in 5-times speed. There are two places (200Hz, 400Hz) close to 0. It causes the joggle.

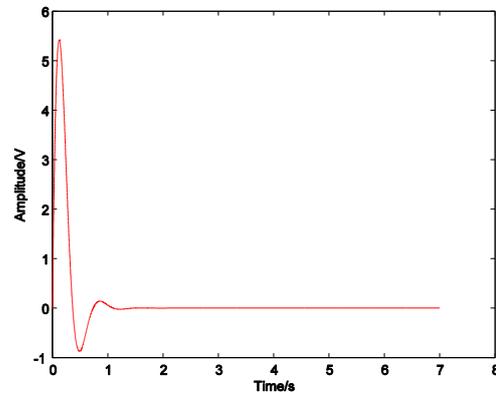


Fig. 3: The spectrum of m sequence autocorrelation is in 5-times speed. There are two places (200Hz, 400Hz) close to 0. It causes the joggle.

We contrast the consequence between the given system and the identification after we utilize different algorithm. Take an example of a second-order system (Fig. 3). There is enormous rise in identification spectrum (Fig. 4) after we observe the spectrum of identification system. Compared the same frequency from Fig. 2 and Fig. 4, we find that the 0 points lead to the rise. The result of identification cannot be recognized after 0.5s.

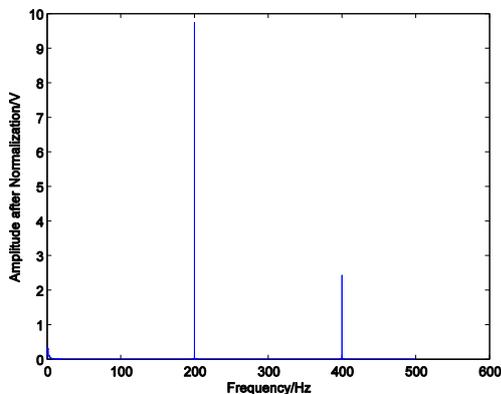


Fig. 4: The correlation identification system of second-order system in frequency domain (no noise)

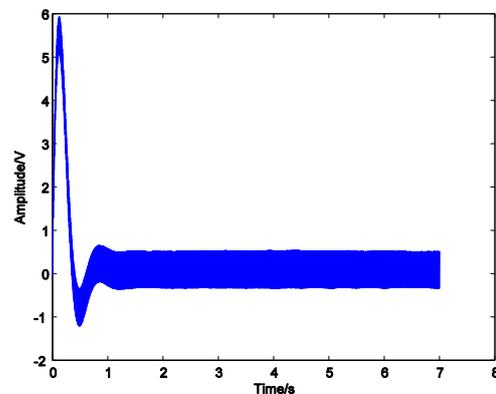


Fig. 5: The tradition correlation identification system of second-order system in time domain (no noise). There is a big jitter in the result after 0.5s.

The sampling of m sequence in the actual always produces certain distortion after launched. There is a large error in the time domain because of approximation. Calculating with precise sampling of m sequence directly can reflect the actual situation in the frequency domain. In the process of Fourier transform, we do not process data approximately. Hence the spectrums are accurate. The zero points problem exist in the traditional correlation identification. This is the ill-conditioned problem. The result of identification as Fig. 5 shown.

Therefore we hope that the amplitude of the points close to 0 could be larger, and it could not affect the result too much. We add a small real number α in the spectrum of autocorrelation. (8) is changed to (9).

$$H_1(w) = \frac{R_{xz}(w)}{R_{xx}(w) + \alpha} \quad (9)$$

The constant α in (9) is tiny so that it would not affect the autocorrelation spectrum too much. The method can make the 0 points vanished. When we transform the spectrum with IFFT, they also will not break the complex-conjugate poles. At the same time, the improved correlation identification can restrain the random noise obviously. Finally, we get the algorithm as (9). Through simulat continuously in terms of (9), the appropriate value of α is 0.01.

2.4. The Comparison of Identification Result

Setting the test function with the second-order system $h(t) = 11.5e^{-5t} \sin(8.6t)$ (as Fig. 3) which is viewed as the earth system, we build the model of the identification simulation. The input of signal is m sequence. Through convolution with the system, the random noise is added to output. The parameters of m sequence such as: 5-times speed, 1 cycle, amplitude of 1V, 14 rank. We get the results of identification with no noise and 5V random noise. Hereby the deviation we adopt is the formula from Qingzhou Shi [9]. $\hat{h}(k)$ is the result of identification system. $h(k)$ is the theoretical values of system. The smaller the value of the result is, the more precise the identification is.

$$Q_0 = 10 \log \left\{ \frac{\left[\sum_{k=0}^{N-1} (\hat{h}(k) - h(k))^2 \right]}{\left[\sum_{k=0}^{N-1} h^2(k) \right]} \right\} \quad (13)$$

We compared with traditional identification method for the purpose of explaining the restrain noise ability of correlation. The calculation is as (14)

$$H(w) = \frac{Z(w)}{X(w) + \alpha_1} \quad (14)$$

- The simulation of traditional identification: The traditional result with no noise is Fig. 6. After adding 5V random noise, there are big jitters in the identification result as Fig. 7. The deviation is -3.6234dB. It can be seen that the traditional identification method could identify the system in the case of no noise, but the result of 5V random is too bad. Thus that is to say that the traditional method with disability of random noise.

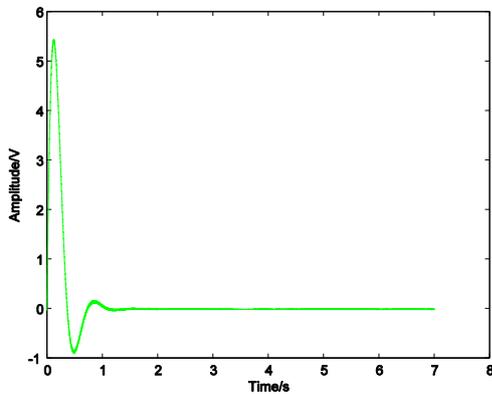


Fig. 7: The traditional result of identification (no noise).

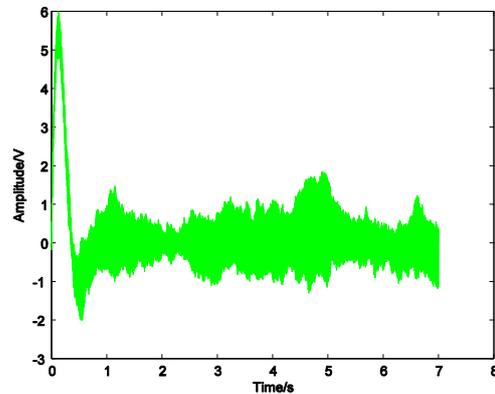


Fig. 8: The traditional result of identification (5V random noise).

- The improved correlational identification: The improved correlation identification result with no noise is Fig. 9. After adding 5V random noise, there are a little jitters in the identification result as Fig. 10. The deviation is -23.5161dB. Compared Fig. 10 with Fig. 8, we intuitively know that the improved correlational identification can restrain the noise. Compared Fig. 9 with Fig. 7, the improved correlational identification greatly reduces identification of pathological effect in the frequency domain.

To explain the excellent abilities of the improved correlation identification further, we do the more simulations. The sampling of system $h(t)$ is 1k Hz. The input of system is m sequence. The specific parameters of m sequence show as the table (1) (the “T” is on behalf of the errors of the traditional identification, the “I” is on behalf of the errors of the improved identification, the unit is dB). Compared Fig.

7 with Fig. 9, the results of two identifications are almost same in no noise. Compared Fig. 8 with Fig. 10, the result of the improved identification is obviously better than the result of the traditional identification after we add 5V random noise. Combined with table 1, we can further know that the improved correlation identification could restrain random noise preferably. The error of improved method is less than the traditional method about 19 dB (almost 100 times).

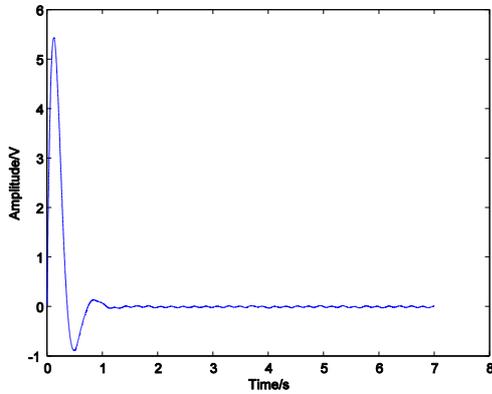


Fig. 9: The result of improved correlation identification (no noise).

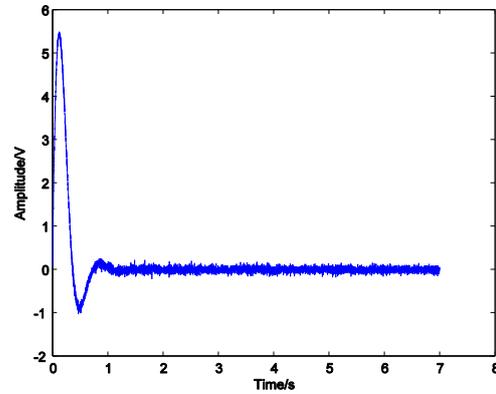


Fig. 10: The result of improved correlation identification (5V random noise).

TABLE I: Errors for the various parameters of traditional identification and the improved correlation identification

α		α_1						
		0.004	0.006	0.008	0.010	0.012	0.014	0.016
Speed (Rank 14 1V)	3 (T)	6.661	3.871	-0.492	-0.820	1.338	2.023	0.007
	3 (I)	-21.031	-22.122	-22.841	-22.898	-13.437	-22.798	-22.951
	4(T)	-0.177	-0.687	0.207	-3.363	-0.200	1.084	-1.127
	4 (I)	-21.323	-22.009	-23.103	-23.567	-23.853	-23.738	-18.242
	5(T)	-1.350	0.315	-1.200	-0.119	-2.114	8.2253	6.953
Rank (5-times Speed 1V)	5 (I)	-21.765	-22.645	-23.360	-23.516	-24.132	-24.510	-24.804
	12(T)	4.533	2.179	-0.225	0.531	0.269	1.529	2.650
	12 (I)	-13.166	-7.266	-4.624	-12.154	-14.205	-14.955	-15.423
	13(T)	2.591	6.244	0.101	0.471	0.058	1.051	-0.066
	13 (I)	23.770	-18.909	-19.826	-19.311	-11.095	-13.901	-18.454
Amplitude (Rank 14 5-times Speed)	14(T)	-0.698	0.734	-0.748	-1.191	-0.480	3.971	2.892
	14 (I)	-21.416	-22.788	-22.999	-23.489	-24.185	-24.545	-24.722
Amplitude (Rank 14 5-times Speed)	1(T)	2.892	0.605	-0.001	-2.009	0.033	4.398	10.956
	1 (I)	-21.803	-22.428	-23.282	-23.516	-24.249	-23.799	-24.689
	5 (T)	-11.463	-9.285	-10.566	-6.610	-7.545	-7.938	-9.440
	5 (I)	-28.526	-29.104	-29.861	-30.478	-30.540	-27.846	-31.383

The result of identification is a little random due to the impact of the random noise. The phenomenon shows that the same parameter gets different errors. But the errors would not be too high in a range.

From the simulations and the table I, we get the conclusion as follow:

- Generally, the data of improved traditional identification errors are higher than the improved correlational identification's with random noise. That suggests that the traditional identification method is easier to be influenced by random noise. The improved correlation identification is disturbed by random noise lightly. Hence the improved method is able to restrain random noise well.
- In common sense, increasing the times speed will increase the errors of the traditional identification, and the improved method could reduce the errors.

- Increasing range of m sequence could improve the result of identification with the improved method.
- Increasing amplitude of m sequence could improve obviously the result of identification with the improved method.
- The parameter α of improved correlation will choose 0.01 when the full results are considered.

Therefore both the improved method keeps the capacity of noise suppression, and it corrects the shortcomings of zero points in frequency domain.

3. Conclusion

Through the simulations and the theory analysis, we know that the result of the improved correlation identification is obviously superior to the traditional method in frequency domain. The improved method is able to restrain the random well. When we identify the system with the improved method, we try to choose the m sequence of the higher range and the larger amplitude. The higher sampling could get the better result, but too high sampling and range will lead to too long time. Hence the values should be chosen appropriately.

In the process of exploration, there is a lot of high random noise which make the exploration difficult [10]. The improved correlation can restrain the random noise well so that it can be used to get the parameters of the earth in the exploration. We need to test more noise in the simulation because there are many kinds of noise in the field. We will apply the method to realty in future. The input also can use the invert-repeats m-sequence. The auther will do more improved research in order to the method come into use.

4. Acknowledgements

The algorithm research is supported by the Natural Science Foundation of China (Grant No. 41374185 and Grant No. 41572347) and the Excellent Mentors Foundation of University of Geosciences in Beijing (Grant No. 2652015324), the authors thank sincerely.

5. References

- [1] Li Mei, Wenbo Wei, Weibin Luo, Qindong Xu. Time-Domain Spectral Induced Polarization Based on Pseudo-random Sequence, 2013, *Pure and Applied Geophysics* (12):2257~2262.
- [2] Duncan P. M., Hwang A., Edwards R. N. Bailey R. C. Garland G. D.. 1980, The development and applications of a wide band electromagnetic sounding system using a pseudo-noise source. *Geophysics*, 45(8): 1276~1296.
- [3] Jianguo Zhang. Investigation on Time Domain Coded Electromagnetic Exploration Method. Applied Science, *Journal of Radars*, Apr 2014.
- [4] Bao Luo. The Application and Implementation of the correlation identification in IP exploration. Ph.D. Thesis. Chang Sha: Southeast University, 2012.
- [5] Changshan Zhu. The Application Research and Implementation of Correlation Identification in Induced Polarization Prospecting. *Applied Science* 1989 (1).
- [6] Qingzhou Shi. A Improved Frequency-domain Algorithm for the Pseudorandom Correlation Method. *Information and Control*, 1984,(1):23~28.
- [7] Weibin Luo, Li Qingchun. Electromagnetic Exploration Based on System Identification for Seafloor Hydrocarbon Reservoir and Gas Hydrate. In: PIERS Proceedings, Moscow. 2009: 1344~1348.
- [8] Weibin Luo. Study on Pseudorandom Marine Controlled-Source Electromagnetic sounding with multi-offsets. Ph,D. Chang Sha: Central South University, 2007.
- [9] Qingzhou Shi. Simulation Study of Some Identification Errors for the Pseudo-Random Correlation Method. *Information and Control*, 1984, (5):28.
- [10] Jishan He. The 2n Pseudo Random Signal and the Application. The Institute of Geophysics of China. Xi'an: Xi'an Map Press, 1998.