Examining signal estimation methods by including missing signal parts

Kohei Okishio, Yoshiharu Koya
Dept. of Electronic Engineering Kobe City College of Technology
8-3, Gakuen higashimachi, Nishiku, Kobe, Hyogo 651-2194, Japan

E-mail: koya@kobe-kosen.ac.jp

Abstract

A signal often includes missing parts and is often affected by noise when it is obtained; thus, all available data includes errors. In this study, we considered a method in which only highly reliable signals, which are close to the true value, are extracted for estimating the signals. When highly reliable signals (candidate points) are extracted from the acquired signals at an arbitrary ratio, the signals from the remaining parts constitute the missing parts. The signals with missing parts after the extraction of the candidate points were estimated by Fourier series approximation. Herein, we examined the optimal extraction ratio for the estimation, including the order used for approximation. First, we examined the required percentage of the candidate points extracted from the acquired signals for increasing the estimation accuracy. As a result, if at least about 80% candidate points are available, the estimation accuracy was high. Subsequently, we proposed three methods for selecting the candidate points; we compared the errors between the original and the estimated signals, and the estimation iteration times. Consequently, the following method had the highest estimation accuracy. Further, we performed the estimation once using all the obtained signals and then removed the signal with large error between the original and the estimated signals at an arbitrary ratio and performed the estimation again. In this method, we can extract signals close to the true value. This method has fewer iteration times compared with other methods. Furthermore, by using the third order, the error was insignificant and the number of calculations was also limited.

1 Background

The estimation of the border of interest domain is necessary for the extraction of the region of interest. However, all candidate points (i.e., a signal at the edge surrounding the area) of the border are provided only in very rare cases. Therefore, a method for border estimation has been proposed that also includes the missing part [1]-[4]. When we obtain all the candidate points, most of these would include noise. Therefore, in this study, we investigated two factors assuming that all candidate points were obtained. The first is the accuracy of a signal due to the difference in the ratio of missing parts. The influence occurred in the estimated precision and calculation times due to the difference in the distribution and the ratios of the missing parts. We investigated the percentage of the obtained candidate points that can be used to improve the estimation accuracy. The second factor is regarding the order used for the estimation. If the order used for the approximation is high, the calculation time will be affected. Therefore, we investigated which order is suitable for use.

Fig2.1 Principle of approximation using Fourier series
2 Method

Fourier series was used to represent the original signals, including the missing parts for border estimation, as shown in Fig. 2.1. First, the original signals were estimated by the Fourier series and subsequently estimated as signals ($f_1$) lower than the original, as the original signals included the missing parts. Subsequently, the difference between the original and the estimated signal is calculated; these signals ($f_{\text{difference}}$) are the next approximating signals. Similarly, $f_2$ is estimated and this process is repeated. By adding these estimated signals, their sum is approximated as the original signals. If the missing parts are gathered, estimated signals are dented as shown Fig 2.2. Hence, selecting the candidate points must be sparse to some extent.

3 Simulation

3.1 Factors to consider when determining the ratio of missing parts

First, the edge estimation is performed using data, including white noise, to investigate the percentage of candidate points, which are required for high-precision estimation. The estimation is performed by removing data containing white noise from large or small absolute values at an arbitrary rate. We used Equation (1), which is assumed to be signals from the edge of the target area obtained from the primary signal.

The estimation is performed at the third order. Table 3.1 shows different ways of comparing the coefficients of the missing data, and Table 3.2 shows a comparison of the errors and the number of calculations. A total of 1000 trials are conducted, where each value shows the average. The error unit is [pixel/degree], as a signal in an image is assumed. The iteration time shows the estimated number of repetitions. The convergence condition is the case where the value did not change compared with the error of the immediately preceding estimation.

\[
f(\theta) = 100 + 50.4\sin(3\theta) + n(\theta)
\]  

As shown in Table 3.1, when noise data are deleted in descending order of absolute value, a value closer to the original coefficient is obtained, as compared to when estimation is performed without the missing parts. Further, even if the absolute values are deleted in the ascending order, the values are close when there are 20%-30% missing parts. A value similar to the original coefficient is obtained. However, the accuracy is lower as compared to the descending order.

Considering the errors and the number of calculations in Table 3.2 in both cases, the estimation is possible with high accuracy even if there are 20% missing parts.

The above results show that if about 80% of the candidate points are obtained, a highly accurate estimation can be performed.

However, depending on the selection criteria of the candidate points, problems exist such as increase in both the error and number of calculations by removing data with a small noise component.

Therefore, we investigated a way for selecting the candidate points.

Table 3.1 Results of coefficients of the estimated border due to different missing states.

| Ratio of missing parts | Value of coefficient | $a_0$ | $b_1$ |
|------------------------|----------------------|------|------|
| 0%                     | 100.3                | 50.5 |
| 10%                    | 99.9                 | 50.3 |
| 20%                    | 99.9                 | 50.4 |
| 30%                    | 99.9                 | 50.4 |

(a) In descending order of absolute value

| Ratio of missing parts | Value of coefficient | $a_0$ | $b_1$ |
|------------------------|----------------------|------|------|
| 0%                     | 100.3                | 50.5 |
| 10%                    | 99.8                 | 50.4 |
| 20%                    | 99.7                 | 50.3 |
| 30%                    | 99.6                 | 50.2 |

(b) In ascending order of absolute value
Table 3.2 Error result and number of calculations.

(a) In descending order of absolute value

| Ratio of missing parts | Error       | Iteration times |
|------------------------|-------------|-----------------|
|                        | mean variance | mean variance |
| 0%                     | 1.04        | 0.08            | 1.00        | 0.00 |
| 10%                    | 0.93        | 0.06            | 2.38        | 0.93 |
| 20%                    | 0.85        | 0.03            | 3.49        | 1.22 |
| 30%                    | 0.83        | 0.02            | 4.70        | 1.33 |

(b) In ascending order of absolute value

| Ratio of missing parts | Error       | Iteration times |
|------------------------|-------------|-----------------|
|                        | mean variance | mean variance |
| 0%                     | 1.04        | 0.08            | 1.00        | 0.00 |
| 10%                    | 1.16        | 0.10            | 1.72        | 0.99 |
| 20%                    | 1.31        | 0.13            | 1.74        | 1.86 |
| 30%                    | 1.50        | 0.18            | 1.64        | 2.41 |

3.2 Consideration of orders used for determination and estimation of candidate points

In Section 3.1, we obtained an absolute value from the noise data we created and estimated it by making it disappear at an arbitrary ratio from the largest value; however, in an actual signal, noise components cannot be deleted by using such a method. We then investigated a way of selecting the candidate points. In this estimation, we used second order to fifth order, and the missing ratio is 20%.

3.2.1 Method of averaging coefficient

Subsequently, we carried out the estimation using the algorithm shown in Fig. 3.1. In this algorithm, random values were made to disappear from the original data. Next, we applied Fourier transform and inverse Fourier transform to obtain the Fourier coefficient. We added this coefficient to the previous coefficient. The first process was repeated if the arbitrary times have not been reached. When the arbitrary times are reached, we divided the coefficient with the arbitrary times. Finally, we estimated the signal using the coefficient. In this estimation, five trials were conducted. The results are summarized in Table 3.3.

The third order showed high estimation accuracy when this method was used. However, if the data are correct, the accuracy of the estimation may be reduced, as the data are randomly deleted. Furthermore, the increase in the number of trials increases the number of calculations for the entire estimation. Therefore, this method might not be stable for making an accurate determination of the candidate point.

The estimation was then applied by using the method shown in the next section.

![Averaging coefficient algorithm](image-url)
Table 3.3 Results of using an averaging coefficient algorithm.

| Order | Error (pixel/degree) | Iteration times |
|-------|----------------------|-----------------|
| 2     | 32.1                 | 37              |
| 3     | 1.23                 | 22              |
| 4     | 2.62                 | 52              |
| 5     | 2.76                 | 51              |

3.2.2 Method of using difference data of the adjacent data

The signal is assumed to be smooth, and the differences between the current, subsequent, and next data are determined. Thereafter, the data are excluded in descending order and an estimation is made. Equation (1) including the white noise is used. Table 3.4 summarizes these results.

Table 3.4 Errors and calculation times when the method of using difference data of adjacent data is applied.

| Order | Error (pixel/degree) | Iteration times |
|-------|----------------------|-----------------|
| 2     | 32.2                 | 8               |
| 3     | 1.00                 | 5               |
| 4     | 2.73                 | 10              |
| 5     | 2.94                 | 12              |

This method having an order 3, produced the most accurate estimation result. However, the use of the difference data from the adjacent data could not remove noise components with high accuracy, as the reference position is different for each dataset.

3.2.3 Method for estimating candidate points from difference data of the original signal after a single estimation

In this method, the estimation carried out once without the missing parts. Thereafter, the difference between the original and the estimated signals is determined. Then, 20% of the data in order with the largest value is deleted, and the estimation results at each order are compared. The results shown in Table 3.5.

In this estimation, the error is almost the same with fewer calculations, as compared to other methods. The results of this method are almost the same as in Section 3.1.

Therefore, a signal close to the original signal can be extracted as a candidate point in this method.

The third order was the most accurate, based on these results. It can be concluded that the estimation accuracy increases when the order is the same; therefore, the signal needs to be changed and compared to previous results using Equation (1). Then, a similar comparison is made for the signal containing multiple harmonics. We used Equation (2) and compared the signals to find the signals with the third and fifth orders. The results are shown in Table 3.6.

\[
f(\theta) = 100 + 3.4 \cos(2\theta) + 5.1 \sin(3\theta) + 4.7 \sin(5\theta) -3.5 \cos(10\theta) + 2.9 \sin(54\theta) + n(\theta) \quad (2)
\]

From these results, if the signal includes fifth order, it is possible to estimate it with accuracy at third degree.

Table 3.5 Method for determining candidate points from difference data

| Order | Error (pixel/degree) | Iteration times |
|-------|----------------------|-----------------|
| 2     | 32.1                 | 4               |
| 3     | 1.21                 | 6               |
| 4     | 2.69                 | 6               |
| 5     | 2.36                 | 6               |

Table 3.6 Case of inclusion with multiple harmonics

| Order | Error (pixel/degree) | Iteration times |
|-------|----------------------|-----------------|
| 2     | 3.48                 | 6               |
| 3     | 1.40                 | 5               |
| 4     | 1.02                 | 6               |
| 5     | 3.22                 | 6               |

| Order | Error (pixel/degree) | Iteration times |
|-------|----------------------|-----------------|
| 2     | 3.97                 | 6               |
| 3     | 3.11                 | 6               |
| 4     | 3.50                 | 6               |
| 5     | 2.05                 | 6               |

4 Conclusions

We can estimate the original signal with high accuracy if about 80% of the candidate points are available, as shown in Tables 3.1 and 3.2. However, even if 80% points are available, a reliable method for determining the candidate points is important, as the accuracy varies depending on the missing state. In some cases, more accurate estimation might be required by using a method of averaging the coefficients; however, as the data are randomly lost, the correct data might also get lost. Therefore, accurate deletion of the incorrect data is not possible. Additionally, the number of calculations also increases as the process is executed numerous times.

In the method, using the difference data of the adjacent data, the estimation accuracy becomes low depending on the case. This is because the reference position of each data is different while the difference is noted.

An approximate signal is first estimated without the missing parts. The data with a possibility of noise could then be removed with high probability from the difference of the original signal. Negligible error is found, and the number of calculations is also limited, from the estimation results in Table 3.5. Currently, this is the method that can remove the
The order used for the estimation can be estimated to some extent by using at the third order.

From Table 3.6, it is confirmed that the estimation of the fifth-order signals at the third order has fewer errors than the estimation of the fourth order and the number of calculations is the same as the other orders. Therefore, when the estimation is performed at the third order, the original signal can be roughly estimated.

References

[1] Yoshiharu Koya, Masashige Hori, Isao Mizoshiri: Border Detection of Echocardiogram using Projection Operator, T.IEEE Japan, Vol.120-C No.8/9 pp.1236-1241 (2000-8) (in Japanese)

[2] Isao Mizoshiri, Naoto Yasugi, Kuniaki Kamiya: Border Detection of Legt Ventricle on Echocardiogram, T.IEE Japan, Vol.115-C, No.11, pp.1237-1244 (1995) (in Japanese)

[3] Shuichi Adachi, Masayuki Tanji: A System Identification Method for Nonuniform Sampling Data Based on Missing Data Estimation, T.IEE Japan, Vol.122-C, No.8, pp.1383-1384 (2002) (in Japanese)

[4] DANTE C. YOULA: Generalized Image Restoration by the Method of Alternating Orthogonal Projections, IEEE Trans.Circuits and Systems., Vol. Cas-25, No9, pp.694-701(September 1978)