Proceedings of the Research Institute of Atmospherics,
Nagoya University, vol. 26 (1979)

ON THE NON-NEGATIVITY CONSTRAINT IN ITERATIVE RECONSTRUCTION OF ASTRONOMICAL IMAGES

Masato ISHIGURO

1. Introduction

In many astronomical imaging systems, imaging process is represented mathematically by following equations:

$$I_m = \iint T_m(x,y)B(x,y)dxdy + N_m \ , \quad (m=1,2,\ldots, M). \tag{1}$$

$\{I_m\}$ is a set of measured data, $B(x,y)$ is a source brightness distribution, and $\{T_m(x,y)\}$ is a transformation caused by imaging instruments $\{N_m\}$ is a set of measurement noise or error, $\{x,y\}$ are spatial coordinates, and M is a number of independent measurements.

If $\{I_m\}$ is completely sampled with a good signal-to-noise ratio, equation (1) can be solved without difficulty, and $B(x,y)$ is obtained. In actual measurements, however, $\{I_m\}$ cannot be measured beyond the resolution limit of the instruments. Even the data within the limit are missed in some cases. To cope with these difficulties, iterative reconstruction techniques (hereafter to be called IRT) incorporated with *a priori* information on $B(x,y)$ have been used.

The constraint that the brightness should not be negative:

$$B(x,y) \geq 0, \text{ all } x \text{ and } y, \tag{2}$$

is a very useful *a priori* information in image reconstruction from incomplete data, such as in very-long-baseline interferometry (Fort and Yee, 1976). An application of the non-negativity constraint to

the aperture-synthesis data was described by Högbom (1969). A very similar procedure has been used in ART (Algebraic Reconstruction Technique) developed for the computerized tomography.

In these applications, all negative values in B(x,y) are set to zero. This operation will not be an unique way to assure the non-negativity on B(x,y). In this paper, we examined the properties of the IRT in which all negative values are multiplied by some negative factor (loop gain), rather than replaced by zeros. It is expected that this modification will accelerate the convergence of the iteration process and will improve the final results.

As an example of imaging process, Fourier transform is considered:

$$T_m(x,y) = \exp[-j2\pi(u_m x+v_m y)]. \qquad (3)$$

The result is directly applicable to aperture synthesis or Fourier transform spectroscopy.

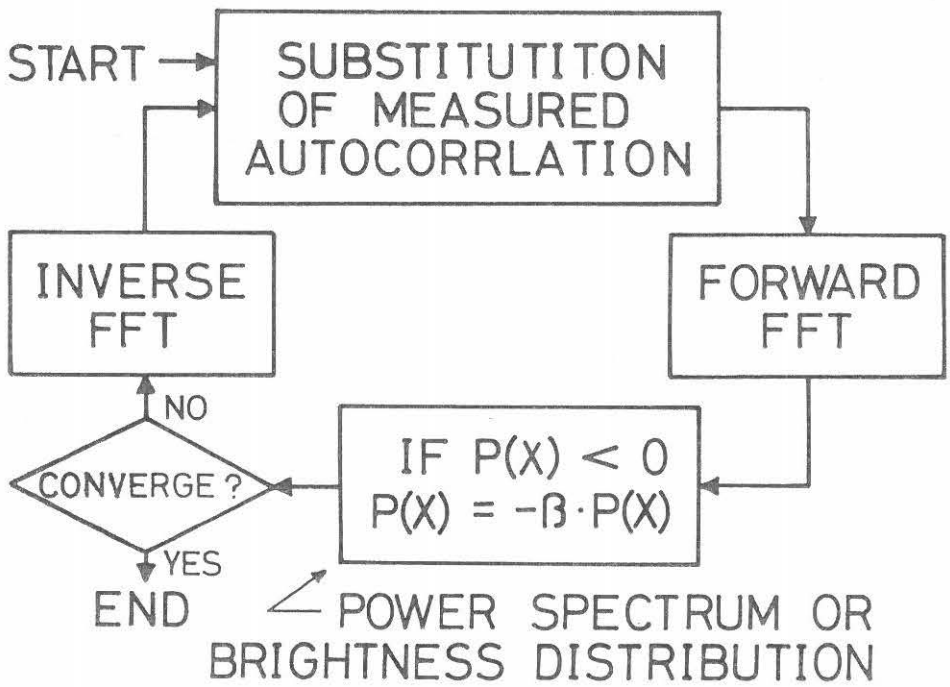

Fig. 1. IRT (Iterative Reconstruction Technique) with non-negativity constraint.

2. The algorithm of IRT and simple simulation experiments

Figure 1 shows the simple algorithm of Fourier iteration as an example of the IRT. For simplicity, one-dimensional case is considered. In aperture-synthesis observations, measured autocorrelation function (hereafter to be called ACF) is a complex function of the spatial frequency. In Fourier transform spectroscopy the ACF is a real function of the time lag. Though, in general, measured spatial frequencies may not be evenly spaced, we used the FFT (Fast Fourier Transform) to reduce the computing time of simulation experiments.

Figure 2 shows an example of estimating power spectrum from the measured ACF:

$$R(u) = \delta(u) + 5\cos(40\pi u/128) + 10\cos(48\pi u/128),$$
$$u = n\Delta u \ , \ n = 0, 1, \ldots, 19. \qquad (4)$$

In this example, sampling is uniform but the maximum lag is limited to $19\Delta u$.

The direct Fourier transform (Equation 5) of the ACF is shown in Figure 2 (a).

$$P(x) = F.T.[\ R(u)\ ] \ , \qquad (5)$$

where P(x) denotes power spectrum (or brightness distribution). It contains many strong sidelobes and the resolution is so poor that the two sinusoids (or two point sources) cannot be distinguished. A conventional approach to reduce the sidelobes in power spectrum is to taper the ACF at large lags. Figure 2 (b) is the result of a triangular taper (Bartlett window). Sidelobes are reduced with a deterioration of resolution.

At this stage, the IRT performs the following operation on P(x):

$$P(x) = -\beta \cdot p(x) \ , \text{ where } p(x) < 0 \ , \qquad (6)$$

where β is a non-negative factor (loop gain). The conventional method corresponds to the case when $\beta = 0$. This is a very simple realization of the non-negativity constraint on P(x). After the operation (6), the inverse Fourier transform of P(x) is performed to obtain a modified ACF which satisfies the non-negativity constraint. The operation (6) not only generates non-zero values in the unmeasured portion of the

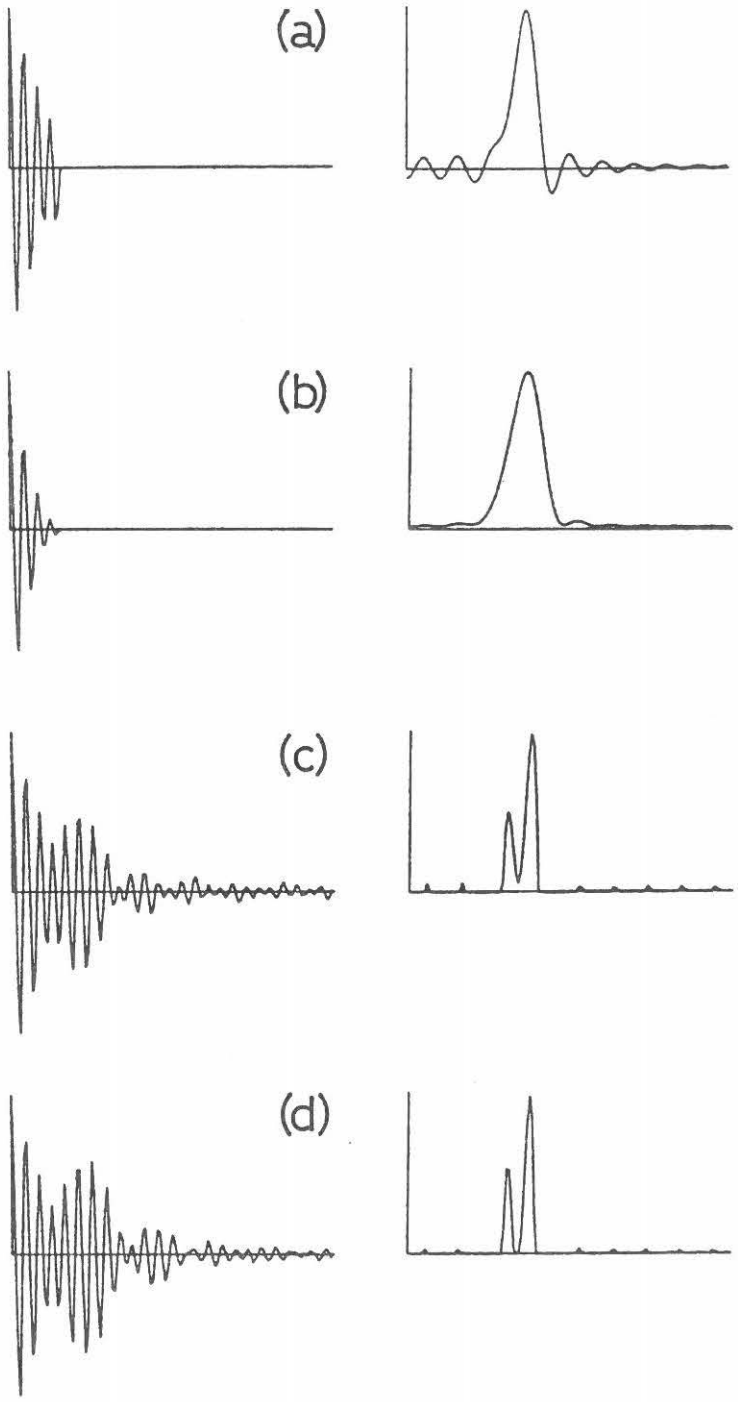

Fig. 2. An example of two point sources.
Left : ACF, right : power spectrum.
(a) Direct Fourier transform.
(b) Fourier transform after tapering ACF at large lags.
(c) Result of IRT with loop gain, $\beta = 0$.
(d) Result of IRT with loop gain, $\beta = 2.4$.

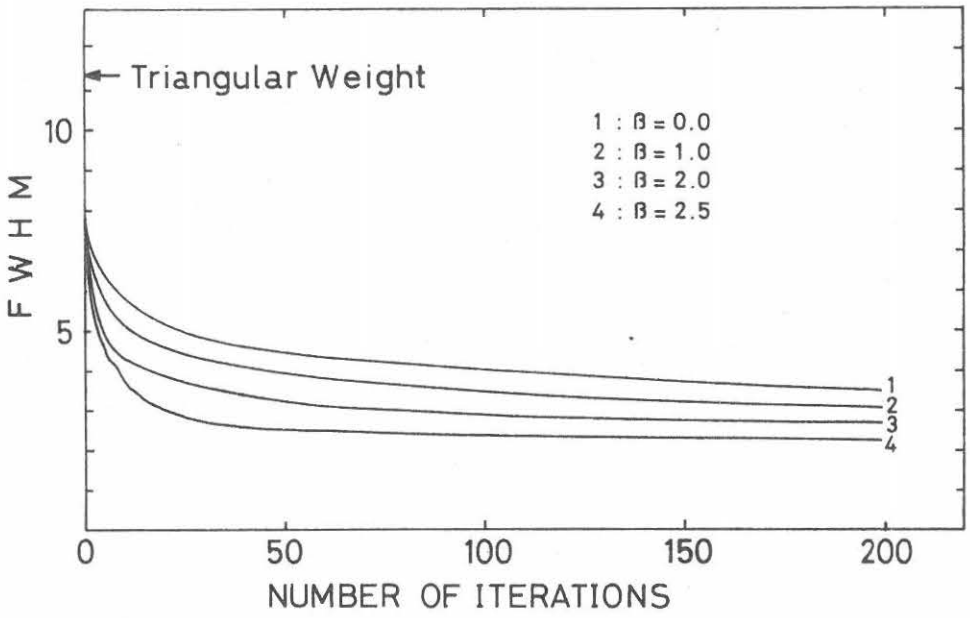

Fig. 3. Convergence of IRT for various loop gains.

ACF, but also modifies the values in the measured portion of the ACF.

We assumed in the simulation experiments that the measured ACF does not contain noise or error. So, it is reasonable to substitute actually measured data for the modified values in the measured portion of the ACF. The unmeasured portion of the ACF is left as it is. The ACF thus obtained is a first approximation to the true ACF, which is consistent with both the non-negativity and the measured data.

This process is repeated many times until the convergence is obtained. Figures 2 (c) and 2 (d) are the results when the loop gains are zero and 2.4, respectively. The number of iteration is about 100. As it is clear from Figure 2 (d), both lower sidelobe-level and higher resolution can be achieved.

3. Convergence

Figure 3 shows the convergence of Fourier iteration for the various values of loop gain. The full width at half maximum (FWHM) of a point source is used as an indication of the convergence. Theoretically, the shape of a point source is δ-function and has infinitely

narrow width. In actual simulation experiments, however, discrete Fourier transform is used, so that the FWHM of a point source is always greater than 1. The FWHM calculated from the direct Fourier transform with Bartlett window is indicated by an arrow in Figure 3 for comparison.

The convergence is fast at the initial stage of iteration, but becomes slower at the later stage. Improvement in convergence can be obtained by increasing the loop gain. But, if a value greater than 2.5 is used for the loop gain, the process becomes unstable, and the convergence cannot be assured (see Figure 4). Though a loop gain greater than 2.5 gives good result for specific cases, it is not applicable to the other cases. In the course of the simulation experiments, it was found that a loop gain of 2.0 is safer in many cases.

4. Other examples

We tested several more complicated distributions, which consist of isolated strong point sources, contiguous point sources, extended sources, and "core and halo" type sources. One of these examples is shown in Figure 5.

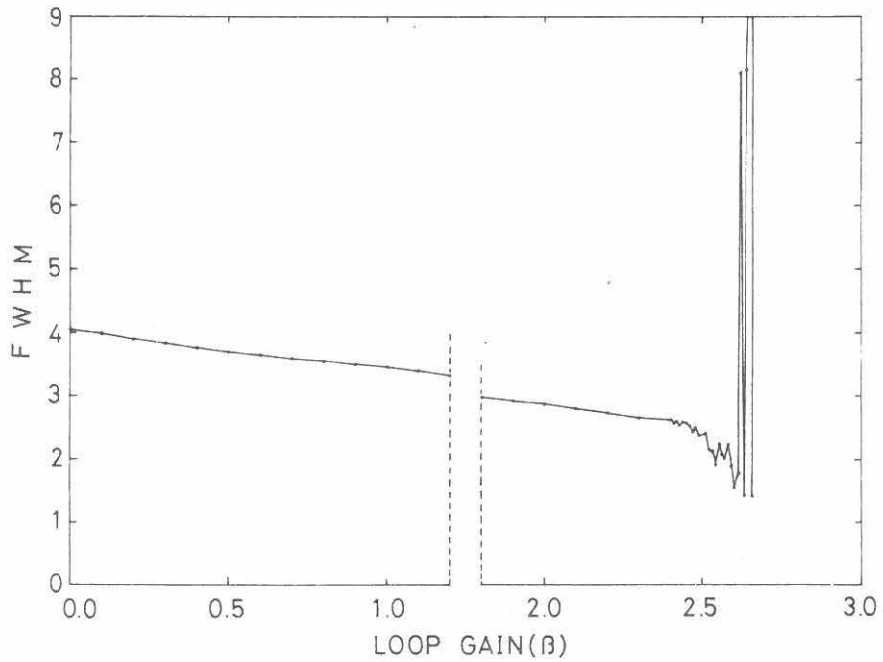

Fig. 4. Effect of the loop gain for the final sharpness of a point source.

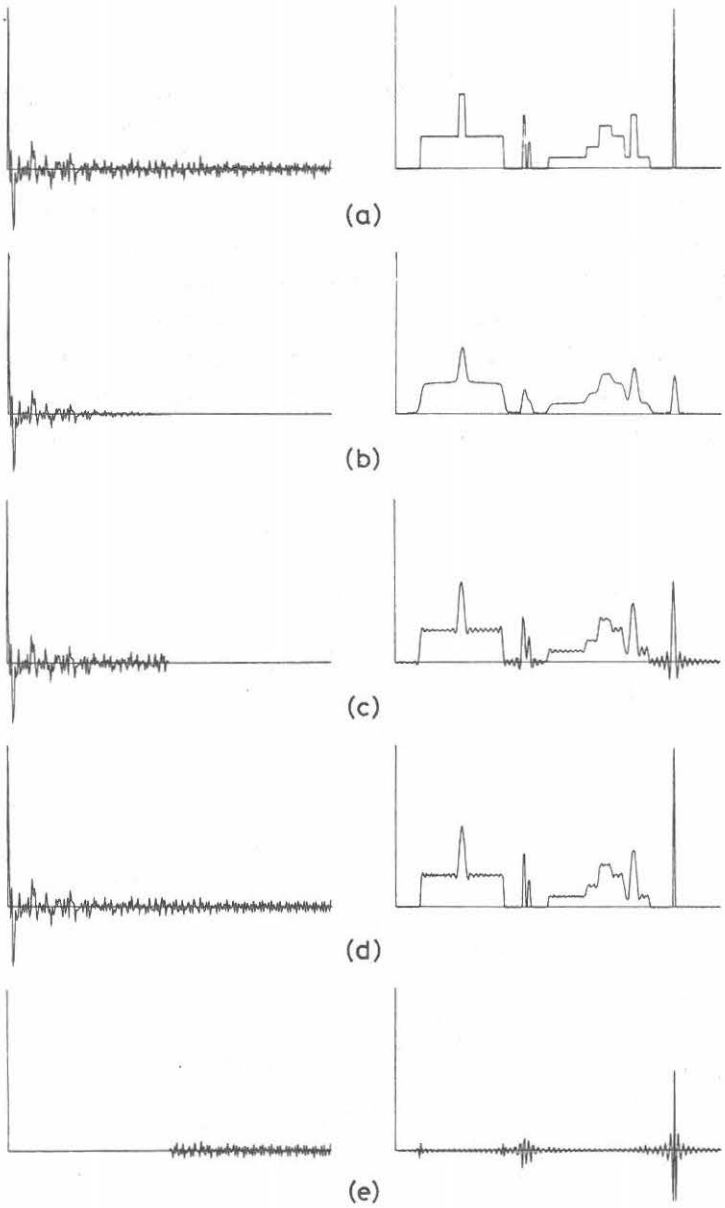

Fig. 5. An example of more complicated distribution.
Left : ACF, right : power spectrum (or image).
(a) A simulated distribution.
(b) Fourier transform after tapering ACF.
(c) Direct Fourier transform after low-pass filtering.
(d) Result of IRT with loop gain, $\beta = 2.4$.
(e) Recovered ACF and it's Fourier transform.

Figure 5 (a) shows the true distribution and it's ACF. The ACF is low-pass filtered intentionally at the half point of the maximum lag as shown in Figure 5 (c). The Fourier transform of the low-pass filtered ACF contains strong sidelobes as is expected, and the resolution is halved. Figure 5 (d) is the result of the IRT, starting with this ACF. The missed portion of the ACF is recovered through this process and the reconstructed distribution is quite similar to the true distribution. The recovered portion of the ACF is shown in Figure 5 (e) with it's Fourier transform. It is clear that the improvement is obtained for both resolution and sidelobes where the steep variations begin from the zero level in the true distribution. The improvement is not so remarkable where the steep variations begin from non-zero levels. This is a weak point of the present method in which the non-negativity is examined around the zero level.

As the computing algorithm is very simple, the extension to the

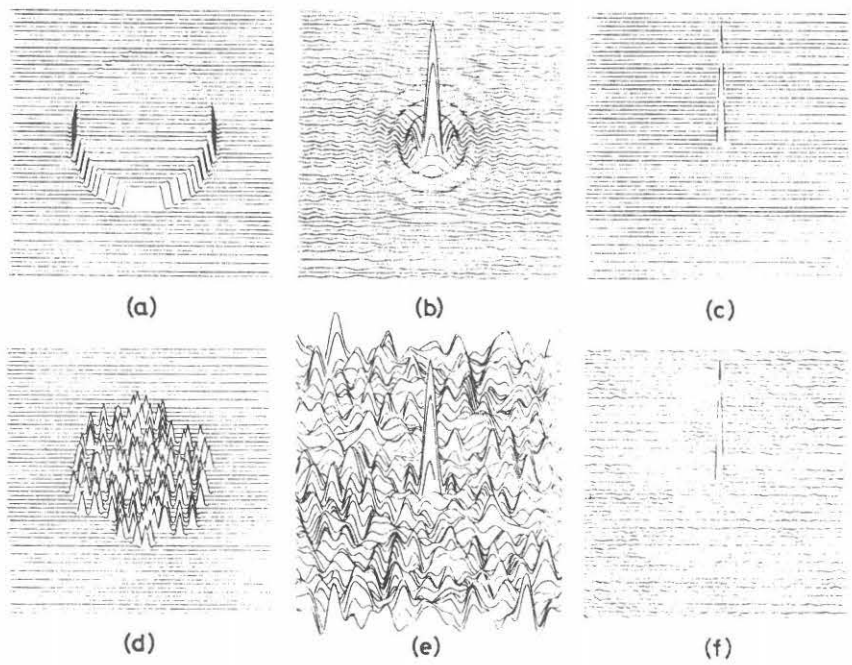

Fig. 6. An example of two-dimensional distribution.
(a) Fully sampled ACF.
(b) Direct Fourier transform of (a).
(c) Result of IRT for (a) with loop gain, $\beta = 2.4$.
(d) An example of sparsely sampled ACF.
(e) Direct Fourier transform of (d).
(f) Result of IRT for (d) with loop gain, $\beta = 2.4$.

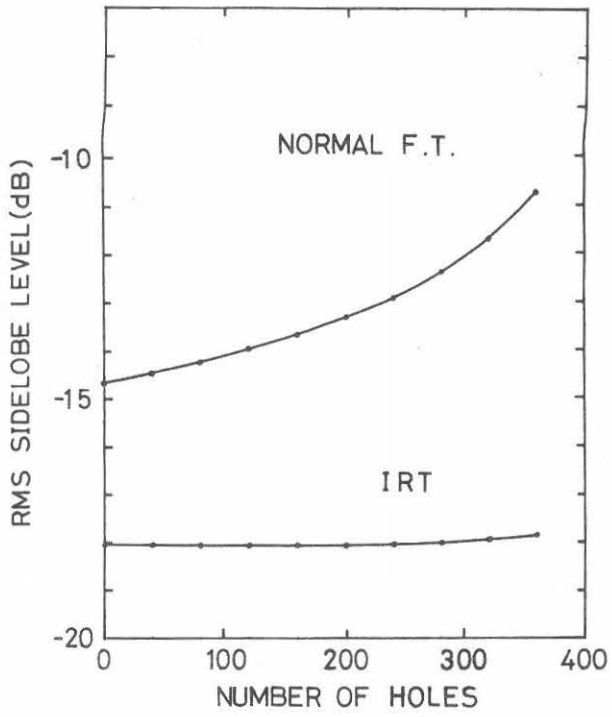

Fig. 7. RMS sidelobe levels for Figs. 6 (e) and (f).

multi-dimensional distribution is straightforward. A two-dimensional example for a single point source is shown in Figure 6, in which comparison is made between the fully sampled ACF and the sparsely sampled ACF. The latter is generated by randomly dropping out the data in the ACF and has only half samples of those for the former. The distribution of holes in the ACF is made almost uniform. Figure 7 shows RMS sidelobe levels against a number of holes in the ACF, for both conventional direct Fourier transform and the IRT. The IRT is effective in case of the large number of holes.

5. Noisy data

In the preceding sections, we neglected the effect of the measurement noise. We tested the case when the ACF is corrupted by random noise. Figure 8 shows an example, in which the standard deviation of the random noise is 10 persent of the maximum value in the ACF. The IRT is still superior to the direct Fourier transform in this case.

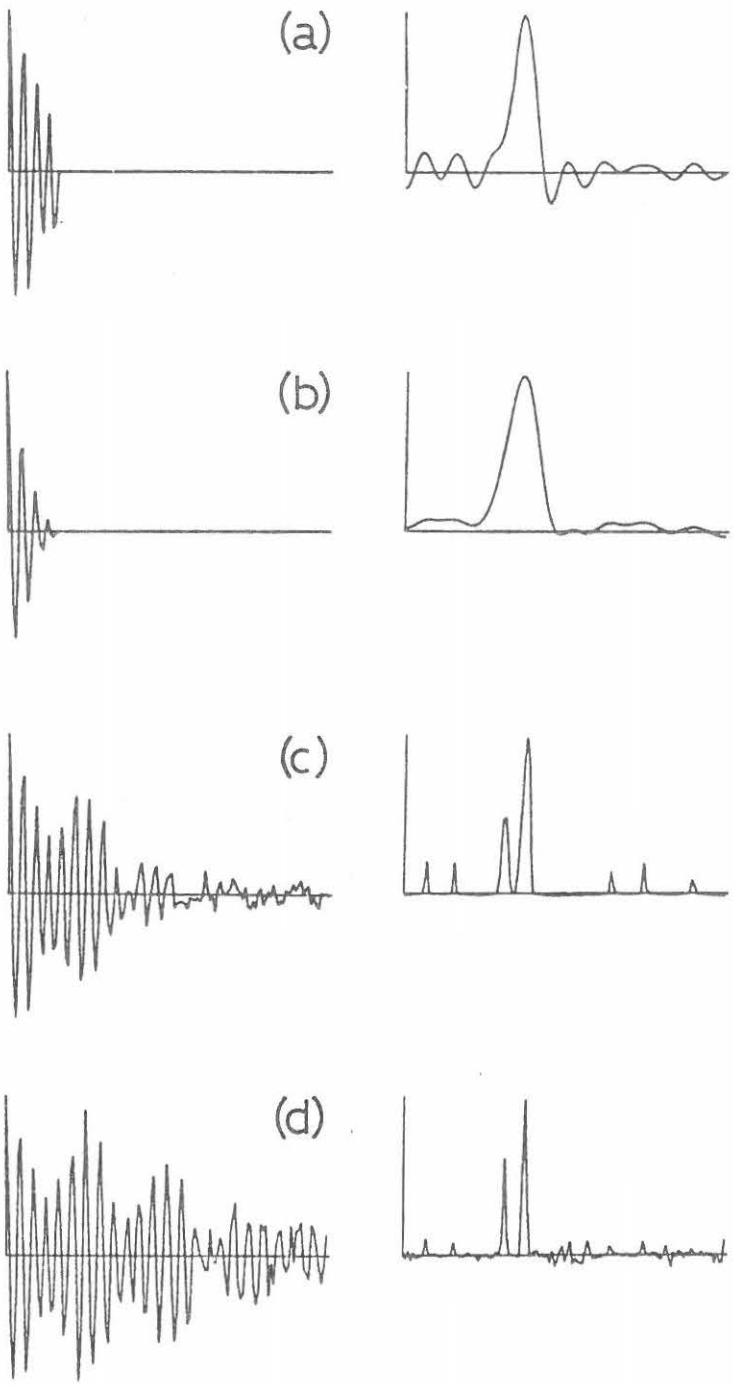

Fig. 8. An example of noisy data (explanation is the same as Fig. 2).

when the signal-to-noise ratio is worse than 10dB, the iteration process becomes unstable for a large loop gain, and spurious distributions appear. It is recommended to use more conservative loop gain for the noisy data.

6. Conclusion

The properties of the IRT incorporated with a non-negativity constraint on the power spectrum (or brightness distribution) are examined by simulation experiments. The non-negativity is assured by multipling all the negative values with some negative factor (loop gain) rather than by replacing by zeros. Improvement is obtained for both resolution and sidelobes when non-zero loop gain is used. The non-negativity is examined around the zero level in the distribution, so that the improvement is remarkable where the steep variations begin from the zero level. Consequently, this method is very sensitive to the DC component of the ACF and measurement noises. We concluded that the effective use of the present method is limited to the case when the distribution consists of isolated point-like sources, and proposed another method of image reconstruction which can be applied to more general cases (Ishiguro and Ishiguro, 1979).

References

Fort, D. N. and Yee, H. K. C.: *A Method of Obtaining Brightness Distributions from Long Baseline Interferometry*, Astron. & Astrophys., 50, 19 (1976).

Högbom, J. A.: *Aperture Synthesis with a Non-Regular Distribution of Interferometer Basalines*, Astron. & Astrophys. Suppl. 15, 417 (1974).

Ishiguro, M. and Ishiguro, M.: *Fitting a Gaussian Model to Aperture Synthesis Data by Akaike's Information Criterion (AIC)*, in *FORMATION OF IMAGES FROM SPATIAL COHERENCE FUNCTIONS IN ASTRONOMY (C. van Schooneveld, ed.)*, D. Reidel Publishing Company (1979).