On a Parameter Estimation Algorithm for MA Time Series *

Petre Stoica and Girish Ganesan
Department of Systems and Control,
Uppsala University,
P. O. Box 27,
S-751 03, Uppsala, Sweden
Tel: +46-18-471 3153
Fax: +46-18-503611

*This work was partly supported by the Senior Individual Grant Program of the Swedish Foundation for Strategic Research (SSF).

Abstract

We take a close look at an existing algorithm for MA(1) parameter estimation. We provide a thorough convergence analysis of that algorithm in the scalar case, discuss its main advantages and disadvantages, present a possible extension of it to the MA(n) case, and compare it with a recent alternative algorithm that appears to be one of the best available choices for MA parameter estimation.

Keywords

MA Parameter Estimation, Convergence Analysis.

1 Introduction and Preliminaries

Consider the following MA(1) time series:

$$y(t) \; = \; e(t) + \theta e(t-1) \quad t = 1, 2, \ldots \tag{1}$$

where $e(t)$ is (unobservable) white noise with mean zero and unknown variance denoted by σ^2, and θ is the MA parameter the estimation of which from the observations of $\{y(t)\}$ is our main problem.

We will consider the following iterative algorithm for estimation of θ, where the subindex i denotes the iteration number:

$\star$ Given $\theta_0 = 0$, for $i = 1, 2, \ldots$ do:

$$\begin{aligned} e_i(t) &= y(t) - \theta_{i-1} e_{i-1}(t-1) \\ \theta_i &= E[y(t)e_i(t-1)]/E[e_i^2(t-1)] \end{aligned} \tag{2}$$

Here E is either the statistical expectation operator (used in most part of the theoretical analysis) or it denotes temporal averaging along the available sample of $\{y(t)\}$ (when we are interested in implementation issues). To distinguish the former from the latter case we use the notation $\tilde{E}$ for the temporal averaging operator:

$$\tilde{E}(.) \; = \; \frac{1}{N} \sum_{t=1}^{N} (.) \tag{3}$$

Note that for $\theta_0 = 0$ and *any* finite sequence $\{e_0(t)\}$ we obtain $e_1(t) = y(t)$ (which is why $\{e_0(t)\}$ was not specified). Hence the first iteration of (2) consists of computing

$$\theta_1 = E[y(t)y(t-1)]/E[y^2(t-1)] \tag{4}$$

which is recognized as the least-squares estimate of the parameter of an AR(1) time series model fitted to $\{y(t)\}$.

The above algorithm was introduced in (Monti, 1996) and (Monti, 1998) for scalar, respectively multivariate MA(1) time series. Here we consider the scalar case to simplify the exposition and make our points as clear as possible, but similar conclusions are expected to hold in the multivariate case as well. We should note that the original proposal of the above algorithm is initialized differently from (2). More exactly, the algorithm in (Monti, 1996; Monti, 1998) starts by computing θ_1 in (2) using an *arbitrary* white noise sequence $\{e_1(t)\}$, which looks like a curious choice at first sight. However a moment of reflection shows that this initialization will yield $\theta_1 = 0$ (in the asymptotic case; or $\theta_1 = O(N^{-1/2})$ in the finite (but large) sample case). Hence if we used this initialization we would gain nothing but lose one iteration compared with directly computing θ_1 as in (4).

Our goal in this report is to answer a number of questions concerning the algorithm in (2), such as:

- What are the convergence properties of the algorithm ? In (Monti, 1996; Monti, 1998) only the convergence of (2) for the specific initialization $\theta_0 = 0$ and for $|\theta| < 1$ was established. The convergence properties for other initializations were unknown. Here we show that the algorithm converges to θ for many other initializations but not for all. Furthermore in (Monti, 1998) it was stated that if $|\theta| > 1$ then (2) does not converge. In fact it does converge but to $1/\theta$, as we show later.

- The convergence of (2) in the finite sample case (when E in (2) is replaced by $\tilde{E}$) was not clarified in (Monti, 1996; Monti, 1998). Since the analysis of (2) in such a case appears to be hard, we recourse to Monte-Carlo simulations to show that the implementable version of (2) may have convergence problems.

- The extension of the algorithm to MA(n) time series (with $n > 1$) is straightforward. (See Section 4 and also (Monti, 1996)). However the extension of the convergence analysis of the MA(1) case is not. We present the extended algorithm and illustrate its performance by means of a number of Monte-Carlo simulations. We also compare the performance of the extended algorithm with that of a recently proposed MA parameter estimator (Stoica et al., 2000) that appears to have benchmark properties.

2 Convergence Analysis

Since

$$e_i(t) = y(t) - \theta_{i-1}y(t-1) - \theta_{i-1}\theta_{i-2}y(t-2)\ldots$$

it follows readily (assuming that $|\theta_{i-1}| < \infty$ etc.) that

$$E[y(t)e_i(t-1)] = E[y(t)y(t-1)] = \sigma^2\theta \quad \text{(for any } i \geq 1\text{)} \tag{5}$$

Next we use the definition of θ_{i-1} along with (5) to write:

$$\begin{aligned} E[e_i^2(t)] &= E\left[(y(t) - \theta_{i-1}e_{i-1}(t-1))y(t)\right] \\ &= \sigma^2(1+\theta^2) - \theta_{i-1}\sigma^2\theta \end{aligned} \tag{6}$$

Inserting (5) and (6) in (2) yields the following first-order (non-linear) recursion for θ_i:

$$\theta_i = \frac{\theta}{1+\theta^2 - \theta\theta_{i-1}} \tag{7}$$

For $\theta = 0$ the above recursion converges in one iteration to θ regardless of the initialization $|\theta_0| < \infty$. Hence we let $\theta \neq 0$. Assume that

$$0 < |\theta| < 1 \tag{8}$$

and let

$$z_i = \theta(\theta_i - \theta) \tag{9}$$

(the case of $|\theta| > 1$ will be discussed shortly). A straightforward calculation shows that (7) can be rewritten in terms of $\{z_i\}$ as:

$$z_i = \theta^2 \frac{z_{i-1}}{1 - z_{i-1}} \tag{10}$$

Our first important observation is that in the case of $|\theta| > 1$ we can use $\tilde{\theta} = 1/\theta$ and $\tilde{z}_i = \tilde{\theta}(\theta_i - \tilde{\theta})$ to rewrite (7) as

$$\tilde{z}_i = \tilde{\theta}^2 \frac{\tilde{z}_{i-1}}{1 - \tilde{z}_{i-1}} \tag{11}$$

which has the same form, and hence the same convergence properties, as (10). In particular this means that *if the recursion in (7) converges to θ for $|\theta| < 1$ then it will converge to $1/\theta$ for $|\theta| > 1$.* This observation corrects a statement in (Monti, 1998) where it was claimed that the recursion under consideration does not converge if $|\theta| > 1$. In the sequel we assume that (8) holds (i.e. $|\theta| < 1$), as the case of $|\theta| > 1$ can be similarly analyzed using (11).

The stationary points of (10) are the solutions to the equation:

$$z_\infty = \theta^2 \frac{z_\infty}{1 - z_\infty} \tag{12}$$

Hence there are two stationary points:

$$z_\infty = 0 \;\Leftrightarrow\; \theta_\infty = \theta \tag{13}$$

and

$$z_\infty = 1 - \theta^2 \;\Leftrightarrow\; \theta_\infty = 1/\theta \tag{14}$$

In the vicinity of 0, the recursion (10) is approximately given by $z_i = \theta^2 z_{i-1} + o(|z_{i-1}|)$ and hence $\theta_\infty = \theta$ is a *stable stationary point* (under (8)). In the vicinity of $1 - \theta^2$, the approximate linear recursion is given by:

$$z_i - (1 - \theta^2) = (1/\theta^2)[z_{i-1} - (1 - \theta^2)] + o(|z_{i-1} - (1 - \theta^2)|)$$

which shows that $\theta_\infty = 1/\theta$ is *an unstable stationary point* (under (8); of course if we had $|\theta| > 1$ then $\theta_\infty = \theta$ would be unstable whereas $\theta_\infty = 1/\theta$ would be stable, as implied by the discussion around (11).

To analyze the convergence of (7) to $\theta_\infty = \theta$ we consider two cases:

Case 1

$$z_0 < 0 \text{ or } \; z_0 > 1 \tag{15}$$

Note that for $\theta_0 = 0$ we get $z_0 = -\theta^2 < 0$. Hence the specific initialization in (2) is a most convenient way to satisfy (15)! It is shown in the sequel that *the recursion (2) converges exponentially to θ when the initial estimate satisfies (15).*

To prove the previous claim note that $z_0 > 1$ implies $z_1 < 0$, and hence it is sufficient to consider (10) for $z_0 < 0$. It is easily checked that

$$z_0 < 0 \;\Rightarrow\; z_i < 0 \quad \text{for } i = 1, 2, \ldots \tag{16}$$

Using this fact we immediately obtain the inequality:

$$|z_i| < \theta^2 |z_{i-1}| \tag{17}$$

which establishes the exponential convergence to $z_\infty = 0$. Apparently, the closer $|\theta|$ is to 1, the slower the convergence rate.

Case 2

$$0 < z_0 < 1 \tag{18}$$

Since it is so easy to find an initialization for which the recursion converges to the desired stationary point (see Case 1 above), we do not discuss the case of (18) in detail. We only point out that the recursion initialized as in (18) does *not* necessarily converge to θ. As an example, for $z_0 = 1 - \theta^2$ (which satisfies (18) for $|\theta| < 1$) the recursion gets stuck in the undesired stationary point $1/\theta$.

Example 1 *In this example we consider an MA(1) time-series with true parameter $\theta = 0.5$. The driving white noise is Gaussian distributed with mean zero and variance $\sigma^2 = 0.1$. The algorithm in (2) is used to compute the MA parameter. One hundred data samples are used. The estimated value is plotted against the number of iterations in Figure 1. The results shown are the average over 100 Monte-Carlo runs. The mean and variance of the estimated parameter, after 100 iterations, are: Mean= 0.4994, Variance=0.0114. From Figure 1 it can be seen that in some 10 iterations the estimate converges to a stable point, which is close to the true value.*

Example 2 *In this example the conditions are the same as in Example 1 except for the fact that $\theta = 2.5$. The results of the estimation algorithm averaged over 100 Monte-Carlo runs are shown in Figure 2. The mean and variance of the estimated parameter, after 100 iterations, are: Mean= 0.3902, Variance=0.0106. From the figure it can be seen that in about 10 iterations the estimate converges to a stable point that is close to $1/\theta = 0.4$.*

3 Convergence in Finite Samples

The convergence of (10) to θ does not necessarily imply the convergence of (2) with E replaced by $\tilde{E}$ to a point close to θ, not even for $N >> 1$. Indeed the small errors of order $O(N^{-1/2})$ made at each iteration when replacing E by $\tilde{E}$ may accumulate destructively as the iterations proceed. Since the analysis of (2) in the finite-sample case is hard we recourse to simulations to show the convergence problems that may occur in such a case.

Example 3 *In this example we set $\theta = 0.95$. The driving noise is the same as in the previous examples. One hundred data samples are used. Out of the 100 Monte-Carlo runs conducted it was seen that in roughly 5% of the runs even after 10000 iterations the algorithm did not converge to a stable point, but oscillated. For one such run θ_i has been plotted for the iterations 1 to 500 and 9500 to 10000 in Figures 3 and 4. In Figure 5 θ_i has been plotted for the iterations 9500 to 10000 for another such run. Thus in the finite sample case the algorithm may not converge.*

4 Extension to MA(n) and an Alternative Algorithm

Consider an MA(n) time series:

$$y(t) = e(t) + c_1 e(t-1) + \ldots + c_n e(t-n) \tag{19}$$

Let $\theta = [c_1, \ldots, c_n]^T$ and let θ_i denote the estimate of the parameter vector at iteration i. The direct extension of (2) to the present case yields the algorithm below (see also (Monti, 1996)):

$\star$ Given $\theta_0 = 0$, for $i = 1, 2, \ldots$ do:

$$\begin{aligned} e_i(t) &= y(t) - \Phi_{i-1}^T(t-1)\theta_{i-1} \\ \Phi_i(t-1) &= [e_i(t-1), \ldots, e_i(t-n)]^T \\ \theta_i &= \left[\tilde{E}\{\Phi_i(t-1)\Phi_i^T(t-1)\}\right]^{-1}\left[\tilde{E}\{\Phi_i(t-1)y(t)\}\right] \end{aligned} \tag{20}$$

Note that, exactly as in the MA(1) case, the first iteration of the above algorithm computes the least squares estimate of an AR(n) model fitted to $\{y(t)\}$. This is somewhat similar to what is done in the Durbin algorithm for MA parameter estimation (see, e.g, (Durbin, 1959; Stoica and Moses, 1997)) in the first step of which, however, an AR model of an order (much) larger than n is fitted to $\{y(t)\}$.

Since the analysis of the convergence properties of (20) (even with $\tilde{E}$ replaced by E) appears to be difficult, we make use of Monte-Carlo simulations to illustrate what kind of performance one could expect from (20). In the simulations we also compare (20) with the MA parameter estimation algorithm recently introduced in (Stoica et al., 2000). The latter algorithm is rather unique in that it produces consistent parameter estimates of an MA(n) in a time that is only a polynomial function of n, namely $O(n^4)$. The basic implementation block of that algorithm is a semidefinite program solver, which computes the MA parameters in a computationally efficient and reliable manner that is not plagued by any convergence problems or the like.

Example 4 *In this example we consider an MA(3) time series with the following parameters:*

$$c_1 = -1.5000,\ c_2 = 1.2000,\ c_3 = -0.4480$$

and a white Gaussian driving noise with mean zero and variance $\sigma^2 = 0.1$. *The sample size is* $N = 200$. *A Monte-Carlo simulation is performed over 100 independent data sets. For the algorithm in (20) 400 iterations are performed. (Fewer iterations produced less accurate estimates). The Overparametrized Signal (OS) algorithm proposed in (Stoica et al., 2000) is also used to estimate the MA parameters. From the estimated parameters the zeros of the MA polynomial are calculated. The results are shown in Figures 6 and 7. From these figures it can be seen that the iterative algorithm in (20) has a much larger mean square error than that of the Overparametrized Signal method. Additionally , the number of flops (as counted in MATLAB) required by (20),* 7240×10^3, *is much larger than the* 104×10^3 *flops needed by the OS algorithm. Finally observe that the MA zeros estimated by (20) tend to cluster around two quite distinct sets of zeros. This may be indicative of the fact that in this case there are two stable convergence points.*

5 Conclusions

In this report we considered the algorithm for MA parameter estimation in (Monti, 1996; Monti, 1998). A thorough convergence analysis was performed in the MA(1) case. We also discussed the possible advantages and disadvantages of the method. Our conclusion is that this method does not seem to stand comparison with the much more accurate and faster MA estimation algorithm introduced in (Stoica et al., 2000).

References

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika*, 46, 306–316.

Monti, A. C. (1996). A new preliminary estimator for MA(1) models. *Computational Statistics & Data Analysis*, 21, 1–15.

Monti, A. C. (1998). A proposal for estimation of the parameters of multivariate moving-average models. *Journal of Time Series Analysis*, 19(2), 209–219.

Stoica, P., McKelvey, T. and Mari, J. (2000). MA-estimation in polynomial time. *IEEE Transactions on Signal Processing*, 48(7), 1999–2012.

Stoica, P. and Moses, R. (1997). *Introduction to Spectral Analysis.* NJ, USA: Prentice Hall.

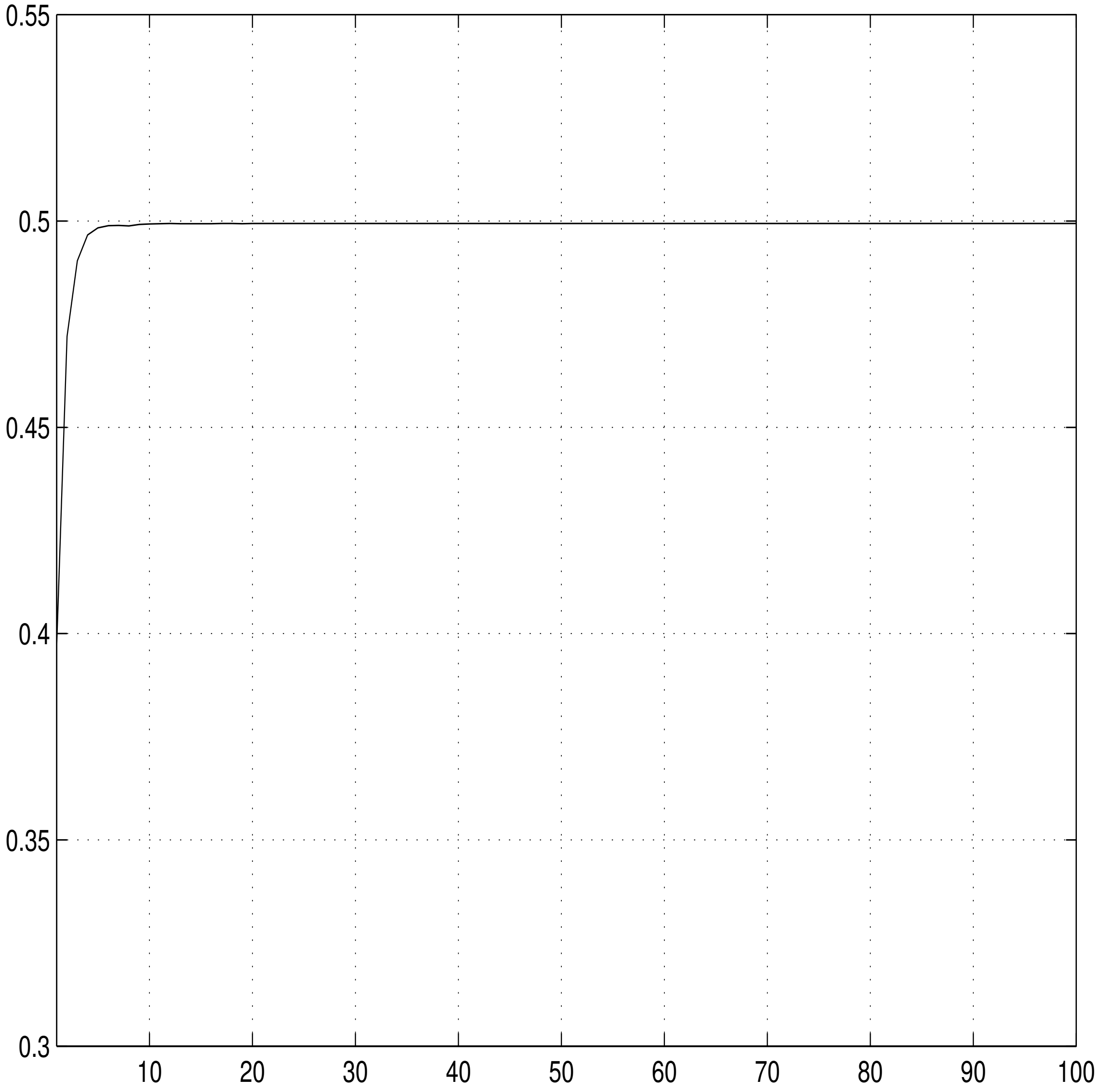

Figure 1: Results of Example 1. The estimated parameter in the ith iteration, θ_i, is plotted against the number of iterations. The true value is $\theta = 0.5$.

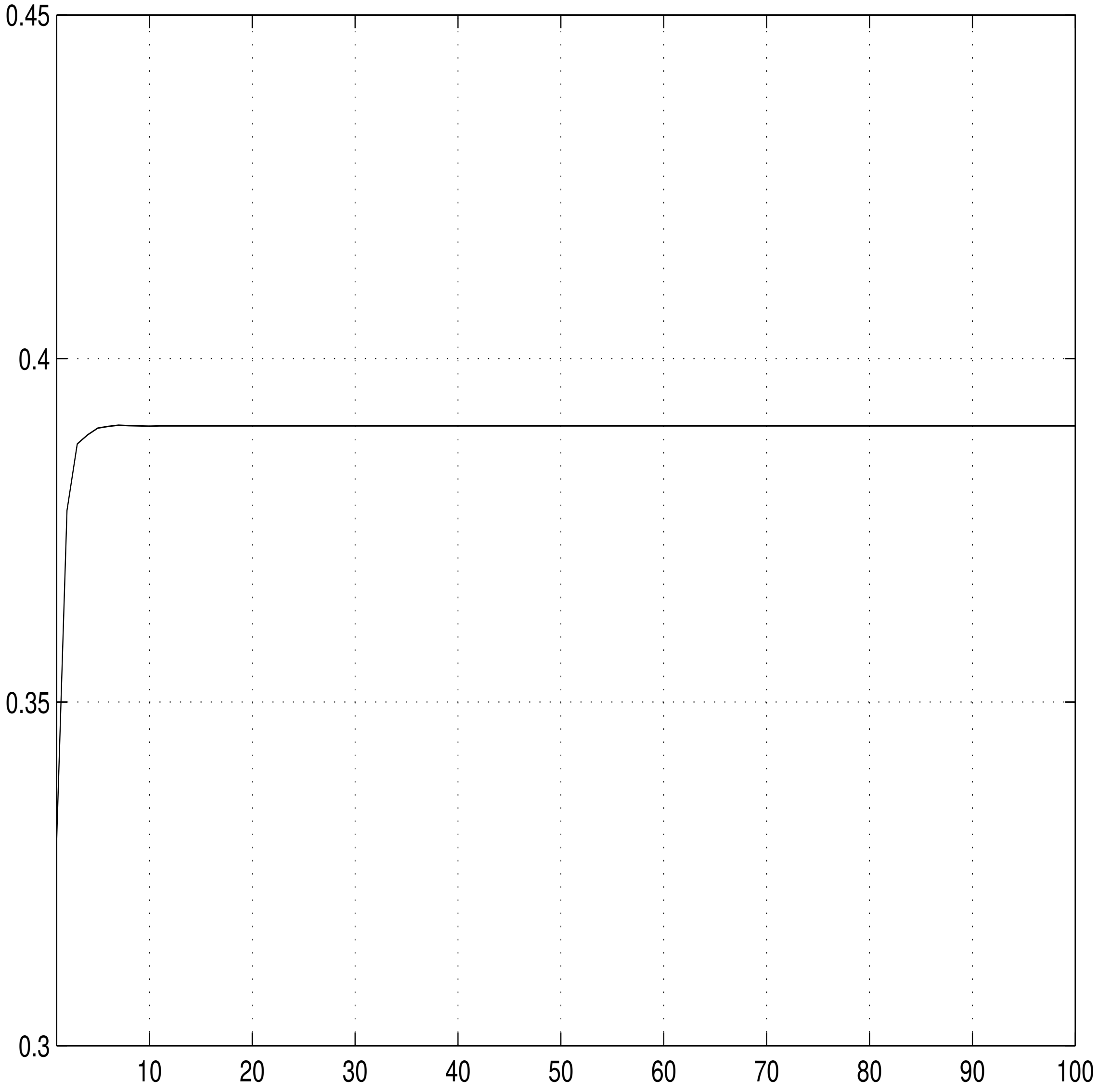

Figure 2: Results of Example 2. The estimated parameter in the ith iteration, θ_i, is plotted against the number of iterations . The true value is $\theta = 2.5$. Since $|\theta| > 1$ the iteration converges to a point close to $1/\theta = 0.4$.

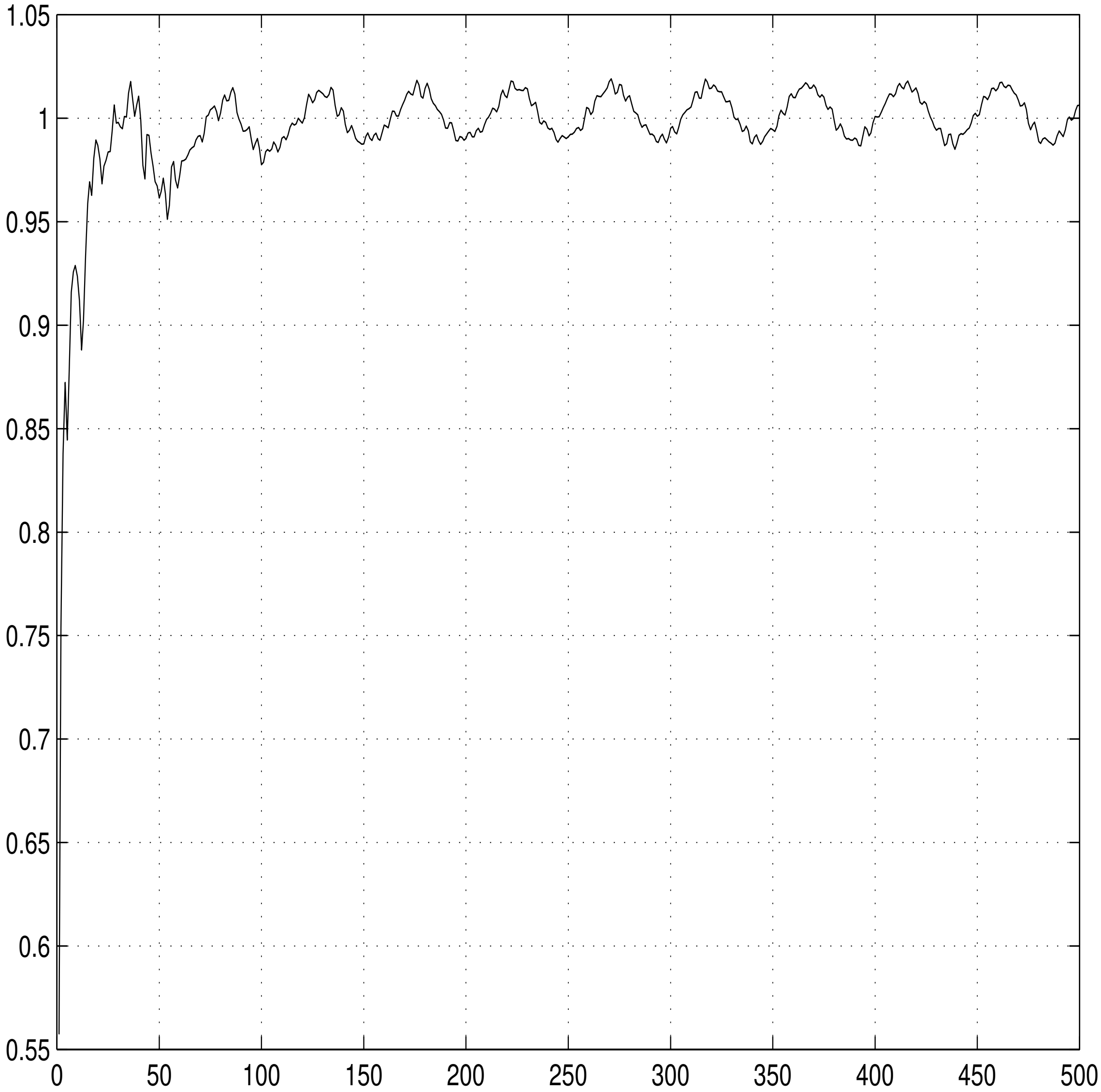

Figure 3: Results of Example 3 for a run in which the algorithm did not converge. The estimated parameter in the ith iteration, θ_i, is plotted against the number of iterations. The true value is $\theta = 0.95$.

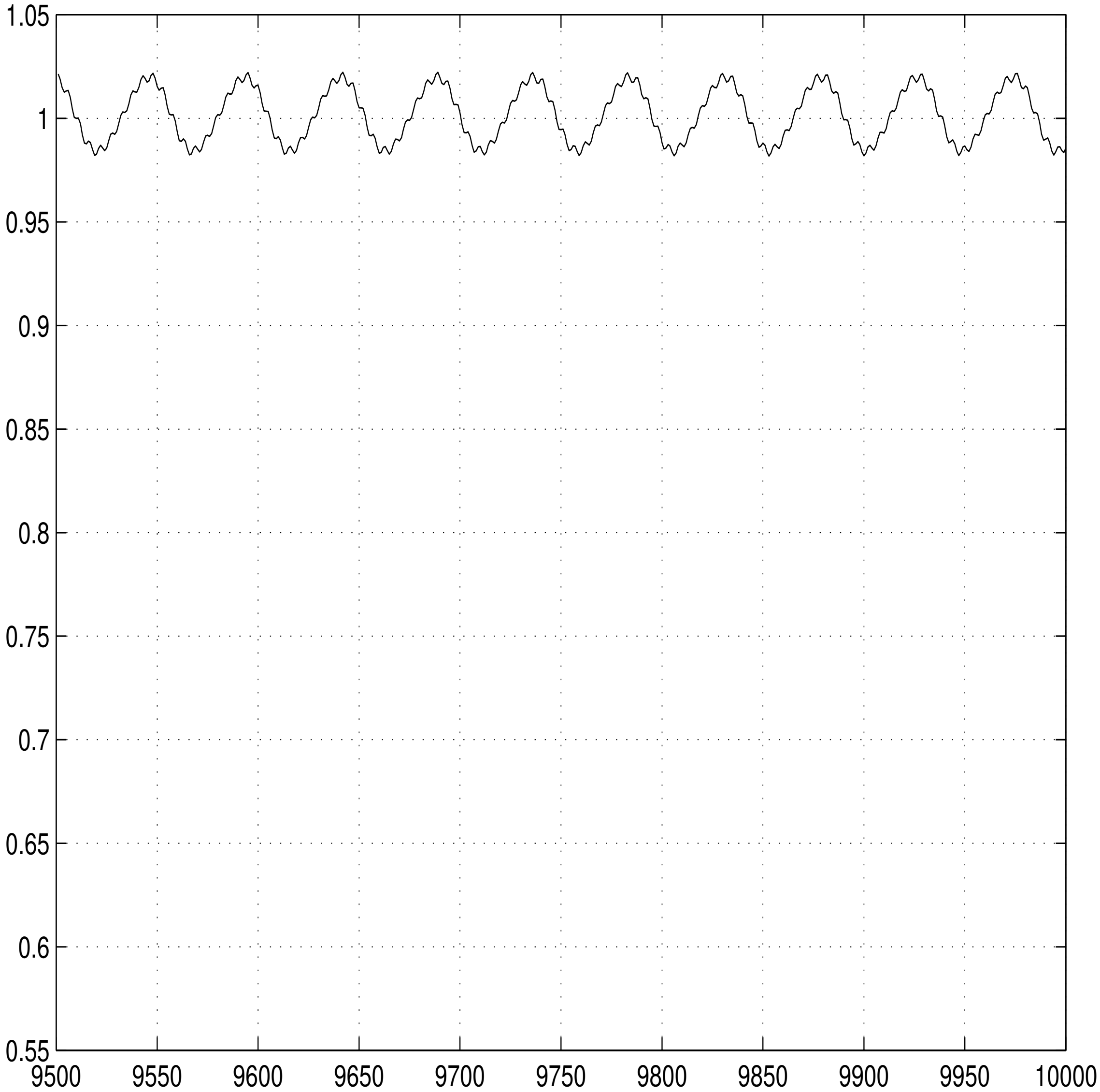

Figure 4: Continuation of Figure 3 for iterations 9500 to 10000.

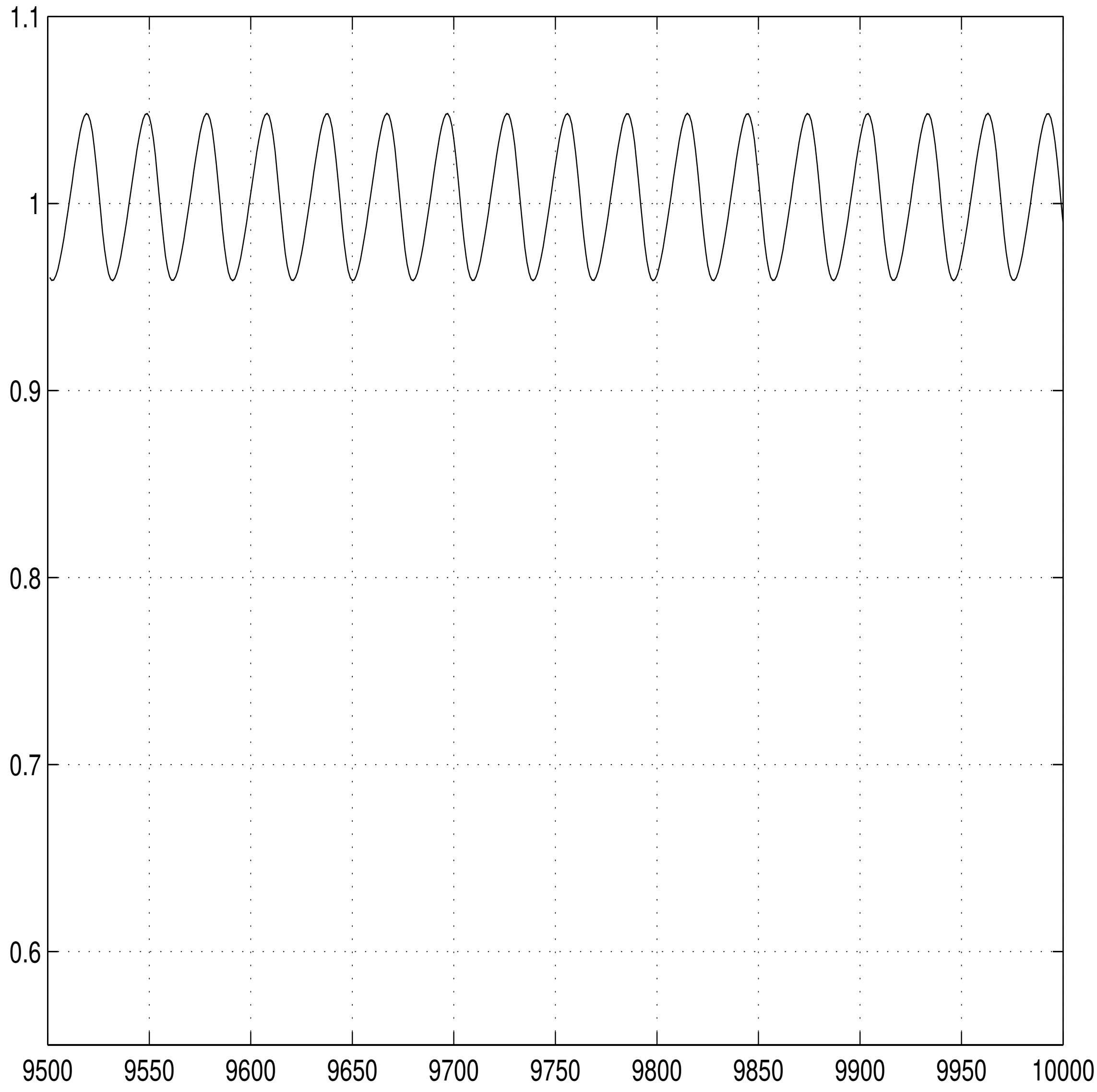

Figure 5: Results of Example 3 for another run in which the algorithm did not converge. The estimated parameter in the ith iteration, θ_i, is plotted against the number of iterations. The true value is $\theta = 0.95$.

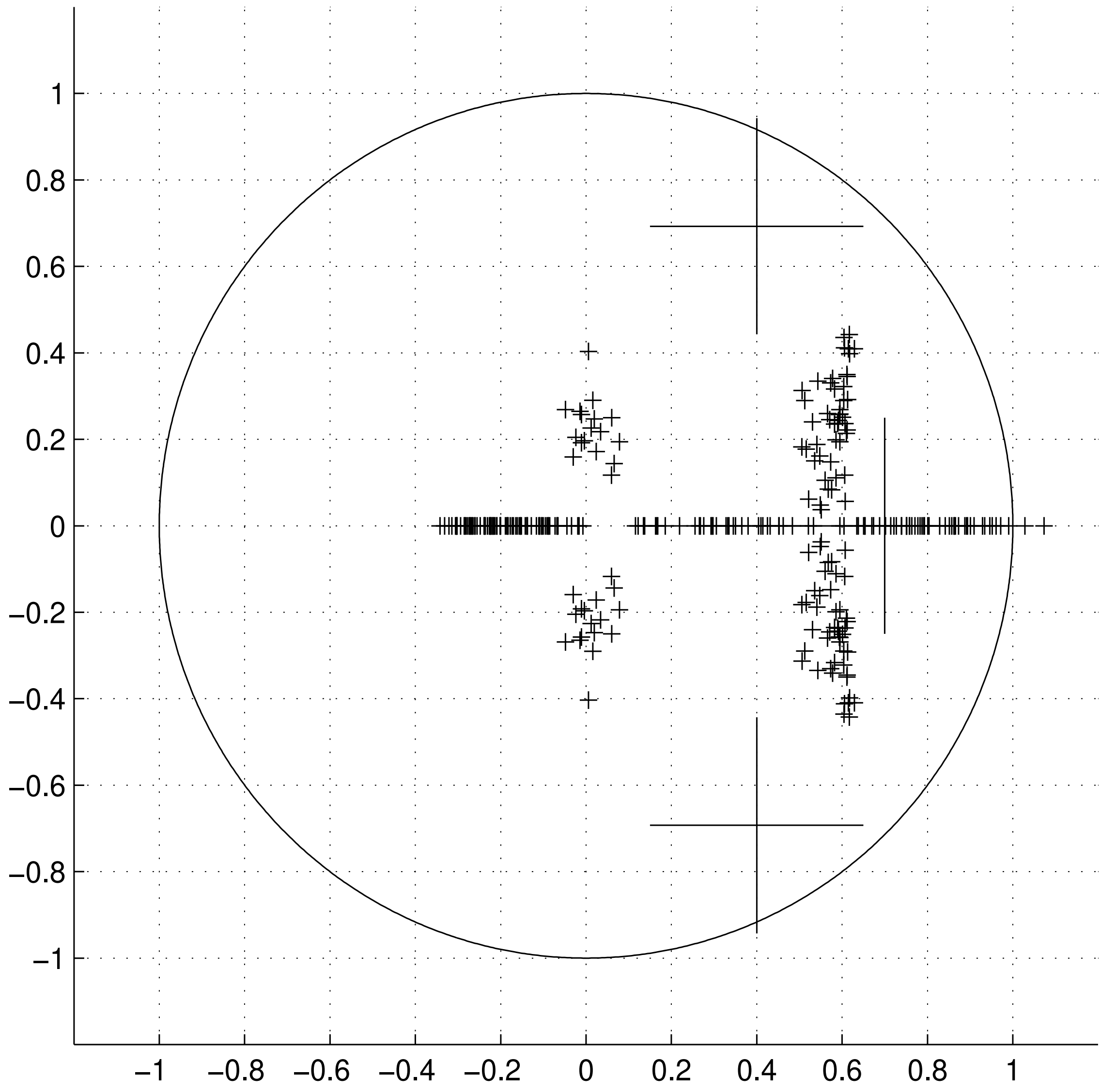

Figure 6: Results of Example 4 for the algorithm in (20). Estimated MA zeros are shown in a superimposed manner for 100 independent data sets. The large crosses indicate the true zero locations.

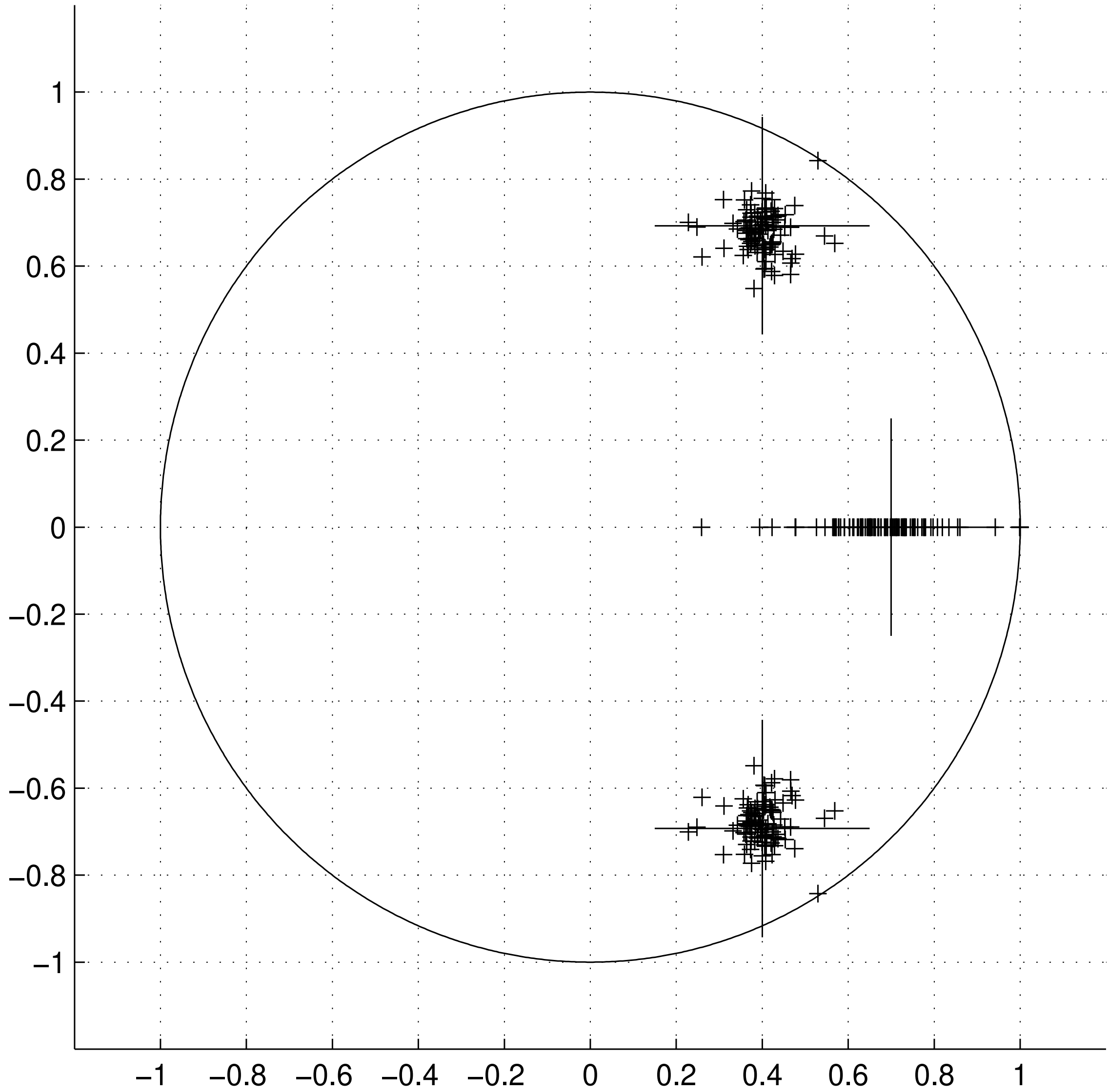

Figure 7: Results of Example 4 for the OS algorithm Estimated MA zeros are shown in a superimposed manner for 100 independent data sets. The large crosses indicate the true zero locations.