Capon Estimation of Covariance Sequences *

Hongbin Li † Petre Stoica ‡ Jian Li †

Abstract

Estimating the covariance sequence of a wide-sense stationary process is of fundamental importance in digital signal processing (DSP). A new method, which makes use of Fourier inverting the Capon spectral estimates and is referred to as the *Capon method*, is presented in this paper. It is shown that the Capon power spectral density (PSD) estimator yields an equivalent autoregressive (AR) or autoregressive moving-average (ARMA) process, hence the *exact* covariance sequence corresponding to the Capon spectrum can be computed in a rather convenient way. Also, without much accuracy loss, the computation can be significantly reduced via an approximate Capon method that utilizes the fast Fourier transform (FFT). We show by using a variety of ARMA signals that the Capon covariance estimates are generally better than the standard sample covariance estimates and can be used to improve the performances in the DSP applications that are critically dependent on the accuracy of the covariance sequence estimates.

*This work was supported in part by the National Science Foundation Grant MIP-9308302, the Advanced Research Project Agency Grant MDA-972-93-1-0015, and the Swedish Research Council for Engineering Sciences (TFR).

†Hongbin Li and Jian Li are with the Department of Electrical Engineering, University of Florida, Gainesville, FL 32611, USA.

‡P. Stoica is with the Systems and Control Group, Department of Technology, P.O.Box 27, Uppsala University, S–751 03, Uppsala, Sweden.

1 Introduction

Covariance sequence estimation is a ubiquitous task in digital signal processing. A standard technique for estimating the covariance sequences is the so-called the *standard sample covariance estimator*. The standard covariance estimates are consistent provided that the given signals are ergodic to the second order. However, there is a major concern of using the standard estimator due to the unrealistic *windowing* of the observed data it assumes; that is, it assumes that the data beyond the observed duration either is zero or repeats itself periodically. Partly for this reason, there have been several attempts in the literature to derive more accurate covariance estimates than the standard ones. A notable example is the approach based on the Burg autoregressive (AR) spectral estimator. However the so-obtained covariance estimator turned out to be even less accurate than the standard sample covariance estimator [1]. To be more exact, the Burg approach was shown to have larger variances than the standard method. Another approach, which has generated a whole new research direction, relies on the maximum likelihood (ML) principle [2]. However, the ML estimation of covariance sequences is a computationally involved problem and there are several theoretical questions about its solution that are yet to get a satisfactory answer. Apparently, there exist no compelling alternatives to the standard method that can be recommended for general use.

In this paper we present a new method, namely the *Capon method*, for covariance sequence estimation. The Capon method obtains the covariance sequence estimates by Fourier inverting the Capon power spectral density (PSD) estimates. There are basically two Capon PSD estimators, referred to as Capon-1 [3] [4] and Capon-2 [5] in this paper. We find that, while Capon-2 is capable of finer spectral resolution around the peaks of a spectrum, it is generally a *globally* poorer spectral estimator than Capon-1. We hence concentrate our interest on Capon-1 for covariance sequence estimation in this paper. Since the Capon spectra, i.e., the PSD estimates, are shown to be equivalent to AR or autoregressive moving-average (ARMA) spectra, the inversion procedure for computing the *exact* covariance sequences corresponding to the Capon spectra can be implemented in a rather convenient way. (Note that the calculation of the covariance sequences corresponding to the Capon spectra is an interesting problem by itself.) We also present an FFT-based approximate method to compute the covariance sequences from the Capon spectra. It has been found that the approximate method provides covariance estimates that are almost identical with those obtained by the exact method, while the computational complexity is greatly reduced.

Since our primary interest is to apply the Capon method as well as the standard approach to ARMA signals, a few typical ARMA signals having a variety of pole and zero locations are studied in our numerical examples. The studies show that considerable improvements are attained by the new Capon method.

The Capon covariance estimation method can be readily used in many applications. One important class is the ARMA spectral estimation. Since most ARMA spectral estimators rely on the Yule-Walker equations to determine the AR coefficients, it may be expected that the better the covariance estimates used, the more accurate the AR coefficient estimates yielded. We examine how the Capon covariance estimates can be used with the *overdetermined modified Yule-Walker (OMYW) method* [6] [7] to compute more accurate AR coefficients. We find that the performances of the usage are critically dependent on the pole and zero locations and generally better AR coefficient estimates are obtained by using the Capon covariance estimates than by the standard ones. Another application discussed in this paper is the moving-average (MA) model order determination by making use of the Capon and the sample covariance estimates, where we

find that the former performs better than the latter.

2 Standard Covariance Estimator and Outlook

With no other assumptions made on the signal under study, except for assuming the second order ergodicity, there are two ways to obtain the standard sample covariances of the signal, namely, the *biased* and the *unbiased* covariance estimators. However, the biased covariance estimator is more commonly used since it provides smaller mean-squared errors (MSE) than the unbiased one and guarantees the covariance estimates to be positive semidefinite [8].

The biased sample covariance estimator of a wide-sense stationary signal with zero-mean has the form

$$\hat{r}(k) = \frac{1}{N} \sum_{n=1}^{N-k} y^*(n) y(n+k), \quad k = 0, 1, \ldots, K. \tag{1}$$

where $\{y(n)\}_{n=1}^{N}$ are the observed data samples, N is the number of samples, $\hat{r}(k)$ denotes the estimate of the covariance function $r(k)$, K is the largest lag desired ($0 \leq K \leq N-1$), and $(\cdot)^*$ denotes the complex conjugate. Note that (1) is asymptotically unbiased.

The estimator given in (1) is consistent if the ergodicity assumption is satisfied. A study on whether the standard sample covariance estimator is also asymptotically statistically efficient, i.e., whether it asymptotically achieves the Cramér-Rao bound (CR bound), has been undertaken in [9]. Let $y(t)$ be an ARMA(p, q) signal. If $p \geq q$, then the sample covariance estimate $\hat{r}(k)$ is asymptotically statistically efficient if and only if $0 \leq k \leq p - q$; in particular, for AR processes of order p, $\hat{r}(k)$ is asymptotically efficient for $0 \leq k \leq p$, but inefficient for all other k. If $p < q$, none of $\hat{r}(k)$ is asymptotically efficient; in particular, none of $\hat{r}(k)$ is asymptotically efficient for an MA process.

It is known that the sample covariance sequence $\{\hat{r}(0), \ldots, \hat{r}(N-1), 0, 0, \ldots\}$ and the data periodogram constitute a Fourier transform pair. It is also known that the periodogram is a statistically inefficient (in fact inconsistent) estimator of the PSD [8]. This observation suggests that better covariance estimators might be obtained by Fourier inverting better PSD estimators. However, this is not necessarily so. Briefly stated, the reason is that the Fourier transform and the inverse Fourier transform are integral transforms and hence small errors in one domain may be associated with large errors in the other domain or vice versa. The fact that the covariance estimates in (1) are consistent whereas the periodogram is not illustrates this observation. Also note that, while the Burg estimate of the PSD is typically more accurate than the periodogram, the corresponding Burg estimate of the covariance sequence is generally poorer than (1) [1].

In spite of the fact briefly discussed above, in the following we consider estimating the covariance sequences by inverting a PSD estimate that is often much more accurate than the periodogram, namely the Capon PSD estimate. Like the periodogram, no model is assumed in the Capon PSD estimator, which makes it more robust than the parametric estimators in many situations. Although it has lower spectral resolution than the AR spectral estimator, it generally exhibits less variance than the latter [4] [8]. Another reason that we consider the Capon PSD estimator is that it does not exhibit the so-called *correlation matching property* [8] [10]; that is, the inverse Fourier transform of the Capon PSD estimates does not yield the same covariance sequences used to obtain the Capon PSD estimates. This fact allows us to obtain a new covariance estimator from the Capon spectra.

3 Capon Spectral Estimator

The Capon spectral estimator belongs to the class of filterbank approaches. Filterbank approaches to power spectral estimation are generally implemented as follows: (a) the observed signal is passed through a bandpass filter with varying center frequency ω (the *steering frequency*); (b) the output power in the filter's passband is measured; (c) the power spectral estimates are obtained by dividing the measured power by the equivalent bandwidth of the filter. Hence from the viewpoint of the filterbank analysis, the spectral estimation is a problem of filter design subject to some specific constraints [5]. The constraints adopted in the Capon method are that the signal at the current steering frequency ω is passed undistorted (with unit gain) and that the output power of the overall frequency domain is minimized. It has been shown that by choosing such constraints, the Capon filter is actually a matched filter [11] [12].

Consider an $(M+1)$-tap finite impulse response (FIR) filter given by

$$\mathbf{h}^H = \left[\begin{array}{cccc} h_0 & h_1 & \dots & h_M \end{array} \right], \tag{2}$$

where $(\cdot)^H$ denotes the conjugate transpose. The output of the filter, at time n, when the input is the raw data sequence $\{y(n)\}$, is given by

$$y_F(n) = \sum_{m=0}^{M} h_m y(n-m) = \mathbf{h}^H \left[\begin{array}{c} y(n) \\ y(n-1) \\ \vdots \\ y(n-M) \end{array} \right] \triangleq \mathbf{h}^H \mathbf{y}(n). \tag{3}$$

Let $\mathbf{R}$ denote the covariance matrix of the data vector $\mathbf{y}(n)$; that is, $\mathbf{R} = E\left\{\mathbf{y}(n)\mathbf{y}^H(n)\right\}$, where $E\{\cdot\}$ denotes the expectation operator. Then the power of the filter output can be written as

$$E\left\{|y_F(n)|^2\right\} = \mathbf{h}^H \mathbf{R} \mathbf{h}. \tag{4}$$

The filter frequency response is given by

$$H(\omega) = \sum_{m=0}^{M} h_m e^{-i\omega m} = \mathbf{h}^H \mathbf{a}(\omega), \tag{5}$$

where $\mathbf{a}(\omega)$ is the *steering vector* defined by

$$\mathbf{a}(\omega) = \left[\begin{array}{cccc} 1 & e^{-i\omega} & \dots & e^{-iM\omega} \end{array} \right]^T, \tag{6}$$

with $(\cdot)^T$ denoting the transpose.

The Capon method uses a bandpass filter satisfying

$$\mathbf{h} = \arg\min_{\mathbf{h}} \mathbf{h}^H \mathbf{R} \mathbf{h}, \qquad \text{subject to} \quad \mathbf{h}^H \mathbf{a}(\omega) = 1. \tag{7}$$

The solution of (7) is standard and is given by [8]:

$$\mathbf{h} = \frac{\mathbf{R}^{-1}\mathbf{a}(\omega)}{\mathbf{a}^H(\omega)\mathbf{R}^{-1}\mathbf{a}(\omega)}. \tag{8}$$

Inserting (8) into (4) gives the filter output power

$$E\left\{|y_F(n)|^2\right\} = \frac{1}{\mathbf{a}^H(\omega)\mathbf{R}^{-1}\mathbf{a}(\omega)}. \tag{9}$$

Let β denote the bandwidth of the filter given by (8). Then the Capon PSD estimate has the form

$$\hat{\phi}(\omega) = \frac{E\left\{|y_F(n)|^2\right\}}{\beta} = \frac{1}{\beta\mathbf{a}^H(\omega)\mathbf{R}^{-1}\mathbf{a}(\omega)}. \tag{10}$$

Since the (equivalent) time-bandwidth product is equal to unity, one way is to choose β as the reciprocal of the temporal length of the Capon filter; that is

$$\beta = \frac{1}{M+1}. \tag{11}$$

By choosing the filter bandwidth as given by (11), we obtain the so-called *Capon-1* PSD estimator [4] [8]:

$$\text{Capon-1:} \qquad \hat{\phi}(\omega) = \frac{M+1}{\mathbf{a}^H(\omega)\hat{\mathbf{R}}^{-1}\mathbf{a}(\omega)}, \tag{12}$$

where $\hat{\mathbf{R}}$ is the sample covariance matrix (to be discussed later on).

Another more elaborate choice of β is obtained as the equivalent bandwidth of $|H(\omega)|^2$, where $H(\omega)$ is the filter's frequency response defined in (5) [5]. This specific bandwidth choice leads to the *Capon-2* PSD estimator [5] [8]:

$$\text{Capon-2:} \qquad \hat{\phi}(\omega) = \frac{\mathbf{a}^H(\omega)\hat{\mathbf{R}}^{-1}\mathbf{a}(\omega)}{\mathbf{a}^H(\omega)\hat{\mathbf{R}}^{-2}\mathbf{a}(\omega)}. \tag{13}$$

There are several ways to calculate the sample covariance matrix $\hat{\mathbf{R}}$. A simple one is the Toeplitz and Hermitian matrix formed from $\{\hat{r}(j-i)\}, i, j = 0, \ldots, M$, where $\hat{r}(k)$ is obtained by using (1). Since the forward-backward method generally achieves better estimation performance than the forward-only method as well as the Toeplitz and Hermitian $\hat{\mathbf{R}}$ [8] [10], we use the following sample covariance matrix:

$$\hat{\mathbf{R}} = \frac{1}{2(N-M)} \sum_{n=M+1}^{N} \left\{\mathbf{y}(n)\mathbf{y}^H(n) + \mathbf{J}\mathbf{y}^*(n)\mathbf{y}^T(n)\mathbf{J}\right\}, \tag{14}$$

where $\mathbf{J}$ denotes the exchange matrix whose cross diagonal elements are ones and all the others are zeros. To make sure that $\hat{\mathbf{R}}^{-1}$ exists, we must have $M < N/2$.

Burg showed that the inverse of Capon-1 spectrum is equal to the average of the inverses of the estimated AR spectra of orders from 0 to M [13]. This observation reveals the fact that Capon-1 has less statistical variation as well as lower spectral resolution than the AR estimator. A similar but more involved relationship between Capon-2 and the AR estimators was derived in [8]. Theoretically, the performance of Capon-2 is hard to quantify. However, it is generally believed that Capon-2 possesses finer resolution and hence is a better spectral estimator than Capon-1 [5]. We will show here, with a typical example, that even though Capon-2 has better resolution *locally* around the power peaks, it is *globally* a more biased estimator than Capon-1. Our experience also

shows that Capon-2 generally gives much poorer covariance estimates than Capon-1. Therefore, Capon-2 is not recommended for covariance sequence estimation.

To illustrate the above claim, consider an ARMA(4,2) signal

$$\begin{aligned} y(n) &= 2.76y(n-1) - 3.809y(n-2) + 2.654y(n-3) - 0.924y(n-4) \\ &\quad +e(n) - 0.9e(n-1) + 0.81e(n-2), \end{aligned} \tag{15}$$

where $e(n)$ is a real white Gaussian random process with zero-mean and unit variance. The Capon-1 and Capon-2 spectral estimates with $N = 256$ and $M = 50$ are shown in Figure 1(a), where the dashed curve stands for the true PSD of the ARMA signal, while the solid and the dashdotted lines, respectively, indicate the Capon-1 and Capon-2 estimates. It is obvious that Capon-2 is highly biased and suffers from a significant power loss. This observation has also been made with other signals, especially if the signal is a narrowband signal. An explanation of this behavior follows.

The calculation of the filter bandwidth in Capon-2 is applicable only if the Capon filter is a narrowband filter. Recall that the Capon method aims to find the Capon filter that minimizes the total output power of the overall frequency band while passes the current frequency ω undistorted. No effort has been taken to make sure that the Capon filter is narrowband. Let the lobe of the Capon filter frequency response where the current frequency of interest ω is located be called as the *mainlobe*, while all the others are called the *sidelobes*. It has been found that the steering frequency ω is not necessarily at the maximum or the center of the mainlobe [8] [11]. Furthermore, if the input signal is a narrowband signal, there may exist "sidelobes", located at frequency bands where the power level of the input signal is low, that are even larger than the "mainlobe" of the Capon filter frequency response. Note that the large "sidelobes" do not make any significant contributions to the filter output power so that the filter design criterion is still satisfied; that is, the output power is minimized, while the frequency response at ω is one. In all these cases, the Capon filter is not a narrowband filter and hence it calculates an *overestimated* filter bandwidth. Hence the Capon-2 PSD estimates become highly biased. However, it is interesting to note that Capon-2 does possess higher resolution capability, around the power peaks, than Capon-1. This is better illustrated in Figure 1(b), which shows the PSD estimates of the same ARMA signal as used in Figure 1(a) but with $N = 32$ and $M = 10$. The Capon-1 estimator cannot resolve the two power peaks this time, while Capon-2, albeit biased, still can.

For the reasons above, we do not consider using Capon-2 for covariance sequence estimation in the sequel.

4 Capon Covariance Estimator

We describe below how the Capon PSD estimates can be Fourier inverted in a rather convenient manner yielding the Capon covariance sequence estimates. The study of the covariance sequences corresponding to the Capon spectra is an interesting endeavor by itself, which apparently has not been undertaken in the literature before. We also present an approximate but computationally more efficient method to calculate the Capon covariance estimates from the Capon PSD estimates.

4.1 Exact Method

Lemma 1 *Let* $\mathbf{\Gamma} = \{\Gamma_{i,j}\} \in \mathcal{C}^{(M+1)\times(M+1)}$ *and let* $\mathbf{a}(\omega) = [1 \quad e^{-i\omega} \dots e^{-iM\omega}]^T$*, then*

$$\mathbf{a}^H(\omega)\mathbf{\Gamma}\mathbf{a}(\omega) = \sum_{s=-M}^{M} \mu_s e^{is\omega}, \tag{16}$$

where

$$\mu_s = \sum_{k=\max(0,s)}^{\min(M+s,M)} \Gamma_{k,k-s}. \tag{17}$$

If $\mathbf{\Gamma}$ *is Hermitian, then* $\mu_s = \mu_{-s}^*$.

Proof: See Appendix A.

It is obvious that whenever $\mathbf{\Gamma}$ is non-negative definite we have $\mathbf{a}^H(\omega)\mathbf{\Gamma}\mathbf{a}(\omega) \geq 0$ for any ω. Thus $\mathbf{a}^H(\omega)\mathbf{\Gamma}\mathbf{a}(\omega)$ is a valid power spectrum. Furthermore, Lemma 1 indicates that in such a case $1/\mathbf{a}^H(\omega)\mathbf{\Gamma}\mathbf{a}(\omega)$ is in fact the power spectrum of an Mth-order AR process. Consequently, Capon-1 yields an equivalent AR(M) process (whereas Capon-2 yields an equivalent ARMA(M, M) process). By making use of (17), we can find the coefficients of the equivalent AR process. The calculation of the exact covariance sequences from the AR coefficients is a standard problem and can be solved, for example, via the inverse Levinson-Durbin algorithm (See [8] [10] and the references therein for more details). Hence the implementation of the Capon method for covariance estimation runs as outlined below:

Step 1: Pick up a value for $M(M \leq N/2)$ and compute $\hat{\mathbf{R}}$ by (14).

Step 2: Compute μ_s associated with $\mathbf{\Gamma} = \hat{\mathbf{R}}^{-1}$ by (17). Factorize $\sum_{m=-M}^{M} \mu_m e^{im\omega}$ (say, by using the Newton-Raphson algorithm) and obtain the (minimum-phase) spectral factor.

Step 3: Compute the corresponding covariance sequence $\left\{\hat{r}_{\text{temp}}(k)\right\}$ from the spectral factor (or, equivalently, the AR model) by, for example, the inverse Levinson-Durbin algorithm.

While using the Fourier inverting method for covariance estimation, it is necessary that the integral of the PSD estimate over all frequencies gives a good estimate of the signal power; otherwise there may be scaling errors in the covariance estimates. So after Step 3, we use a power compensation approach to obtain our ultimate Capon covariance estimates. Our final estimates $\left\{\hat{r}_{\text{Capon}}(k)\right\}$ have the form

$$\hat{r}_{\text{Capon}}(k) = \hat{P}_0 \frac{\hat{r}_{\text{temp}}(k)}{\hat{r}_{\text{temp}}(0)}, \tag{18}$$

where $\hat{P}_0$ is the estimated power of the signal

$$\hat{P}_0 = \frac{1}{N} \sum_{n=1}^{N} |y(n)|^2 . \tag{19}$$

4.2 Approximate Method

Since the Capon-1 spectrum is equivalent to an AR spectrum, the covariance sequence can be computed *exactly* as described above. With some accuracy loss, the computational demand of the

Capon method can be reduced. According to the Wiener-Khintchine theorem, we have

$$r(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\omega) e^{ik\omega} d\omega. \tag{20}$$

We can rewrite (20) as

$$r(k) = \frac{1}{2\pi}\int_{0}^{2\pi} \phi(\omega) e^{ik\omega} d\omega. \tag{21}$$

Let $\bar{N} \gg N$, and let the Capon spectrum be evaluated at $\omega_{\bar{n}} = 2\pi\bar{n}/\bar{N}$; that is, we calculate $\hat{\phi}(\omega_{\bar{n}})$, $\bar{n} = 0, \ldots, \bar{N}-1$. Then we can approximate the computation of the covariance sequence corresponding to $\hat{\phi}(\omega)$ by the following equation

$$\hat{r}(k) = \frac{1}{2\pi}\frac{2\pi}{\bar{N}}\sum_{\bar{n}=0}^{\bar{N}-1} \hat{\phi}(\omega_{\bar{n}}) e^{ik\omega_{\bar{n}}} = \frac{1}{\bar{N}}\sum_{\bar{n}=0}^{\bar{N}-1} \hat{\phi}(\omega_{\bar{n}}) e^{ik\omega_{\bar{n}}}, \tag{22}$$

which can be evaluated by using FFT. Since the error in approximating the integration by the summation above is $O(1/\bar{N})$, the errors introduced are quite small for large enough $\bar{N}$. Our computer simulations also confirm this observation.

The evaluation of $\hat{\phi}(\omega_{\bar{n}})$ in (22) by directly using (12) is computationally burdensome. We can instead make use of (16) as follows. Since $\hat{\mathbf{R}}^{-1}$ is Hermitian, by Lemma 1 we have $\mu_s = \mu_{-s}^*$. Then

$$\begin{aligned}
(M+1)\hat{\phi}^{-1}(\omega_{\bar{n}}) &= \sum_{s=-M}^{M} \mu_s e^{is\omega_{\bar{n}}} \\
&= \sum_{s=0}^{M} \mu_s e^{is\omega_{\bar{n}}} + \sum_{s=-M}^{0} \mu_s e^{is\omega_{\bar{n}}} - \mu_0 \\
&= \sum_{s=0}^{M} \mu_s e^{is\omega_{\bar{n}}} + \sum_{s=0}^{M} \mu_{-s} e^{-is\omega_{\bar{n}}} - \mu_0 \\
&= \hat{\phi}_1^*(\omega_{\bar{n}}) + \hat{\phi}_1(\omega_{\bar{n}}) - \mu_0,
\end{aligned} \tag{23}$$

where

$$\hat{\phi}_1(\omega_{\bar{n}}) = \sum_{s=0}^{\bar{N}-1} \mu_s^* e^{-is\omega_{\bar{n}}}, \tag{24}$$

which, again, can be evaluated by using FFT with zero-padding, i.e., $\mu_s = 0$, for $s = M+1, \ldots, \bar{N}-1$.

4.3 Computational Aspects

We briefly discuss the computational aspects of the standard method and the Capon method for covariance sequence estimation. Recall that K, M, and N denote the largest lag of the covariance estimates, the length of the Capon filter, and the number of data samples, respectively. Assume that the data is real, $N \gg K$, $N \gg M$, and $K \approx M$. Then the standard method involves approximately $2KN$ flops. (A *flop* here is defined as a floating point operation; that is, a flop is either a floating point addition or a floating point multiplication.)

The number of flops needed by the exact Capon method is difficult to determine. For a moderate M, say $N/6 \leq M \leq N/3$, we find that the computationally most demanding part of the exact Capon method is the spectral factorization. Since the algorithm (Newton-Raphson) used to find the minimum-phase spectral factor is iterative, the number of flops needed is hard to quantify. Another significant computational demand is the calculation of the sample covariance matrix (14), which involves approximately $2(N-M)(M+1)^2+4(M+1)^3$ flops. However, we note that the amount of computations needed by the factorization is much larger than that needed for computing the sample covariance matrix for a moderate or large M.

The approximate Capon method no longer needs to perform the spectral factorization, and the computational demand is greatly reduced. The approximate Capon method requires approximately $2(N-M)(M+1)^2+6(M+1)^3$ flops, a large part of which comes from the computation of the sample covariance matrix $\hat{\mathbf{R}}$.

To illustrate quantitatively the computational burdens of the above three methods, we show a few simulation results in Table 1, where we define ζ as the ratio of the flops needed by the exact or approximate Capon method to that corresponding to the standard method. We remark that we did not pay special attention to the coding of the algorithms, hence the numbers provided here should be only taken as indicative of the computational complexities of the methods.

5 Numerical Results

In this section, we present numerical examples showing the performance of the Capon method for covariance estimation. The first problem addressed is the ARMA covariance estimation. A variety of ARMA signals with different pole and zero locations are generated to compare the performances of the standard and the Capon methods. We also consider AR coefficient estimation of the ARMA signals by using the Capon and standard covariance estimates. Finally, we give an example illustrating the MA model determination with the standard and Capon methods.

5.1 ARMA Covariance Estimation

First we comment on the generation of the ARMA signals. To eliminate the initial transient caused by improper initialization of the ARMA system we proceed as follows. Given the coefficients of an ARMA(p,q), we can determine the covariance matrix (see e.g., [8])

$$\mathbf{C} = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r^*(1) & r(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r(1) \\ r^*(p-1) & \cdots & r^*(1) & r(0) \end{bmatrix}. \tag{25}$$

We use $\mathbf{C}^{1/2}\mathbf{e}$ as the initial condition for the ARMA system, where $\mathbf{e} \in \mathcal{C}^{p\times 1}$ is a zero-mean Gaussian random vector with identity covariance matrix.

Four ARMA signals are chosen for our study. In the selection of these ARMA signals, efforts have been made to make them representative of a large class of ARMA signals. The coefficients of the ARMA signals are listed in Tables 2. The pole and zero diagrams are shown in Figures 2(a) to 2(d). Figures 3(a) to 3(d) give the PSD's of the ARMA signals, while Figures 4(a) to 4(d) show the corresponding covariance plots for these signals.

One critical choice is the value of the Capon filter length M. If M is too large and approaches $N/2$, $\hat{\mathbf{R}}$ is likely to be nearly singular. If M is too small, the resolution of the spectral estimates is expected to become worse; on the other hand, the accuracy of $\hat{\mathbf{R}}$ will increase with decreasing M, because more outer products are averaged in (14). Hence M should be chosen by considering the tradeoffs between the resolution and the statistical accuracy of the Capon method. Our studies suggest the following rule of thumb for choosing M: $N/12 \leq M \leq N/3$.

In the following numerical examples, we set $N = 256$ and $M = 50$. (Shorter data lengths have also been considered, and the results are similar.) The performances of the standard and the Capon methods for each of the four ARMA models are shown in Figures 5(a) to 5(d), where the curves show the mean-squared errors (MSE) of the covariance estimates, normalized with respect to $r(0)$, versus the time-lag of the covariance sequences. (Only the results from the exact Capon method are demonstrated owning to space limitation.) The MSE values are based on 100 independent realizations. It has been found that the Capon method generally gives better results than the standard method, especially for large time-lags. We also note that, when the poles are close to the unit circle, as in Figure 5(b), the performance differences between the two methods are not so large as in the other cases. (Though, the Capon method is still better than the standard method.)

5.2 AR Coefficient Estimation for ARMA Signals

As an application of the Capon method for covariance estimation, we include here an example on how to use the Capon covariance estimates to find the AR coefficients of ARMA signals via the *overdetermined modified Yule-Walker (OMYW) method* [6] [7] [8]. We first briefly explain the OMYW method.

For an ARMA(p, q) process, the covariance sequence and the AR coefficients are related by

$$\begin{bmatrix} r(q) & r(q-1) & \cdots & r(q-p+1) \\ r(q+1) & r(q) & \cdots & r(q-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ r(q+L-1) & r(q+L-2) & \cdots & r(q+L-p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} r(q+1) \\ r(q+2) \\ \vdots \\ r(q+L) \end{bmatrix}. \tag{26}$$

If $L = p$ in (26), then we have a system of q equations with q unknowns. These equations are referred to as the *modified Yule-Walker equations* [6] since they constitute a generalization of the Yule-Walker equations for the AR signals. Replacing the theoretical covariances $\{r(k)\}$ by their sample estimates $\{\hat{r}(k)\}$ in (26) yields

$$\begin{bmatrix} \hat{r}(q) & \hat{r}(q-1) & \cdots & \hat{r}(q-p+1) \\ \hat{r}(q+1) & \hat{r}(q) & \cdots & \hat{r}(q-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}(q+L-1) & \hat{r}(q+L-2) & \cdots & \hat{r}(q+L-p) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = - \begin{bmatrix} \hat{r}(q+1) \\ \hat{r}(q+2) \\ \vdots \\ \hat{r}(q+L) \end{bmatrix}. \tag{27}$$

The overdetermined case of $L > p$ in (27) is motivated by the fact that additional "information" in the higher-lag covariances can be exploited to improve the accuracy of the AR coefficient estimates; that is, we can make use of the additional information by choosing $L > p$ and solving the so-obtained overdetermined system of equations, either in a least-squares (LS) or in a total-least-squares (TLS) sense [14]. This is especially the case if there are poles in the ARMA model close to the unit circle, since the covariance sequence decays to zero very slowly in such a case.

The performances of the AR coefficient estimation for the ARMA signals via the OMYW method by using both the standard and Capon covariance estimates are shown in Figures 6(a) to 6(d), where $N = 256$ and $M = 32$, and where the plots are, again, based on 100 independent trials. The OMYW equations are solved by using the LS method. To reduce the number of figures, the curves are the sum of the MSE's of all AR coefficient estimates versus the number of equations used in the OMYW method. When using the OMYW method, we assume that the AR orders are known. The numbers of equations used in the curves are $L = 4, 8, 16, 32, 64$, and 128. It has been found that the AR coefficient estimates obtained by using the Capon covariance estimates are usually better than those obtained by the sample covariance estimates. Note that for the OMYW estimator based on the sample covariance estimates, the estimation performance may significantly deteriorate with increasing L (a large value of L may be used for lack of *a priori* information on the ARMA signal in question), whereas the Capon-OMYW estimator's performance is much less affected by the increase of L.

5.3 MA Model Order Determination

For an MA(q) process, the covariances with lags larger than q are all zeros [8] [10]. However, we can expect that the sample covariance estimates for an MA signal will not decay to zero fast enough, while the Capon covariance estimates will usually give $\hat{r}(k) \approx 0$, for $k > q$, since the Capon method gives better covariance estimates for higher lags. Hence the inference about the type of signal we are dealing with and its order will be easier to make with the Capon method.

We consider an MA(4) signal

$$y(n) = e(n) - 2.76e(n-1) + 3.809e(n-2) - 2.654e(n-3) + 0.924e(n-4). \tag{28}$$

We use both the standard and the Capon methods to determine its covariance estimates. The results are shown in Figures 7(a) and 7(b) where $N = 64$ and $M = 6$, and where 10 superimposed realizations for both methods are displayed. As expected, the Capon covariance estimates decay to almost zero after $k = 5$, while the standard covariance estimates are much more erratic. Consequently the inference that the process under study is an MA(4) is easier to make by using the Capon covariance estimator.

6 Conclusions

We have shown that the Capon method can be used to obtain more accurate covariance estimates than the standard biased sample covariance estimates, especially for large lags. The Capon spectrum is equivalent to an AR spectrum and hence the corresponding covariance sequence can be conveniently calculated via the inverse Levinson-Durbin algorithm. Since the spectral factorization needed for the exact Capon covariance estimation can be computationally expensive, the approximate Capon method proposed in this paper, which avoids the spectral factorization and makes use of FFT, may be more desirable,

Appendix A - Proof of Lemma 1

Proposition

$$\sum_{k=0}^{M}\sum_{j=0}^{M} f(k,k-j) = \sum_{s=0}^{M}\sum_{k=s}^{M} f(k,s) + \sum_{s=1}^{M}\sum_{k=0}^{M-s} f(k,-s) \tag{29}$$

Proof:

$$\begin{aligned}\sum_{k=0}^{M}\sum_{j=0}^{M} f(k,k-j) &= \sum_{j=0}^{M}\sum_{k=0}^{M} f(k,k-j) \\ &= \sum_{j=0}^{M}\sum_{k=j}^{M} f(k,k-j) + \sum_{j=1}^{M}\sum_{k=0}^{j-1} f(k,k-j)\end{aligned} \tag{30}$$

Let $s = k - j$. Note that the first term of the right side of (30) corresponds to $s \geq 0$ and the second term corresponds to $s < 0$. Since $0 \leq k \leq M$ and $0 \leq j \leq M$, it is obvious that

$$\begin{array}{ll} M \geq k = s + j \geq s, & \text{for} \quad s \geq 0; \\ 0 \leq k = s + j \leq s + M = M - |s|\,, & \text{for} \quad s < 0. \end{array} \tag{31}$$

Therefore, we can rewrite (30) as

$$\begin{aligned}&\sum_{s=0}^{M}\sum_{k=s}^{M} f(k,s) + \sum_{s=-M}^{-1}\sum_{k=0}^{M-|s|} f(k,s) \\ = &\sum_{s=0}^{M}\sum_{k=s}^{M} f(k,s) + \sum_{s=1}^{M}\sum_{k=0}^{M-s} f(k,-s),\end{aligned} \tag{32}$$

which concludes the proof of the proposition.

Proof of Lemma 1: By making use of the above proposition we obtain:

$$\begin{aligned}\mathbf{a}^H(\omega)\mathbf{\Gamma}\mathbf{a}(\omega) &= \sum_{k=0}^{M}\sum_{j=0}^{M} a_k^* \Gamma_{kj} a_j \\ &= \sum_{k=0}^{M}\sum_{j=0}^{M} e^{i(k-j)\omega}\Gamma_{k,k-(k-j)} \\ &= \sum_{s=0}^{M}(\sum_{k=s}^{M}\Gamma_{k,k-s})e^{is\omega} + \sum_{s=1}^{M}(\sum_{k=0}^{M-s}\Gamma_{k,k+s})e^{-is\omega} \\ &= \sum_{s=-M}^{M}\mu_s e^{is\omega},\end{aligned}$$

with μ defined in (17).

If $\mathbf{\Gamma}^H = \mathbf{\Gamma}$, then we have

$$\mu_{-s} = \sum_{k=0}^{M-s}\Gamma_{k,k+s}, \qquad \text{for} \quad s \geq 0. \tag{33}$$

Taking complex conjugation of both sides yields

$$\begin{aligned}(\mu_{-s})^* &= \sum_{k=0}^{M-s} \Gamma^*_{k,k+s} = \sum_{k=0}^{M-s} \Gamma_{k+s,k} \\ &= \sum_{l=s}^{M} \Gamma_{l,l-s} = \mu_s, \end{aligned} \tag{34}$$

and the proof of Lemma 1 is complete.

References

[1] T. J. Ulrych and T. N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Reviews of Geophysics and Space Physics*, vol. 13, pp. 183–200, February 1975.

[2] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proceedings of the IEEE*, vol. 70, pp. 963–974, September 1982.

[3] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, August 1969.

[4] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, August 1971.

[5] M. A. Lagunas and A. Gasull, "An improved maximum likelihood method for power spectral density estimation," *IEEE Transactions on Acoustics, Speech Signal Processing*, vol. ASSP-32, pp. 170–173, February 1984.

[6] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proceedings of the IEEE*, vol. 70, pp. 907–939, September 1982.

[7] P. Stoica, B. Friedlander, and T. Söderström, "Approximate maximum-likelihood approach to ARMA spectral estimation," *International Journal of Control*, vol. 45, no. 4, pp. 1281–1310, 1987.

[8] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, 1997.

[9] B. Porat, "Some asymptotic properties of the sample covariance of Gaussian autoregressive moving-average processes," *Journal of Time Series Analysis*, vol. 8, no. 2, pp. 205–220, 1987.

[10] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[11] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1469–1484, June 1996.

[12] P. Stoica, A. Jakobsson, and J. Li, "Matched-filter bank interpretation of some spectral estimators," *Signal Processing*, vol. 66, pp. 45–59, April 1998.

[13] J. P. Burg, "The relationship between maximum entropy and maximum likelihood spectra," *Geophysics*, vol. 37, pp. 375–376, April 1972.

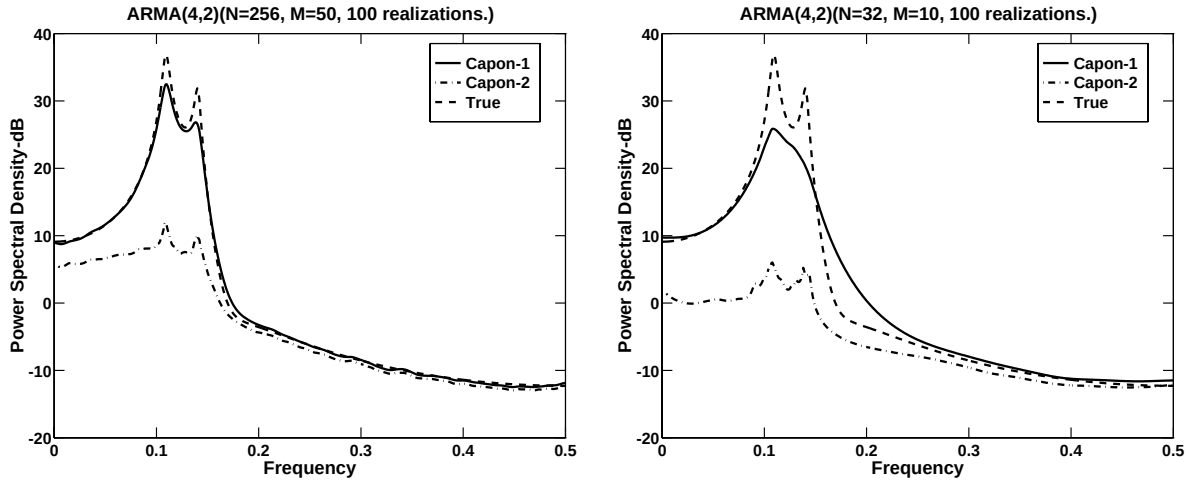
[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins University Press, 3 ed., 1996.

Flop Ratio ζ	Exact Capon	Approximate Capon
$M = K = 32$	281	47
$M = K = 50$	655	82
$M = K = 64$	1020	115

Table 1: Comparison of the computational burdens of the standard and Capon methods with $N = 256$ and $\bar{N} = 512$.

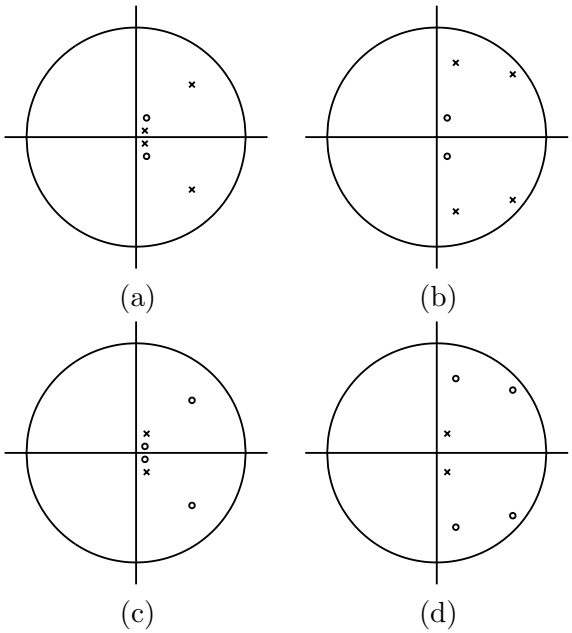
Processes	ARMA1	ARMA2	ARMA3	ARMA4
a_1	-1.1824	-1.7351	-0.2000	-0.2000
a_2	0.6651	1.7829	0.0400	0.0400
a_3	-0.0895	-0.9616	0.0000	0.0000
a_4	0.0049	0.3969	0.0000	0.0000
b_1	-0.2000	-0.2000	-1.1824	-1.7351
b_2	0.0400	0.0400	0.6651	1.7829
b_3	0.0000	0.0000	-0.0895	-0.9616
b_4	0.0000	0.0000	0.0049	0.3969
σ^2	1.0000	1.0000	1.0000	1.0000

Table 2: The ARMA processes used in the numerical simulations.



(a) (b)

Figure 1: Power spectral density estimates for the ARMA(4,2) signal given in Section 3 by using Capon-1 and Capon-2. The plots are the averages of 100 independent realizations. (a) $N = 256, M = 50$; (b) $N = 32, M = 10$.



(a) (b)

(c) (d)

Figure 2: Pole-zero diagrams for ARMA test cases. (a) ARMA1; (b) ARMA2; (c) ARMA3; (d) ARMA4.

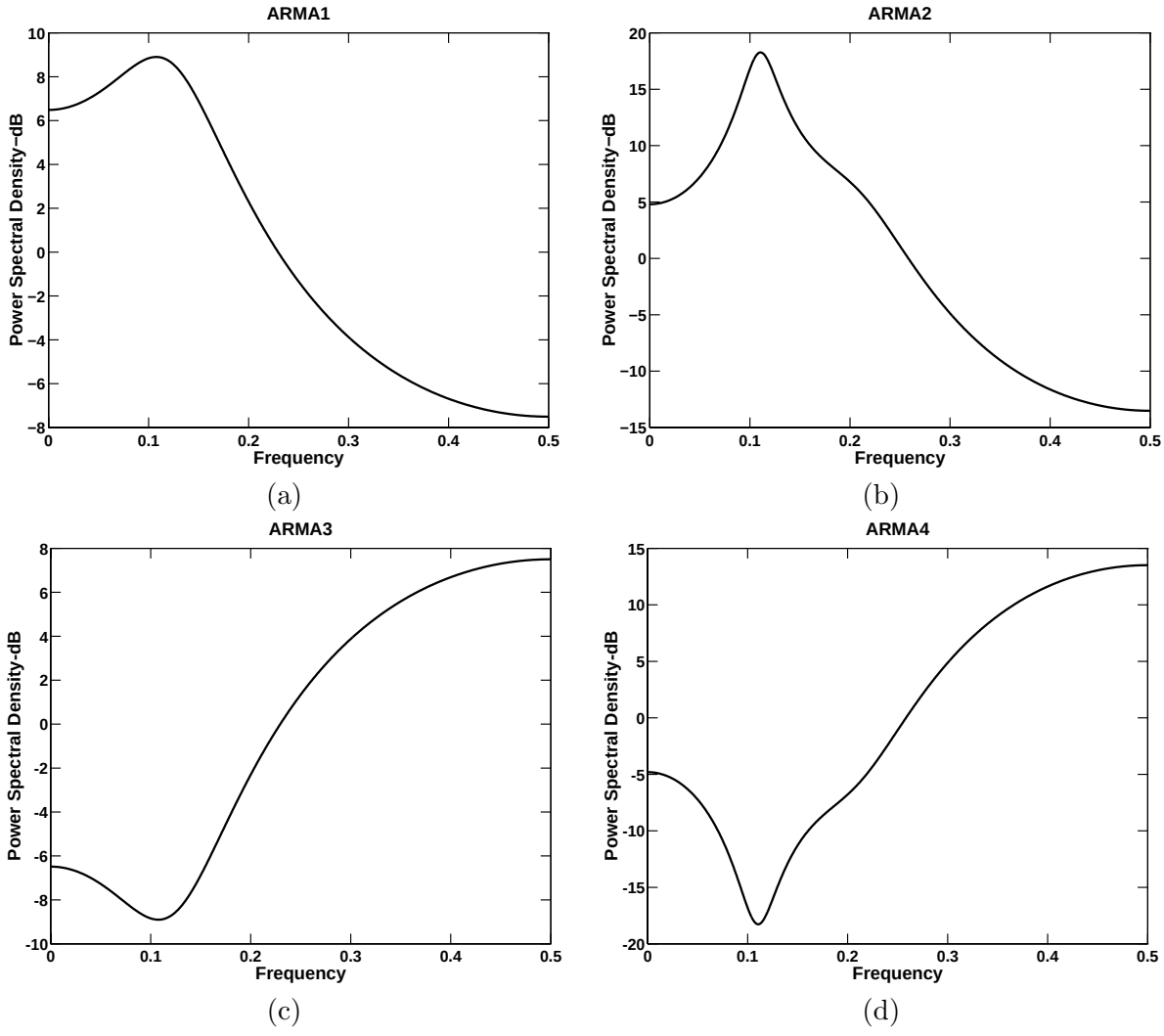


Figure 3: True power spectral densities. (a) ARMA1; (b) ARMA2; (c) ARMA3; (d) ARMA4.

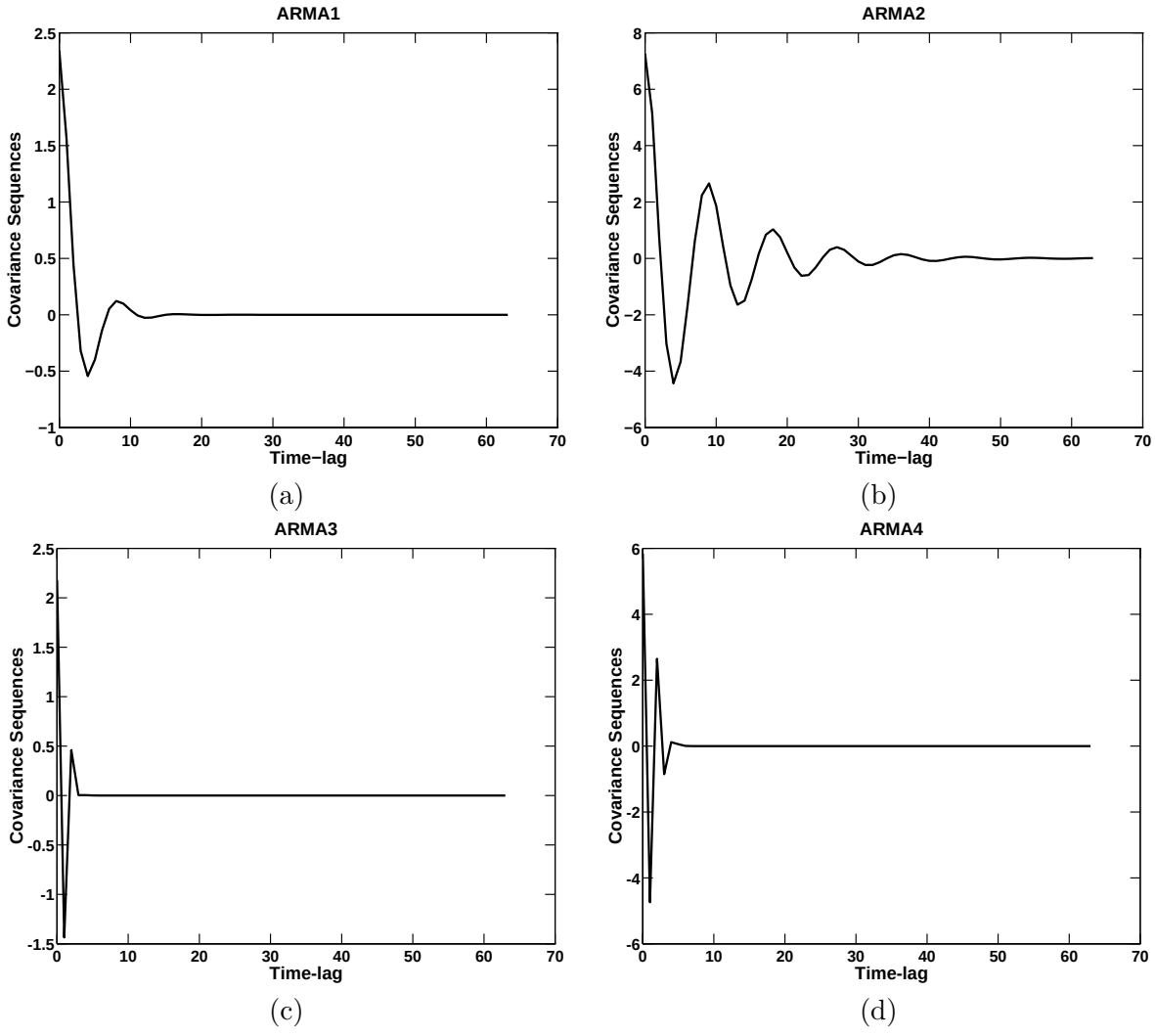


Figure 4: True covariance sequences. (a) ARMA1; (b) ARMA2; (c) ARMA3; (d) ARMA4.

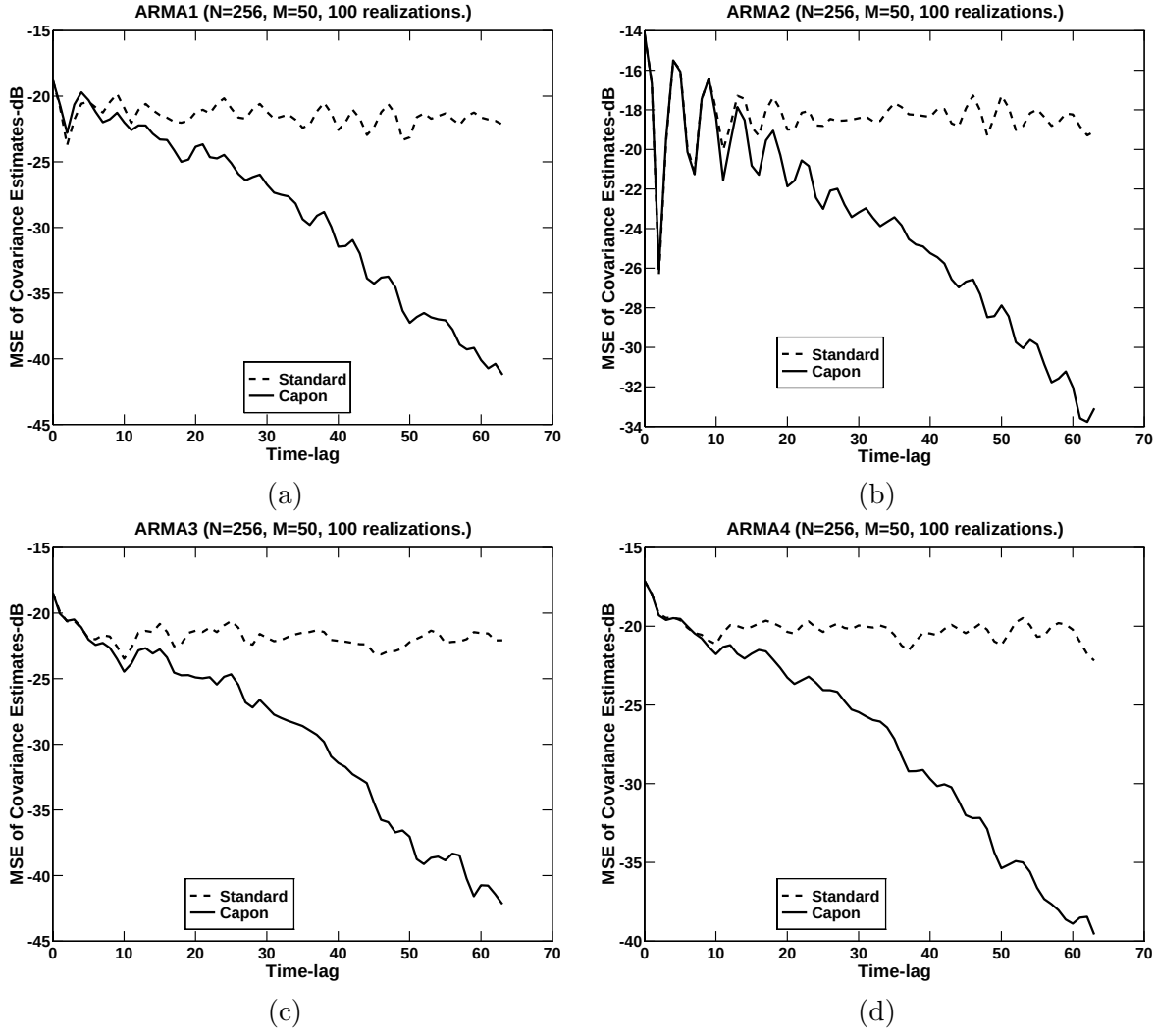


Figure 5: Covariance sequence estimation with $N = 256$ and $M = 50$. The mean-squared errors (MSE) of the covariance estimates, normalized with respect to $r(0)$, are based on 100 independent realizations. (a) ARMA1; (b) ARMA2; (c) ARMA3; (d) ARMA4.

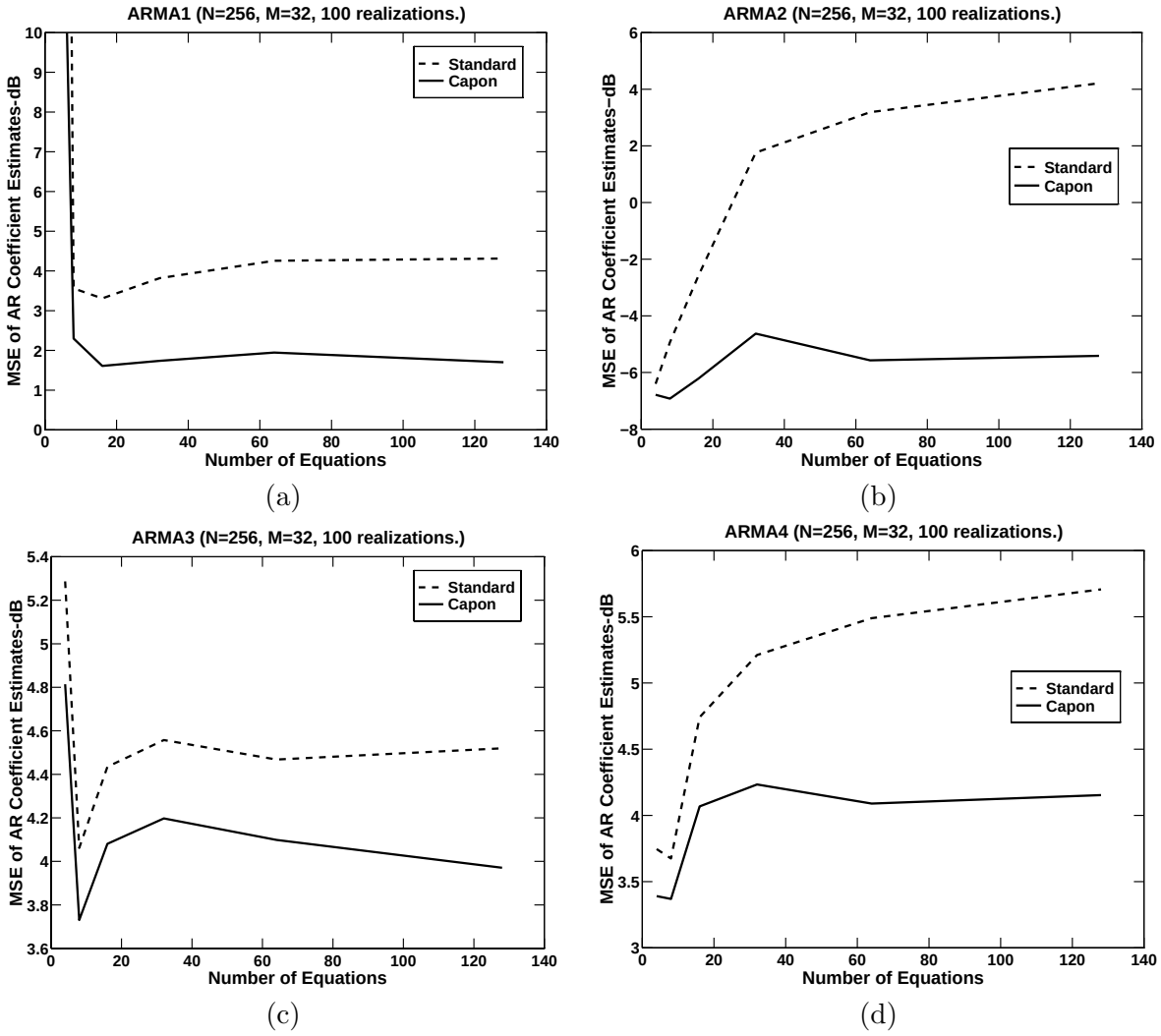
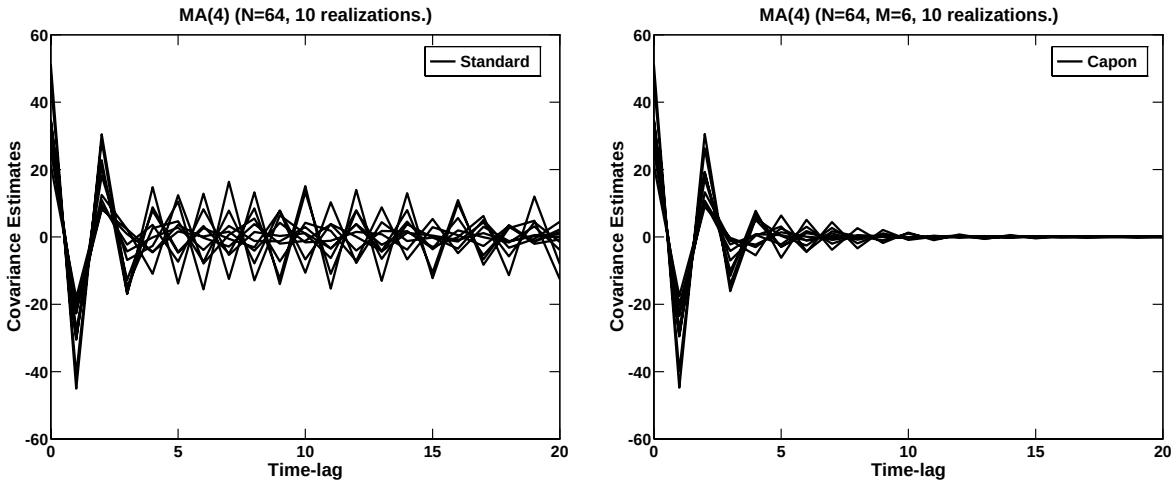


Figure 6: The AR coefficient estimation of the ARMA signals via the overdetermined modified Yule-Walker method with $N = 256$ and $M = 32$. The curves are the summations of the mean-squared errors (MSE) of all the AR coefficient estimates versus the numbers of included equations, which have been set as 4, 8, 16, 32, 64 and 128, respectively. The MSE curves are based on 100 independent realizations. (a) ARMA1; (b) ARMA2; (c) ARMA3; (d) ARMA4.



(a) (b)

Figure 7: 10 superimposed realizations of the MA covariance sequence estimates with $N = 64$ and $M = 6$. (a) The standard method; (b) The Capon method.