On Moving Average Parameter Estimation

Niclas Sandgren* and Petre Stoica*

Contact information: niclas.sandgren@it.uu.se, tel: +46 18 4713392

Abstract

Estimation of the autoregressive moving average (ARMA) parameters of a stationary stochastic process is a problem often encountered in the signal processing literature. It is well known that estimating the moving average (MA) parameters is usually more difficult than estimating the autoregressive (AR) part, especially if the zeros are located close to the unit circle. In this paper we present four linear methods for MA parameter estimation (i.e., methods that involve only linear operations) and compare their performances first in a case when the zeros are located far away from the unit circle and secondly in a presumably harder case when the zeros are located very close to the unit circle.

1 Introduction

Consider the following MA equation

$$\begin{aligned} y(t) &= e(t) + b_1 e(t-1) + \cdots + b_n e(t-n) \\ &= B(z^{-1})e(t) \qquad t = 0, \ldots, N-1 \end{aligned} \tag{1}$$

where

$$B(z^{-1}) = 1 + b_1 z^{-1} + \cdots + b_n z^{-n} \tag{2}$$

and where $\{e(t)\}$ is a white noise sequence with variance σ^2. N is the number of available samples. The model order n is assumed to be known in the following. If n is unknown it can be estimated using, for example, an information criterion rule (see, e.g., [1]).

Our problem lies in estimating the parameters $\{b_k\}$ from the measured signal $\{y(t)\}_{t=0}^{N-1}$. Several methods have been developed in the past for solving this problem (see, e.g., [2] [3] and the references therein). As long as the zeros associated with the polynomial in (2) are located reasonably far away from the unit circle most known MA parameter estimation methods perform satisfactorily. However, when the zeros are located close to the unit circle the problem becomes more intricate, and the accuracy of the parameter estimates usually decreases.

In this paper we will compare four *linear* MA parameter estimation techniques to see how their performances differ from one another, especially in the two cases when the zeros

*Systems and Control Division, Department of Information Technology, Uppsala University, P.O. Box 337, SE-751 05 Uppsala, Sweden.

are located close to respectively far away from the unit circle. We consider the well known Durbin's method (abbreviated DM in the following) [4] which is commonly used for estimation of the $\{b_k\}$ in (1). The second method considered is called the inverse covariance (or correlation) method (abbreviated ICM here) (see, e.g., [5] [6]). Finally, we consider two MA parameter estimation techniques based on the cepstrum, or the *vocariance sequence* (see, e.g., [7] - [13]). We call these two techniques the vocariance recursion method (VRM) and the vocariance ESPRIT method (VEM) (see [11] [12] and also below). The outline of VEM below includes a novel parameter estimation approach specifically designed for the problem when the zeros are located close to the unit circle. All methods are described in a concise and simple manner.

In Section 2 brief descriptions of these four methods (DM, ICM, VRM, and VEM) are presented. In Section 3, we show a simulation study based on two second-order MA examples in which we compare the four techniques with a nonlinear least squares (NLS) search method, in terms of estimation accuracy and computational speed. Finally, in Section 4, we give some conclusions and suggestions on which technique of those above should be preferred in a specific scenario.

2 Methods

2.1 Durbin's method (DM)

Durbin's method [4] is one of the most widely used techniques for MA parameter estimation. It is also known as the 2-stage LSM (least squares method) (see, e.g., [2] [3] and the references therein for more information). The two stages can be outlined as follows:

Step 1. The first step consists of fitting an AR model of order $m > n$ to $\{y(t)\}$. Once m has been specified, the estimated AR parameters $\{\hat{a}_k\}_{k=1}^m$ can be obtained via LSM. Hence, estimates $\{\hat{e}(t)\}$ of the noise sequence $\{e(t)\}$ can be computed as

$$\hat{e}(t) = \hat{A}(z^{-1})y(t) \qquad t = 0, \ldots, N-1 \tag{3}$$

where

$$\hat{A}(z^{-1}) = 1 + \hat{a}_1 z^{-1} + \cdots + \hat{a}_m z^{-m}.$$

Step 2. Using $\{\hat{e}(t)\}$ we can write

$$y(t) - \hat{e}(t) \approx [b_1 \ \ldots \ b_n] \begin{bmatrix} \hat{e}(t-1) \\ \vdots \\ \hat{e}(t-n) \end{bmatrix} \tag{4}$$

for $t = 0, \ldots, N-1$, from which estimates $\{\hat{b}_k\}$ of $\{b_k\}$ can be obtained via LSM. The model order m can be selected via the Akaike's information criterion (AIC) or the Bayesian information criterion (BIC) (see, e.g., [1]). However, a more expedient rule for selecting m is $m = 2n$, which we will use in the following.

2.2 Inverse covariance method (ICM)

The main idea behind this technique is outlined below. For more information, see, e.g., [5] [6] and the references therein.

The standard covariance sequence of a data string $\{y(t)\}$ is given by

$$r_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(\omega) e^{ik\omega} d\omega \qquad k = 0, \pm 1, \pm 2, \ldots \tag{5}$$

where $\Phi(\omega)$ is the power spectral density of $\{y(t)\}$. As the name suggests, the inverse covariance sequence is given by

$$\rho_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\Phi(\omega)} e^{ik\omega} d\omega \qquad (\text{assuming } \Phi(\omega) > 0 \ , \ \forall \omega). \tag{6}$$

For notational convenience, let $\{\Phi_p\}$ denote the values taken by the spectrum at the Fourier frequency grid points:

$$\omega_p = \frac{2\pi}{N} p \qquad p = 0, \ldots, N-1. \tag{7}$$

The periodogram estimate of Φ_p is given by (see, e.g., [3]):

$$\hat{\Phi}_p = \frac{1}{N}\left| \sum_{t=0}^{N-1} y(t) e^{-i\omega_p t} \right|^2 \tag{8}$$

for $p = 0, \ldots, N-1$. Using (8) we can estimate $\{\rho_k\}$ as

$$\hat{\rho}_k = \frac{1}{N}\sum_{p=0}^{N-1} \frac{1}{\hat{\Phi}_p} e^{i\omega_k p} \qquad k = 0, \ldots, \frac{N}{2} \tag{9}$$

(for $k > N/2$ the sequence is mirror symmetric). An alternative spectral estimator $\hat{\Phi}_p$ based on the long AR polynomial in (3) has also been tested. However, the difference in performance between the two estimators was small, and hence we will use the estimator in (8) since it has lower computational complexity.

Since

$$\frac{1}{\Phi(\omega)} = \frac{1}{\sigma^2 |B(e^{i\omega})|^2}, \tag{10}$$

which is an AR spectrum, it follows that $\{b_k\}$ and $\{\rho_k\}$ are related via the Yule-Walker equations. Hence we can get estimates $\{\hat{b}_k\}$ from $\{\hat{\rho}_k\}$ via the Yule-Walker method (see, e.g., [2] [3]).

2.3 Vocariance recursion method (VRM)

This technique is based on estimation of the *cepstrum*, or the *cepstral coefficients*, or, yet, the *vocariance sequence* (see, e.g., [7] - [10]).

By definition, the vocariance sequence $\{c_k\}$ satisfies

$$\ln \Phi(z) = \sum_{k=-\infty}^{\infty} c_k z^{-k} \tag{11}$$

where, in the present scenario, $\Phi(z) = \sigma^2 B(z) B(z^{-1})$. It follows that

$$\ln B(z) = \sum_{k=1}^{\infty} c_k z^k \tag{12}$$

which implies, by differentiation with respect to z, that

$$\frac{B'(z)}{B(z)} = \sum_{k=1}^{\infty} k c_k z^{k-1} \qquad \Leftrightarrow$$

$$\sum_{k=1}^{n} k b_k z^{k-1} = \sum_{k=0}^{n} b_k z^k \sum_{k=1}^{\infty} k c_k z^{k-1}. \tag{13}$$

The vocariances $\{c_k\}$ can be consistently estimated via

$$\hat{c}_k = \frac{1}{N} \sum_{p=0}^{N-1} \ln(\hat{\Phi}_p) e^{i\omega_k p} \qquad k = 1, \ldots, \frac{N}{2} \tag{14}$$

(for $k > N/2$ the sequence is mirror symmetric). Note that the right hand side in (13) can be rewritten as

$$\begin{aligned} &\sum_{p=0}^{n} \sum_{k=1}^{\infty} k c_k b_p z^{p+k-1} \\ &= \sum_{j=0}^{\infty} \left[\sum_{p=0}^{n} (j-p+1) c_{j-p+1} b_p \right] z^j \end{aligned} \tag{15}$$

(assuming $c_j = 0$ for $j \leq 0$). The left hand side in (13) can be rewritten as

$$\sum_{j=0}^{n-1} (j+1) b_{j+1} z^j. \tag{16}$$

Equating (15) and (16) leads to the recursion ($b_0 = 1$):

$$b_j = \frac{1}{j} \sum_{p=0}^{j-1} (j-p) c_{j-p} b_p \qquad j = 1, \ldots, n \tag{17}$$

which is a *triangular* linear system in $\{b_j\}$. Replacing $\{c_k\}$ in (17) by their estimates $\{\hat{c}_k\}$ obtained from (14) gives us the desired estimates $\{\hat{b}_j\}$ of $\{b_j\}$.

More details about this type of MA parameter estimation method, and its extension to ARMA signals, can be found in [11].

2.4 Vocariance ESPRIT method (VEM)

The final technique is also based on estimation of the vocariance sequence. The main idea behind this approach was partly described in [12]; however, [12] includes a nonlinear estimation step which is here replaced by the ESPRIT method (see, e.g., [3]). The resulting technique is not really a linear method since it involves a singular value decomposition (SVD) step. However, it can be considered to be "quasi-linear", since the SVD is such a reliable operation. For more information about the relationship (20) below between $\{b_k\}$ and the

vocariance sequence, which lies at the basis of VEM, see [13].

Let $\{1/z_p\}$ be the zeros of $B(z)$

$$B(z) = \prod_{p=1}^{n}(1 - z_p z) \qquad |z_p| < 1. \tag{18}$$

Then

$$\ln B(z) = \sum_{p=1}^{n} \ln(1 - z_p z) = -\sum_{p=1}^{n}\sum_{k=1}^{\infty} \frac{1}{k} z_p^k z^k. \tag{19}$$

Comparing (12) and (19) gives

$$-kc_k = \sum_{p=1}^{n} z_p^k. \tag{20}$$

Replacing $\{c_k\}$ in (20) by $\{\hat{c}_k\}$ obtained from (14) gives

$$-k\hat{c}_k \approx \sum_{p=1}^{n} z_p^k \qquad k = 1, \ldots, M \tag{21}$$

where $M \geq 2n$ is a user parameter. Note that the right hand side of (21) can be seen as a damped sinusoidal model. Hence we can use ESPRIT (see, e.g., [3]) to estimate $\{z_p\}$ from $\{-k\hat{c}_k\}$. Once $\{z_p\}$ have been estimated we can obtain estimates $\{\hat{b}_j\}$ of $\{b_j\}$ via (18).

Regarding the choice of M we note that even though the errors in $\{\hat{c}_k\}$ have the same variance (see [9]), the errors in $\{k\hat{c}_k\}$ have increasing variance as k increases. Moreover the "modes" $\{z_p^k\}$ in (21) may go quickly to zero with increasing k. Hence we should not choose M too large. In the following we will use $M = 4n$, which is a reasonable choice for several practical values of n.

3 Numerical Examples

The four MA parameter estimation techniques (DM, ICM, VRM, and VEM) will be compared with one another in two rather different numerical examples. We will also compare the performances of the above four linear methods to that of a nonlinear least squares (NLS) search method. The estimation criterion of the NLS method can be written as

$$\min_{\{b_k\}} \sum_{t=0}^{N-1} \left[\frac{1}{B(z^{-1})} y(t)\right]^2. \tag{22}$$

NLS achieves the Cramér-Rao lower bound (CRB) in the Gaussian data case, for $N \gg 1$ (see, e.g., [3] [14]). Due to the high accuracy of the NLS approach, it will be used as a reference in the comparative performance study below. Iterative nonlinear methods require adequate initial parameter estimates to prevent the search algorithm from stopping at local minima. In the numerical examples below the initial parameter values for the iterative NLS search are obtained via a special four-stage LS-IV (least-squares instrumental-variable) algorithm

(see [14] for more information).

As a performance measure we consider the total variance (TV) of $\{\hat{b}_k\}$

$$\mathrm{TV}(\hat{\mathbf{b}}) = \sum_{k=1}^{n} E\left[\hat{b}_k - b_k\right]^2 \tag{23}$$

where

$$\hat{\mathbf{b}} = [1\ \hat{b}_1\ \cdots\ \hat{b}_n]$$

and where E is the expectation operator which is estimated from 1000 Monte Carlo runs for each method. We will show the TV for DM, ICM, VRM, VEM, and NLS for N=128, 256, 512, 1024, and 2048. In addition we will show the computational time (in seconds) required to perform the 1000 Monte Carlo runs for each considered technique, also versus N.

Example 1 Consider the following MA sequence

$$y(t) = e(t) + 0.55\, e(t-1) + 0.15\, e(t-2)$$

$$t = 0, \ldots, N-1$$

where $\{e(t)\}$ is a white Gaussian noise sequence with zero mean and unit variance. The corresponding zeros are located at $z_{1,2} = -0.275 \pm 0.273i$ which correspond to a distance $|z_{1,2}| = 0.387$ from the origin.

In Fig. 1 we show the $\mathrm{TV}(\hat{\mathbf{b}})$ for the five methods for different values of N. In this example, with zeros far away from the unit circle, DM and VRM perform very well, having parameter estimation accuracies comparable to that of NLS. ICM shows a lower performance and VEM does not provide reliable parameter estimates in this case. The low performance of VEM can be explained by the fact that zeros close to the origin corresponds to heavily damped sinusoids which are hard to estimate using an ESPRIT-based method for a low value of M. The computational time required to perform 1000 Monte Carlo runs using each of the five techniques is presented in Fig. 2. The differences between the four linear methods (DM, ICM, VRM, and VEM) are small. The main observation from Fig. 2 is that all four linear methods are significantly faster (about 100 times) than NLS. In addition, the time required for NLS depends significantly on the considered example, the number of parameters to be estimated, the location of the zeros, and the initial estimates, whereas the computational time required for the other techniques does not depend as much on the data.

Example 2 Next consider the MA sequence

$$y(t) = e(t) - 1.4\, e(t-1) + 0.98\, e(t-2)$$

$$t = 0, \ldots, N-1$$

where $\{e(t)\}$ is again a white Gaussian noise sequence with zero mean and unit variance. The corresponding zeros are located at $z_{1,2} = 0.7 \pm 0.7i$ which correspond to a distance $|z_{1,2}| = 0.99$ from the origin.

The obtained values of $\mathrm{TV}(\hat{\mathbf{b}})$ are presented in Fig. 3 for the five methods for different values of N. In this example the zeros are located very close to the unit circle, which usually leads to a harder estimation problem than that of Example 1. In this case VRM and VEM

perform better than DM and ICM. None of the linear methods achieves the performance of the NLS approach. The required time to perform 1000 Monte Carlo runs for each of these five techniques is similar to Example 1 and the corresponding results are therefore omitted.

4 Conclusions

VRM was the only linear method which performed satisfactorily in both examples above. Even though DM is a commonly used MA estimation technique it is well known that its estimation accuracy degrades when the zeros are close to the unit circle for a fixed value of m. For such scenarios VRM or VEM can be appealing alternatives. The performance of DM can be improved by selecting a larger value of m, but at the expense of an increased computational complexity. From the simple simulation study above we conclude that the parameter estimation accuracy of ICM is lower than e.g. that of VRM. The significant difference in the performance of VEM in the two examples can be explained by the fact that an ESPRIT-based method can estimate sinusoidal components (in this case corresponding to zeros close to the unit circle) very accurately, however the estimation performance for heavily damped sinusoids (corresponding to zeros close to the origin here) is poor since we have to keep M rather small.

References

[1] P. Stoica and Y. Selén. Model-Order Selection: a Review of Information Criterion Rules. *IEEE Sign. Proc. Mag.*, vol. 21, pp.36-47, 2004.

[2] P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting* Springer, New York, 2002.

[3] P. Stoica and R. Moses. *Spectral Analysis of Signals.* Prentice Hall, Upper Saddle River, NJ, 2005.

[4] J. Durbin. Efficient Estimation of Parameters in Moving-Average Models. *Biometrika*, vol. 46, pp.306-316, 1959.

[5] C. Chatfield. Inverse Autocorrelations. *J. R. Statist. Soc. A*, vol. 142, pp.363-377, 1979.

[6] R.J. Bhansali. A Simulation Study of Autoregressive and Window Estimators of the Inverse Correlation Function. *Applied Statist.*, vol. 32, pp.141-149, 1983.

[7] C.I. Byrnes, P. Enqvist and A. Lindquist. Cepstral Coefficients, Covariance Lags, and Pole-Zero Models for Finite Data Strings. *IEEE Trans. Sign. Proc.*, vol. 49, pp.677-693, 2001.

[8] P. Enqvist. A Convex Optimization Approach to ARMA(n, m) Model Design from Covariance and Cepstral Data. *SIAM J. Control Optim.*, vol. 43, pp.1011-1036, 2004.

[9] P. Stoica and N. Sandgren. Smoothed Nonparametric Spectral Estimation via Cepstrum Thresholding. *IEEE Sign. Proc. Mag.*, Submitted.

[10] P. Stoica and N. Sandgren. Total-Variance Reduction via Thresholding: Application to Cepstral Analysis. *IEEE Trans. Sign. Proc.*, Submitted.

[11] A. Kaderli and A.S. Kayhan. Spectral Estimation of ARMA Processes Using ARMA-Cepstrum Recursion. *IEEE Sign. Proc. Lett.*, vol. 7, pp.259-261, 2000.

[12] S.M. Kay, L.B. Jackson, J. Huang and P.M. Djurić. An Approximate Maximum Likelihood ARMA Estimator Based on the Power Cepstrum. In *Proc. Acoust. Speech Sign. Proc. (ICASSP)*, pp.2344-2347, 1988.

[13] M.R. Schroeder. Direct (Nonrecursive) Relations Between Cepstrum and Predictor Coefficients. *IEEE Trans. Acoust. Speech Sign. Proc.*, vol. 29, pp.297-301, 1981.

[14] L. Ljung. *System Identification - Theory For the User.* Prentice Hall, Upper Saddle River, NJ, 1999.

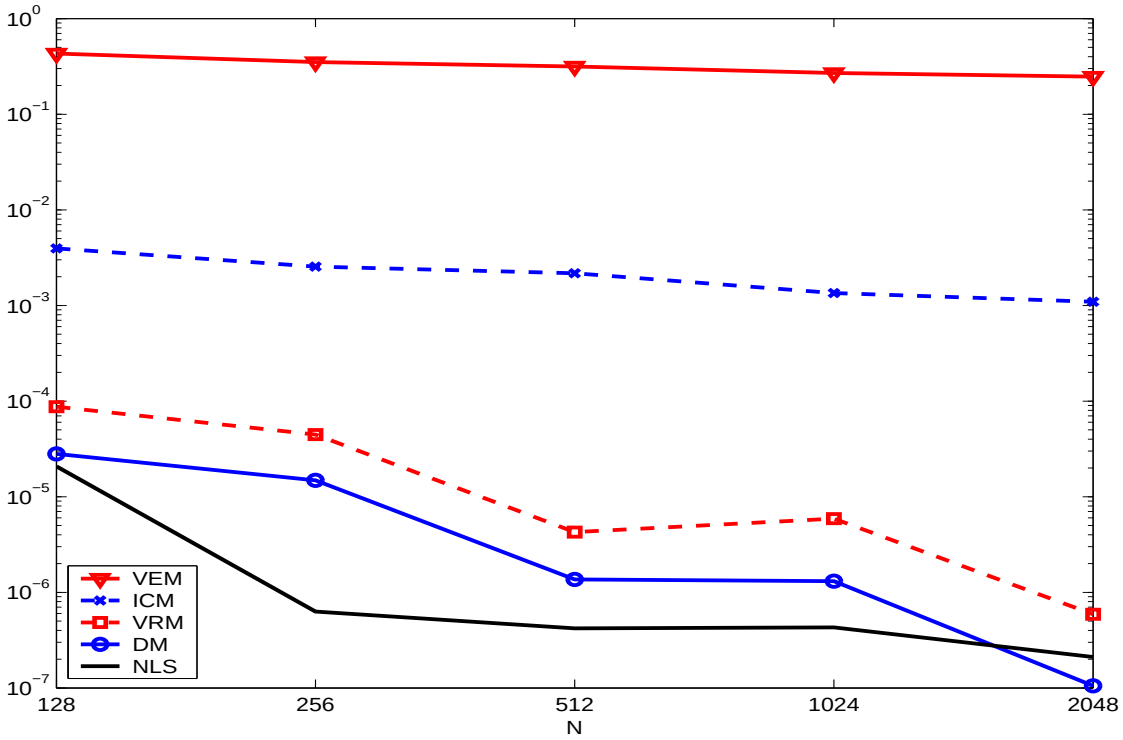

Figure 1: The TV($\hat{\mathbf{b}}$), versus N, for the five methods (DM, ICM, VRM, VEM, and NLS) in Example 1.

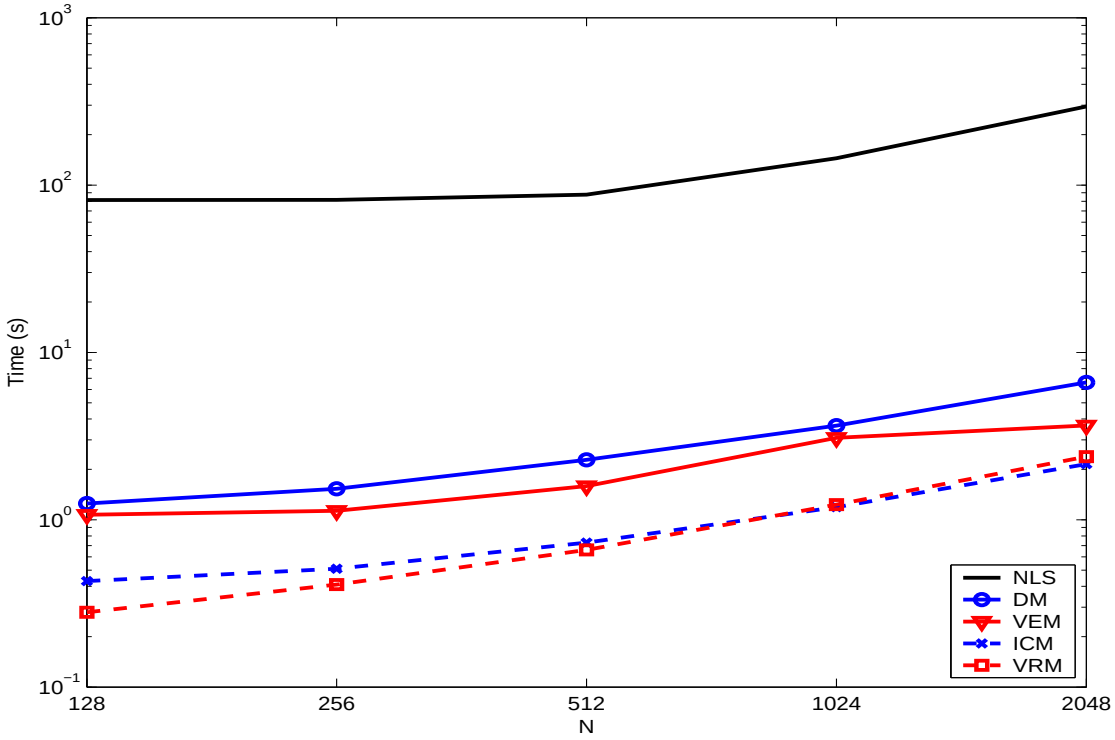

Figure 2: Time, in seconds, required to perform 1000 Monte Carlo runs, versus N, for each of the five methods (DM, ICM, VRM, VEM, and NLS) in Example 1.

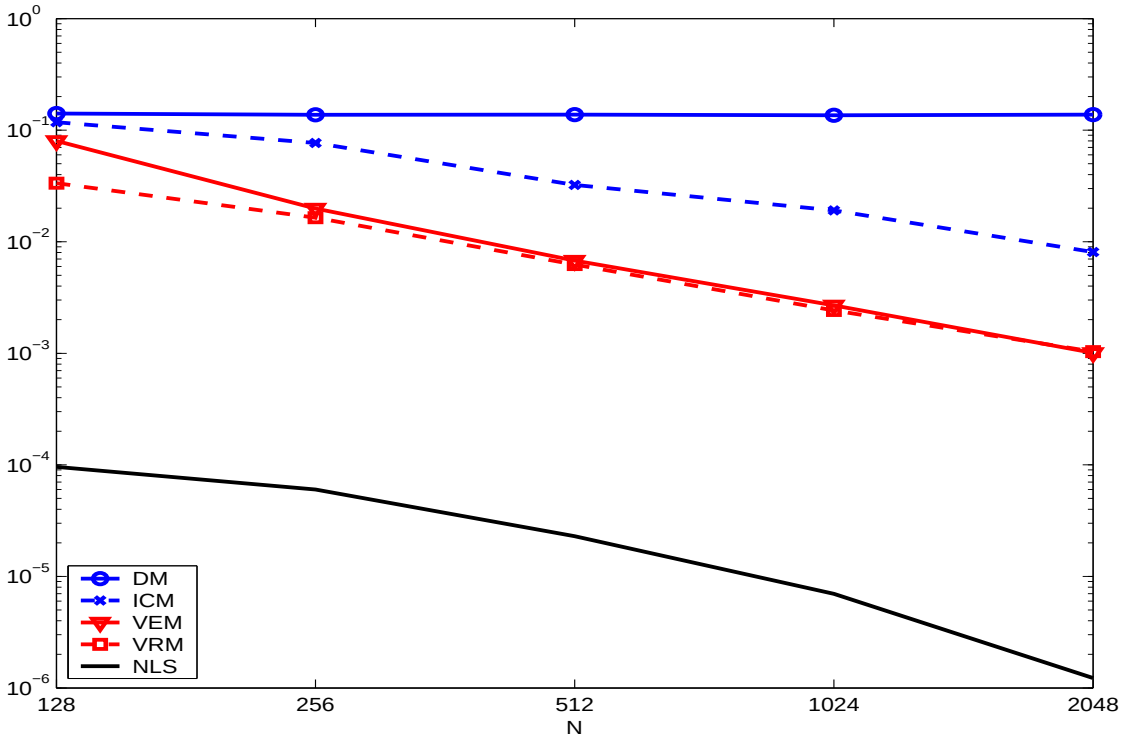

Figure 3: The TV($\hat{\mathbf{b}}$), versus N, for the five methods (DM, ICM, VRM, VEM, and NLS) in Example 2.