# ROBUST ML ESTIMATION FOR UNKNOWN NUMBERS OF SIGNALS

*Pei-Jung Chung*

Institute for Digital Communications
Joint Research Institute for Signal & Image Processing
School of Engineering and Electronics, The University of Edinburgh, UK
peijung.chung@gmail.com

## ABSTRACT

The number of signals plays an important role in array processing. The performance of direction finding algorithms relies strongly on a correctly specified number of signals. When the number of signals is unknown, conventional approaches apply the information theoretic criterion or multiple tests to estimate the number of signals and parameters of interest simultaneously. These methods usually require a series of maximizations over parameter spaces of different dimensions and result in high computational cost. In this work, we develop a novel approach for finding ML estimates without knowing the number of signals. Given an upper bound on the number of signals, the proposed method carries out one maximization and selects relevant components from the estimated parameter vector. Simulation results show that the proposed method provides comparable estimation accuracy as the standard ML method does.

## 1. INTRODUCTION

The problem of estimating direction of arrival (DOA) is a key issue in array processing. Among existing methods, the maximum likelihood (ML) approach has the best statistical properties. It is also known to be robust against small sample numbers, signal coherence and closely located sources.

The standard ML method assumes the number of signals, $m$, to be known and maximizes the concentrated likelihood function over an $m$-dimensional parameter space. In the case of unknown numbers of signals, conventional approaches estimate the number of signals together with unknown parameters using the information theoretic criterion based methods [5, 8] or the multiple hypothesis tests [3, 4]. Given an upper bound on the number of signals, $M$, these methods maximize the likelihood function over a series of parameter spaces with increasing dimensions. The ML estimates obtained from $M$ different parameter spaces are compared with each other. Finally, one of the $M$ estimates is selected by the underlying criterion. The total computational cost can be very high due to the multi-dimensional search for each assumed model order.

In this work, we suggest a novel procedure that requires *only* one maximization over an $M$-dimensional parameter space. According to a recent study [2], the ML estimator derived from an overestimated model order contains components that converge to the true parameters. Furthermore, the components that are associated with the true parameters lead to an increase of the log-likelihood function. Motivated by this observation, we define the relevance value for each component of the estimated parameter vector. With the help of relevance values, we can easily select the relevant estimates that coincide with the true parameters. As a byproduct, we can use the number of the relevant estimates to determine the number of signals. Clearly, the proposed approach requires no sequential maximization over various parameter spaces and is computationally more attractive than the conventional methods.

In the following, we give a brief description of the signal model. Section 3 describes useful properties of ML estimation of misspecified models. In section 4, we develop the robust ML estimation method. Simulation results are presented and discussed in section 5. Our concluding remarks are given in section 6.

## 2. PROBLEM FORMULATION

Consider an array of $n$ sensors receiving $m$ narrow band signals emitted by far-field sources located at $\boldsymbol{\theta}_m = [\,\theta_1, \ldots, \theta_m]^T$. The array output $\boldsymbol{x}(t)$ is described as

$$\boldsymbol{x}(t) = \boldsymbol{H}_m(\boldsymbol{\theta}_m)\boldsymbol{s}_m(t) + \boldsymbol{n}(t), \quad t = 1, \ldots, T, \qquad (1)$$

where the $i$th column $\boldsymbol{d}(\theta_i)$ of the matrix

$$\boldsymbol{H}_m(\boldsymbol{\theta}_m) = [\boldsymbol{d}(\theta_1) \cdots \boldsymbol{d}(\theta_i) \cdots \boldsymbol{d}(\theta_m)] \qquad (2)$$

represents the steering vector associated with the signal arriving from $\theta_i$. The signal vector $\boldsymbol{s}_m(t)$ is considered as a

P.-J. Chung acknowledges support of her position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

stationary, temporally uncorrelated complex normal process with zero mean and covariance matrix $\boldsymbol{C}_s = E\boldsymbol{s}_m(t)\boldsymbol{s}_m(t)'$ where $(\cdot)'$ denotes the Hermitian transpose. The noise vector $\boldsymbol{n}(t)$ is a spatially and temporally uncorrelated complex normal process with zero mean and covariance matrix $\nu\boldsymbol{I}_n$ where $\nu$ is the noise spectral parameter and $\boldsymbol{I}_n$ is an $n \times n$ identity matrix. Thus, the array output $\boldsymbol{x}(t)$ is complex normally distributed with zero mean and covariance matrix

$$\boldsymbol{C}_x = \boldsymbol{H}_m(\boldsymbol{\theta}_m)\boldsymbol{C}_s\boldsymbol{H}_m(\boldsymbol{\theta}_m)' + \nu\boldsymbol{I}_n. \quad (3)$$

Based on array observations $\{\boldsymbol{x}(t)\}_{t=1}^T$ and a pre-specified number of signals, $m$, the ML estimate $\hat{\boldsymbol{\theta}}_m(T)$ is obtained by minimizing the *negative* concentrated likelihood function [1]

$$l_T(\boldsymbol{\theta}_m) = \log\det\big(\boldsymbol{P}(\boldsymbol{\theta}_m)\hat{\boldsymbol{C}}_x\boldsymbol{P}(\boldsymbol{\theta}_m)+\hat{\nu}\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\big), \quad (4)$$

$$\hat{\nu} = \frac{1}{n-m}\mathrm{tr}\big(\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\hat{\boldsymbol{C}}_x\big) \quad (5)$$

where $\boldsymbol{P}(\boldsymbol{\theta}_m)$ represents the projection matrix onto the column space of $\boldsymbol{H}_m(\boldsymbol{\theta}_m)$ and $\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m) = \boldsymbol{I}_n - \boldsymbol{P}(\boldsymbol{\theta}_m)$. $\hat{\boldsymbol{C}}_x = \frac{1}{T}\sum_{t=1}^T \boldsymbol{x}(t)\boldsymbol{x}(t)'$ denotes the sample covariance matrix. The problem of central interest is to estimate the DOA parameters when the true number of signals, $m_0$, is unknown.

The asymptotic behavior of the ML estimator depends strongly on the choice of $m$. For an underestimated $m$ ($< m_0$), the ML estimator provides partial information about the true parameter, $\boldsymbol{\theta}_0$, if widely separated sources are assumed. For an overestimated $m$ ($> m_0$), the ML estimate contains components that coincide with the true parameters. Motivated by this observation, we suggest an algorithm that requires only an upper bound on the number of signals rather than the exact knowledge of $m$.

## 3. USEFUL RESULTS

In this section, we briefly review useful properties of ML estimation under misspecified numbers of signals. In particular, we focus on the case in which the assumed number of signals, $m$, is larger than the true number of signals, $m_0$.

Based on the general theory of ML estimation of misspecified models [6, 7], a recent study [2] shows that under regularity conditions, for $m \neq m_0$, the ML estimator obtained by minimizing (4) converges to a well defined point $\boldsymbol{\theta}_m^*$ as $T$ increases. This limiting parameter $\boldsymbol{\theta}_m^*$ is characterized by the following properties.

**Property 1** The limiting point $\boldsymbol{\theta}_m^*$ minimizes a criterion similar to the concentrated likelihood function:

$$Q(\boldsymbol{\theta}_m) = \log\det\big(\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{C}_{x_0}\boldsymbol{P}(\boldsymbol{\theta}_m)+\tilde{\nu}\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\big), \quad (6)$$

$$\tilde{\nu} = \frac{1}{n-m}\mathrm{tr}\big(\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\boldsymbol{C}_{x_0}\big), \quad (7)$$

where $\boldsymbol{C}_{x_0}$ represents the true covariance matrix of the array output. It can be computed from (3) at the true parameters $\boldsymbol{\theta}_0$, $\boldsymbol{C}_{s0}$ and $\nu_0$.

**Property 2** For $m > m_0$, the *assumed* signal subspace $\mathrm{sp}\{\boldsymbol{H}(\boldsymbol{\theta}_m^*)\}$ is related to the *true* signal subspace $\mathrm{sp}\{\boldsymbol{H}_{m0}(\boldsymbol{\theta}_0)\}$ as follows:

$$\mathrm{sp}\{\boldsymbol{H}_m(\boldsymbol{\theta}_m^*)\} \supset \mathrm{sp}\{\boldsymbol{H}_{m0}(\boldsymbol{\theta}_0)\}, \quad (8)$$

where $\mathrm{sp}(\boldsymbol{A})$ denotes the column space of a matrix $\boldsymbol{A}$.

**Property 3** For $m > m_0$, the minimum value of $Q(\boldsymbol{\theta}_m)$ is equal to that of $Q(\boldsymbol{\theta}_0)$, i.e.

$$Q(\boldsymbol{\theta}_m^*) = Q(\boldsymbol{\theta}_0). \quad (9)$$

Property 3 implies that for an overestimated model order, the value of $Q(\boldsymbol{\theta}_m^*)$ does not change with increasing number of parameters.

## 4. ROBUST ML ESTIMATION

From Property 2 in Section 3 we learn that if we assume more signals than necessary, the ML estimator results in a signal subspace that contains the true signal subspace. This implies that given proper array manifold, $\boldsymbol{\theta}_m^*$ contains $m_0$ components that coincide with those of the true parameter vector $\boldsymbol{\theta}_0$. Motivated by this observation, we proposed the following procedure that requires only an upper bound on the number of signals and still provides reliable results.

Let $M$ denote an upper bound on the number of signals and $\hat{\boldsymbol{\theta}}_M$ be the ML estimator obtained by minimizing the negative log-likelihood function (4) with $m = M$:

$$\hat{\boldsymbol{\theta}}_M = \arg\min_{\boldsymbol{\theta}_M} l_T(\boldsymbol{\theta}_M). \quad (10)$$

Since $M \geq m_0$, the $M \times 1$ vector $\hat{\boldsymbol{\theta}}_M = [\hat{\theta}_1, \ldots, \hat{\theta}_M]^T$ contains more elements than the $m_0 \times 1$ true parameter vector $\boldsymbol{\theta}_0$ does. As discussed previously, for large $T$, a subset of the elements in $\hat{\boldsymbol{\theta}}_M$ may coincide with those of $\boldsymbol{\theta}_0$. The elements of $\hat{\boldsymbol{\theta}}_M$ that are associated with those of $\boldsymbol{\theta}_0$, are referred to as *relevant* estimates. According to Property 2 and Property 3, the remaining $(M - m_0)$ components of $\boldsymbol{\theta}_M^*$ are not predictable and do not change the value of the criterion (6). Thus the next step is to identify the *relevant* elements of $\hat{\boldsymbol{\theta}}_M$.

Motivated by Property 3, we define the *relevance* value for each component of $\hat{\boldsymbol{\theta}}_M$ as follows:

$$R(\hat{\theta}_i) = \frac{1}{l_T(\hat{\boldsymbol{\theta}}_M)}[l_T(\tilde{\boldsymbol{\theta}}_i) - l_T(\hat{\boldsymbol{\theta}}_M)], \quad (11)$$

where $\tilde{\boldsymbol{\theta}}_i$ contains all elements of $\hat{\boldsymbol{\theta}}_M$ except the $i$th component $\hat{\theta}_i$:

$$\tilde{\boldsymbol{\theta}}_i = [\hat{\theta}_1, \ldots, \hat{\theta}_{i-1}, \hat{\theta}_{i+1}, \ldots, \hat{\theta}_M]. \quad (12)$$

The log-likelihood $l_T(\tilde{\boldsymbol{\theta}}_i)$ is computed by (4) using the $(M-1)$ dimensional vector $\tilde{\boldsymbol{\theta}}_i$. The value of $R(\hat{\theta}_i)$ indicates the contribution of the $i$th element $\hat{\theta}_i$ to the log-likelihood function. We use the normalizing factor $1/l_T(\hat{\boldsymbol{\theta}}_M)$ to improve numerical stability. Comparing (4) and (6), we notice that $l_T(\hat{\boldsymbol{\theta}}_m)$ is an estimate for $Q(\boldsymbol{\theta}_m^*)$.

From Property 3, we know that the $(M - m_0)$ elements of $\boldsymbol{\theta}_M^*$ that are not associated with the true parameter $\boldsymbol{\theta}_0$ do not change the value of $Q(\boldsymbol{\theta}_0)$. Thus, a high relevance value $R(\hat{\theta}_i)$ implies that $\hat{\theta}_i$ is associated with one of $\boldsymbol{\theta}_0$'s components. A low relevance value (usually close to zero) is an indication that the underlying estimate does not correspond to any true parameter.

Let $R_{(1)} \geq R_{(2)} \geq \ldots \geq R_{(M)}$ denote the ordered relevance values with corresponding $\{\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(M)}\}$. Using this ordering, one can easily recognize the significance of each individual estimate. Furthermore, we can choose the relevance estimates if the associated relevance values exceed a given threshold $\epsilon$. More precisely, the set of relevant estimates is given by

$$S = \{\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(k)}\} \quad (13)$$

where

$$R_{(1)} \geq R_{(2)} \geq \ldots \geq R_{(k)} \geq \epsilon. \quad (14)$$

Note that the number of components in $S$ provides indirectly an estimate for the number of signals. How to choose the threshold $\epsilon$ in a systematic way is still under investigation.

In summary, given an upper bound on the number of the signals $M$, the proposed algorithm proceeds as follows.

Input: $\{\boldsymbol{x}(t) : t = 1, \ldots, T\}$, $M$, $\epsilon$.

1. Find the ML estimate $\hat{\boldsymbol{\theta}}_M$ by (10).
2. Compute the relevance values
   $$R(\hat{\theta}_i) = l_T(\tilde{\boldsymbol{\theta}}_i) - l_T(\hat{\boldsymbol{\theta}}_M), \quad i = 1, \ldots, M.$$
3. Use the ordered relevance values
   $$R_{(1)} \geq R_{(2)} \geq \ldots \geq R_{(k)} \geq \epsilon$$
   to find the set of relevant estimates
   $$S = \{\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(k)}\}.$$

Output: $\{\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(k)}\}$

Table 1: Robust ML Estimation Algorithm for Unknown Numbers of Signals.

## 5. SIMULATION

In this section, we apply the robust ML estimation to simulated data and assess its performance. In particular, we shall investigate whether the relevant estimates provide useful information about the true parameters.

In our experiment, a uniform linear of 10 sensors with inter-element spacings of half a wavelength $\lambda/2$ is employed. The narrow band signals are generated by $m_0 = 2$ uncorrelated signals located at $[\ 28^\circ\ 36^\circ]$ of various strengths. The two signals are separated about half of the beamwidth. The difference of signal strengths is $[\ 1 \quad 0\ ]$ dB where 0 dB corresponds to the reference signal. The signal to noise ratio (SNR), defined as $10\log\left(E(|s_i(t)|^2)/\nu\right)$ for the $i$th signal, varies from $-6$ to 10 dB in a 2 dB step. We generate $T = 100$ snapshots for each of the 200 trials performed. The upper bound on the number of signals is chosen to be $M = 4$. The threshold $\epsilon$ is chosen to be $0.01$. For comparison, we apply the ML approach to the same batch of data using the correct number of signals, $m_0 = 2$. The estimates obtained under the true number of signals are denoted by $\hat{\boldsymbol{\theta}}_0 = [\hat{\theta}_{01}\ \ \hat{\theta}_{02}]^T$.

Fig. 1 shows the ordered relevance values averaged over 200 trials. The first and second largest relevance values, $R_{(1)}$ and $R_{(2)}$, are much higher than the remaining relevance values, $R_{(3)}$ and $R_{(4)}$. This is a strong indication that the corresponding estimates are associated with the true DOA parameters. Furthermore, one can observe that while the largest two relevance values become larger with increasing SNRs, the two smallest relevance values decrease slightly. The difference between the relevance values associated with relevant estimates and the remaining relevance values become larger at high SNRs.

The sample mean and standard deviation of the estimates $\hat{\theta}_{(i)}$, $i = 1, \ldots, 4$ and $\hat{\theta}_{0j}$, $j = 1, 2$ are plotted in Fig. 2 and Fig. 3, respectively. In Fig. 2, one can observe that the components of $\hat{\boldsymbol{\theta}}_M$ corresponding to the two largest relevance values, $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$, have the same mean value as the ML estimates obtained under the true number of signals, $m_0 = 2$. Furthermore, the average values of $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$ remain constant over the entire SNR range and coincide with the true parameter vector $\boldsymbol{\theta}_0 = [\ 28^\circ\ 36^\circ]$. The mean values of redundant estimates, $\hat{\theta}_{(3)}$ and $\hat{\theta}_{(4)}$, lie between $-20^\circ$ and $10^\circ$ and have large variations over different SNRs.

In Fig. 3, the relevant estimates, $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$ show a slightly higher variance than those obtained under the true number of signals. Therefore, if one chooses the relevant estimates correctly, the performance is only slightly worse than the standard ML method. We can also observe that the redundant estimates, $\hat{\theta}_{(3)}$ and $\hat{\theta}_{(4)}$, have a much higher variance than $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(2)}$. This implies that the redundant estimates spread over a large interval of the parameter space.

As shown in Fig. 2 and Fig. 3, the relevant value is a good criterion for finding the estimates that correspond to the true parameters. Using the threshold $\epsilon = 0.01$, we have a

| | Probability of Correct Detection | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SNR | $-6$ | $-4$ | $-2$ | $0$ | $2$ | $4$ | $6$ | $8$ |
| Prob. | $0.98$ | $0.98$ | $0.96$ | $0.97$ | $0.96$ | $0.97$ | $0.96$ | $0.99$ |

**Table 1**. Empirical Probability of Correct Detection vs. SNR.

high probability of correct detection. By *correct detection* we mean that the number of relevant estimates equals the true number of signals, i.e. $k = 2$. Table 1 shows that the empirical probability of correct detection is as high as $0.98$ at SNR as low as $-6$ dB. Overall, the probability of correct detection is higher than $95\%$.

In summary, the proposed algorithm provides good estimation performance without knowing exact numbers of signals. Based on the relevance values, we can easily choose the set of relevant estimates that coincide with the true parameters. Furthermore, with high probability, the number of selected relevant components is equal to the true number of signals. Thus, the robust ML approach provides accurate estimates of both the parameters of interest and the number of signals.

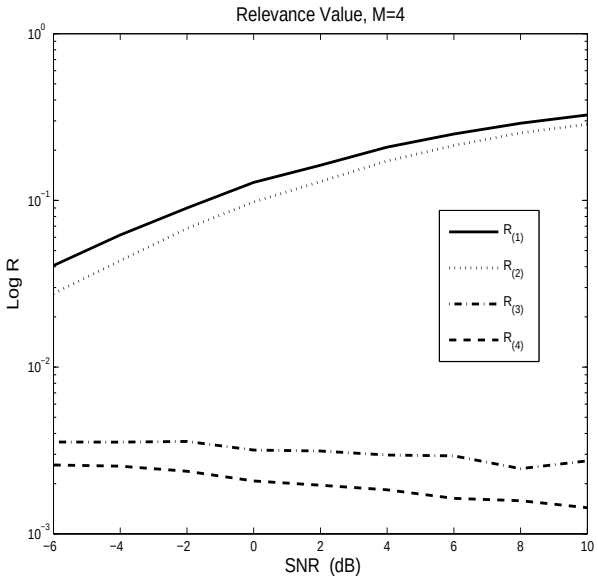


**Fig. 1**. Relevance value. Upper bound on the number of signals $M = 4$. The true DOA parameter $\boldsymbol{\theta}_0 = [28^\circ\ 36^\circ]$, SNR $= [-6 : 2 : 10]$ dB.

## 6. CONCLUSION

We propose a novel robust ML method for unknown numbers of signals. The proposed algorithm computes the ML estimates using an upper bound on the number of signals. To select the relevant components from the estimated parameter vector, we define the relevance value which measures the contribution of each element to the likelihood function. The set of relevant estimates consists of components with relevance values larger than a given threshold. Numerical results show that the proposed algorithm achieves comparable estimation accuracy as the standard ML approach does. Furthermore, the number of signals can be accurately determined by the number of relevant estimates. Compared to conventional methods based on the information theoretic criterion or multiple tests, the proposed algorithm provides a computationally more attractive solution.

## 7. REFERENCES

[1] Johann F. Böhme. Statistical array signal processing of measured sonar and seismic data. In *Proc. SPIE 2563 Advanced Signal Processing Algorithms*, pages 2–20, San Diego, Jul 1995.

[2] P.-J. Chung. Stochastic maximum likelihood estimation under misspecified numbers of signals. *IEEE Trans. Signal Processing*, 55(8), August 2007. to be published.

[3] D. Maiwald. *Breitbandverfahren zur Signalentdeckung und –ortung mit Sensorgruppen in Seismik– und Sonaranwendungen*. Dr.–Ing. Dissertation, Dept. of Electrical Engineering, Ruhr–Universität Bochum, Shaker Verlag, Aachen, 1995.

[4] P.-J.Chung, J. F. Böhme, C. F. Mecklenbräuker, and A. O. Hero. Detection of the number of signals using the Benjamini-Hochberg procedure. *IEEE Trans. Signal Processing*, 55(6):2497–2508, June 2007.

[5] M. Wax and I. Ziskind. Detection of the number of coherent signals by the MDL principle. *IEEE ASSP*, 37(8):1190–1196, August 1989.

[6] H. White. Maximum likelihoood estimation of misspecified models. *Econometrica*, 50(1):1–25, Jan 1982.

[7] H. White. *Estimation, Inference and Specification Analysis*. Cambridge University Press, 1996.

[8] M. Wong, Q. T. Zou, and J. P. Reilly. On information theoretic criterion for determining the number of signals in high resolution array processing. *IEEE Trans. Acoust., Speech, Signal Processing*, 38(11):1959–1971, November 1990.

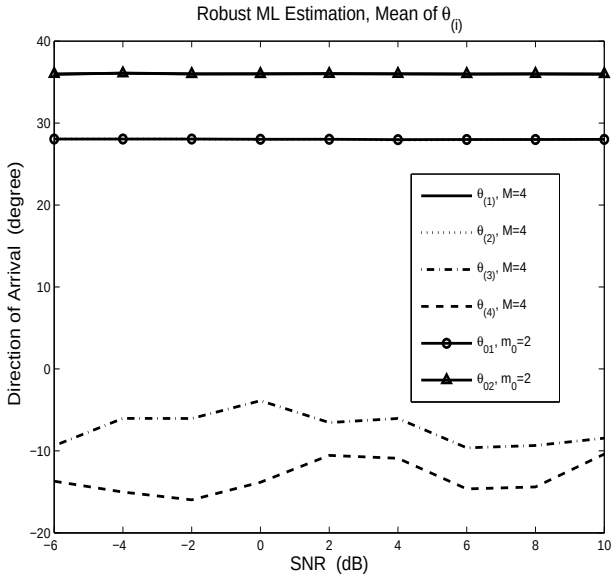


**Fig. 2**. Mean value of $\hat{\theta}_{(i)},\ i = 1, \ldots, M$.

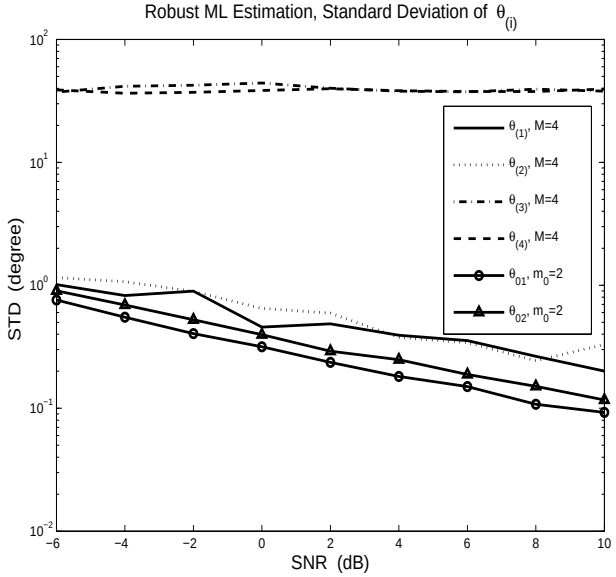


**Fig. 3**. Standard deviation of $\hat{\theta}_{(i)},\ i = 1, \ldots, M$.