

A)  $\sigma^2$  is known and hyperparameter  $\alpha$  is deterministic.

First of all, we consider the ideal case that the variance of noise  $\sigma^2$  is known and the hyperparameter  $\alpha = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_{2K})$  which is associated to the source is deterministic. In this case, the source to be estimated is the only unknown variable, we let  $\mathbf{x} = [\mathbf{s}]$ , where  $\mathbf{s}$  denotes the source to be estimated. With the Bayes' rule and the assumption for  $\sigma^2$  and  $\alpha$ , the likelihood of the measurement  $\mathbf{y}$  can be written as

$$P(\mathbf{y}, \mathbf{s}, \alpha, \sigma^2) = P(\mathbf{y}|\mathbf{s}, \sigma^2)P(\mathbf{s}|\alpha)P(\alpha)P(\sigma^2), \quad (1)$$

where  $P(\mathbf{y}|\mathbf{s})$  is the likelihood of the measurement and  $P(\mathbf{s}|\alpha)$  is the prior estimate for the source which appears in the first stage of the hierarchical model,  $P(\alpha)$  and  $P(\sigma^2)$  are constant under the assumption and they will be ignored in future analysis. The natural logarithm of the likelihood (1) is written as

$$\log P(\mathbf{y}, \mathbf{s}, \alpha, \sigma^2) = \log P(\mathbf{y}|\mathbf{s}) + \log P(\mathbf{s}|\alpha), \quad (2)$$

the expression for  $P(\mathbf{y}|\mathbf{s})$  is derived in () and  $P(\mathbf{s}|\alpha)$  is derived in (), (2) can be written as

$$\begin{aligned} \log P(\mathbf{y}, \mathbf{s}, \alpha, \sigma^2) = & -M \log \frac{1}{2\pi} - M \log(\sigma) - \frac{(\mathbf{y} - \mathbf{A}\mathbf{s})^T(\mathbf{y} - \mathbf{A}\mathbf{s})}{2\sigma^2} \\ & - 2K \log \frac{1}{2\pi} - \sum_{i=1}^{2k} \log \sqrt{\alpha_i} - \sum_{i=1}^{2k} \frac{s_i^2}{2\alpha_i}, \end{aligned} \quad (3)$$

where  $M$  is the number of elements and  $2K$  is the number of entries in  $\mathbf{s}$ , the derivative of  $\log P(\mathbf{y}, \mathbf{s}, \alpha)$  with respect to  $\mathbf{x}$  is

$$\frac{\partial \log P(\mathbf{y}, \mathbf{s}, \alpha, \sigma^2)}{\partial x} = \frac{\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{s})}{\sigma^2} - \sum_{i=1}^{2k} \frac{s_i}{\alpha_i}, \quad (4)$$

from (4), we can have the second derivative of the likelihood with respect to  $\mathbf{x}$ ,

$$\frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}, \alpha, \sigma^2)}{\partial x^2} = -\frac{\mathbf{A}^T \mathbf{A}}{\sigma^2} - \alpha^{-1}. \quad (5)$$

According to the definition of the Fisher's matrix (), the Fisher's matrix is obtained by taking the negative expectation,

$$I(\mathbf{x}) = \frac{\mathbf{A}^T \mathbf{A}}{\sigma^2} + \alpha^{-1}, \quad (6)$$

from (6), it can be seen the Fisher's matrix depends on the measurement conditions (i.e.,  $\mathbf{A}$  and  $\sigma^2$ ) and the hyperparameters. It is known that the lower bound of the estimated variance for the unknown variable is inversely proportional to the Fisher's matrix. Therefore, it can be concluded that the bound is reduced with the lower  $\sigma^2$ , the more diagonally dominant matrix  $\mathbf{A}^T \mathbf{A}$  and  $\alpha$  with small values.

B)  $\sigma^2$  is known and hyperparameter  $\alpha$  is random.

Secondly, the sources are drawn from the prior distribution () and the hyperprior distribution is defined in () with the hyperparameter  $\lambda$ . The relationship between the source and the hyperparameter  $\lambda$  is obtained by marginalizing out  $\alpha$  where  $\alpha$  is seen as a nuisance parameter, with using the identity (), we can see the distribution of the source is a Laplace distribution with  $\lambda$ , the distribution is derived in () :

$$P(\mathbf{s}|\lambda) = \frac{\lambda^K}{2^{2K}} \exp(-\sqrt{\lambda} \sum_{i=1}^{2K} |s_i|). \quad (7)$$

In this case, there are two unknown variables:  $\mathbf{s}$  and  $\lambda$ , we use  $\mathbf{x} = [\mathbf{s}^T, \lambda]^T$  to denote the two unknown variables. By the definition of the Fisher matrix, it now consists of four blocks,

$$I(\mathbf{x}) = \begin{bmatrix} I(\mathbf{s}) & I(\mathbf{s}, \lambda) \\ I(\mathbf{s}, \lambda)^T & I(\lambda) \end{bmatrix}, \quad (8)$$

$I(\mathbf{s}) \in \mathbb{R}^{2k \times 2K} = -\mathbb{E}_{\mathbf{x}} \left\{ \frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)}{\partial \mathbf{s}^2} \right\}$ ,  $I(\mathbf{s}, \lambda) \in \mathbb{R}^{2k \times 1} = -\mathbb{E}_{\mathbf{x}} \left\{ \frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)}{\partial \mathbf{s} \partial \lambda} \right\}$  and  $I(\lambda) = -\mathbb{E}_{\mathbf{x}} \left\{ \frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)}{\partial \lambda^2} \right\}$ .  $I(\mathbf{s}, \lambda)$  is calculated by the mixed partial derivative, it indicates the correlations between  $\mathbf{s}$  and  $\lambda$  instead of the variance of these unknown parameters, therefore the expression for  $I(\mathbf{s}, \lambda)$  will not be derived. The joint distribution is obtained by combining the likelihood in () and the distribution in (),

$$P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda) \propto P(\mathbf{y}|\mathbf{s}, \sigma^2)P(\mathbf{s}|\lambda)P(\lambda), \quad (9)$$

the natural logarithm of the joint distribution is

$$\log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda) = -\frac{(\mathbf{y} - \mathbf{A}\mathbf{s})^T(\mathbf{y} - \mathbf{A}\mathbf{s})}{2\sigma^2} + (K+1)\log \lambda - \sqrt{\lambda} \sum_{i=1}^{2K} |s_i|, \quad (10)$$

since  $|s_i|$  appears in  $\log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)$ , that makes  $\log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)$  not differentiable with respect to  $s_i$  when  $s_i = 0$ , here we ignore the case of  $s_i = 0$  and consider the other two situations,

$$\frac{\partial \log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)}{\partial s_i} = \begin{cases} \frac{\mathbf{A}_i^T(\mathbf{y} - \mathbf{A}\mathbf{s})}{\sigma^2} - \sqrt{\lambda} & \text{if } s_i > 0 \\ \frac{\mathbf{A}_i^T(\mathbf{y} - \mathbf{A}\mathbf{s})}{\sigma^2} + \sqrt{\lambda} & \text{if } s_i < 0 \end{cases}, \quad (11)$$

the second term at the right hand side of (11) is independent on  $s_i$  under the two situations, therefore we can obtain the second derivative of  $\log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)$  with respect to  $\mathbf{s}$  as

$$\frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)}{\partial \mathbf{s}^2} = -\frac{\mathbf{A}^T \mathbf{A}}{\sigma^2}, \quad (12)$$

$I(\mathbf{s})$  is derived by taking the negative expectation of (12),

$$I(\mathbf{s}) = \frac{\mathbf{A}^T \mathbf{A}}{\sigma^2}. \quad (13)$$

It is clear to see compared with bound derived in (6), the tighter bound for the variance of estimated  $s$  is provided in this case since

$$\left(\frac{\mathbf{A}^T \mathbf{A}}{\sigma^2}\right)^{-1} \succeq \left(\frac{\mathbf{A}^T \mathbf{A}}{\sigma^2} + \boldsymbol{\alpha}^{-1}\right)^{-1}, \quad (14)$$

another thing we noticed is with  $\alpha$  being marginalized out, the blocks related to  $s$  can be computed without knowing the realization of  $\alpha$ . The derivative of  $\log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)$  with respect to  $\lambda$  is

$$\frac{\partial \log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)}{\partial \lambda} = \frac{K+1}{\lambda} - \frac{1}{2\sqrt{\lambda}} \sum_{i=1}^{2K} |s_i|, \quad (15)$$

then we can have the second derivative is

$$\frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}, \sigma^2; \lambda)}{\partial \lambda^2} = -\frac{K+1}{\lambda^2} + \frac{1}{4} \lambda^{-\frac{3}{2}} \sum_{i=1}^{2K} |s_i|, \quad (16)$$

we take the negative expectation of (16) to obtain  $I(\lambda)$  :

$$I(\lambda) = -\mathbb{E} \left\{ -\frac{K+1}{\lambda^2} + \frac{1}{4} \lambda^{-\frac{3}{2}} \sum_{i=1}^{2K} |s_i| \right\} = \frac{K+1}{\lambda^2} - \frac{1}{4} \lambda^{-\frac{3}{2}} \mathbb{E} \left( \sum_{i=1}^{2K} |s_i| \right), \quad (17)$$

with the derived distribution for source in (), we can calculate the expected value of  $|s_i|$

$$\mathbb{E}(|s_i|) = \int_{-\infty}^{\infty} |s_i| \cdot p(s_i, \lambda) ds_i = \int_{-\infty}^{\infty} |s_i| \frac{\sqrt{\lambda}}{2} \exp(-\sqrt{\lambda}|s_i|) ds_i \quad (18)$$

$$= 2 \int_0^{\infty} s_i \frac{\sqrt{\lambda}}{2} \exp(-\sqrt{\lambda}s_i) ds_i \quad (19)$$

$$= \sqrt{\lambda} \int_0^{\infty} s_i \exp(-\sqrt{\lambda}s_i) ds_i \quad (20)$$

$$= \frac{1}{\sqrt{\lambda}}, \quad (21)$$

plugging (21) into (17), we have

$$I(\lambda) = -\mathbb{E} \left\{ -\frac{K+1}{\lambda^2} + \frac{1}{4\lambda^{-\frac{3}{2}}} \sum_{i=1}^{2K} |s_i| \right\} = \frac{K+2}{2\lambda^2}. \quad (22)$$

In this case, the lower bound of the variance of sources ( $\mathbf{s}$ ) only depends on measurement matrix  $\mathbf{A}$  and noise variance  $\sigma^2$ , the lower bound of the variance for hyperparameter  $\lambda$  is proportional on the square of its true value and inversely proportional to the size of the angle grid  $K$ .

c)  $\sigma^2$  is unknown

In this case, the variance of noise  $\sigma^2$  is unknown, we use  $\mathbf{x} = [\mathbf{s}^T, \sigma^2]^T$  to denote the unknown variables. Now the Fisher matrix is written as

$$I(\mathbf{x}) = \begin{bmatrix} I(\mathbf{s}) & I(\mathbf{s}, \sigma^2) \\ I(\mathbf{s}, \sigma^2)^T & I(\sigma^2) \end{bmatrix}, \quad (23)$$

for the four blocks of the Fisher matrix,  $I(\mathbf{s})$  has been derived in (6) when  $\alpha$  is deterministic and (13) when  $\alpha$  is drawn from the specific distribution,  $I(\mathbf{s}, \sigma^2)$  is a zero vector which is not related to the variance of the estimated variable. Therefore, we focus on the block  $I(\sigma^2)$ , let  $\eta = \sigma^2$ ,

$$I(\eta) = -\mathbb{E} \left\{ \frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}; \eta)}{\partial \eta^2} \right\} \quad (24)$$

the second derivative of  $\log P(\mathbf{y}, \mathbf{s}; \eta)$  with respect to  $\eta$  is

$$\frac{\partial^2 \log P(\mathbf{y}, \mathbf{s}; \eta)}{\partial \eta^2} = -\frac{(\mathbf{y} - \mathbf{A}\mathbf{s})^T (\mathbf{y} - \mathbf{A}\mathbf{s})}{\eta^3}, \quad (25)$$

since  $\mathbf{y} - \mathbf{A}\mathbf{s} = \mathbf{n}$ , each entry in  $\mathbf{n}$  is assumed to be independent with each other and it is subject to normal distribution with mean zero and variance  $\eta$ , therefore we can have

$$\mathbb{E} \{ (\mathbf{y} - \mathbf{A}\mathbf{s})^T (\mathbf{y} - \mathbf{A}\mathbf{s}) \} = 2M\eta, \quad (26)$$

where  $M$  is the number of elements. Using (26), we can compute  $I(\eta)$  as :

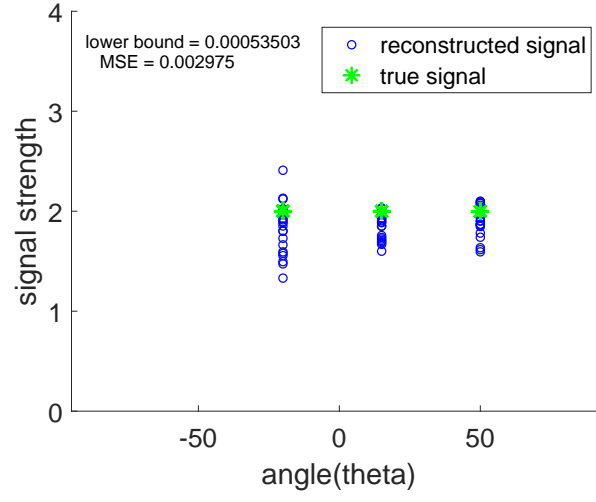
$$I(\eta) = \frac{2M}{\eta^2}. \quad (27)$$

In this case, the lower bound of estimated  $\sigma^2$  is proportional on square of its true variance and inversely proportional to the number of the elements in the receiving system  $M$ .

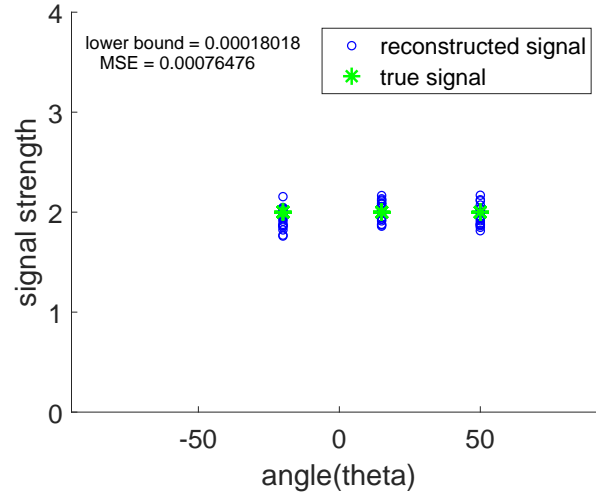
d) multiple measurements

Simulation Results

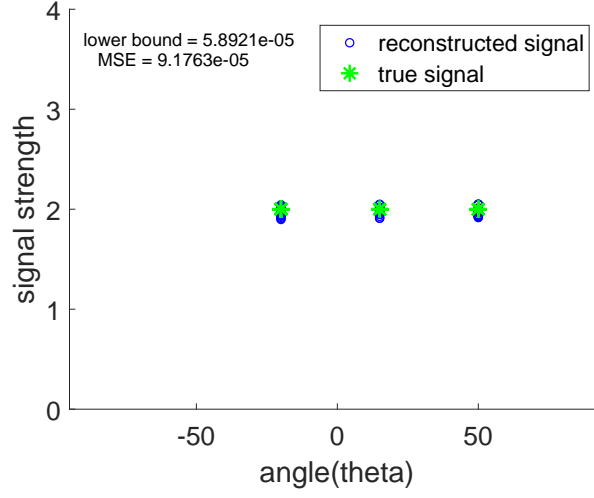
The estimated results with various SNR are presented in Figs. 1 - 4. The normal non-linear array with 12 elements is deployed and 20 experiments are conducted for each case with keeping the incoming source same. In the figures, only the estimated sources at correct angles are plotted for the clarity, but the estimated results at wrong angles are considered when the MSE is calculated with (). The value of lower bound ( i.e., inverse of (13)) and MSE in each case are both indicated in the figures. From the figures it is seen that the MSE is always larger than the lower bound and the MSE drops with the increase of SNR.



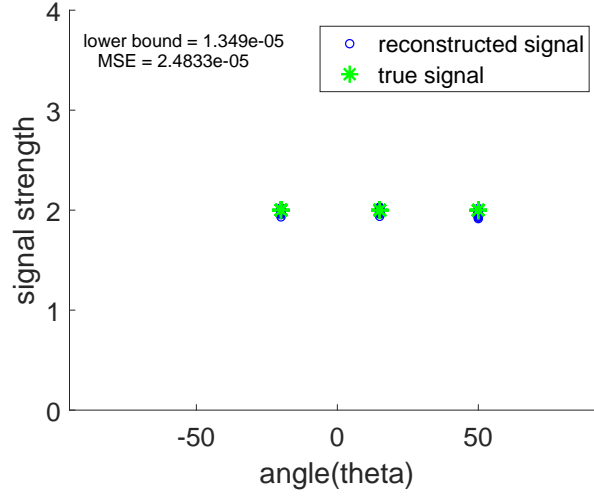
**Figure 1.** SNR = 15 dB,  $N = 3$ ,  $M = 12$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.



**Figure 2.** SNR = 20 dB,  $N = 3$ ,  $M = 12$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.



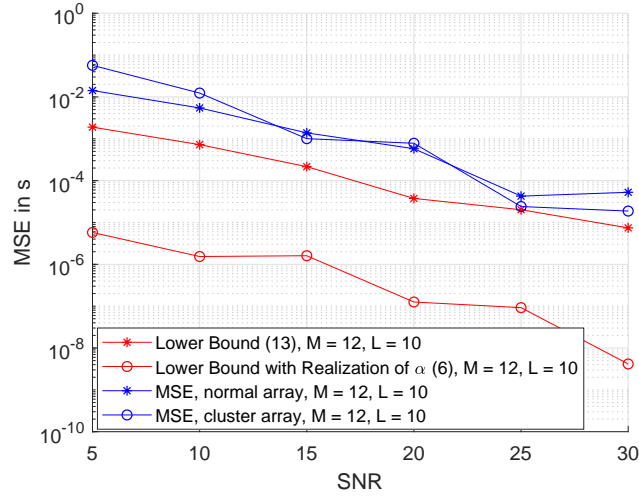
**Figure 3.** SNR = 25 dB,  $N = 3$ ,  $M = 12$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.



**Figure 4.** SNR = 30 dB,  $N = 3$ ,  $M = 12$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.

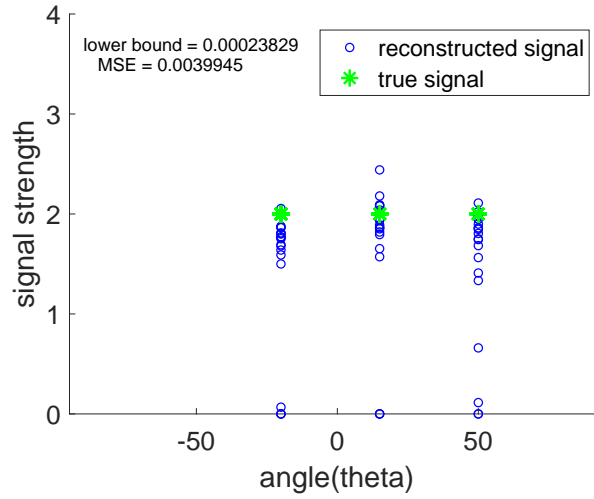
The cluster non-linear array is also used to test the effect of noise to the variance of estimated sources.

Fig. 5 presents performance of the MSE of source under different factors.

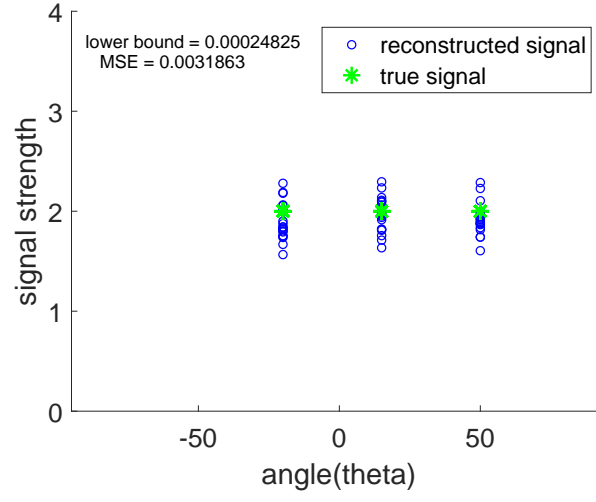


**Figure 5.**  $N = 3$ ,  $L = 10$

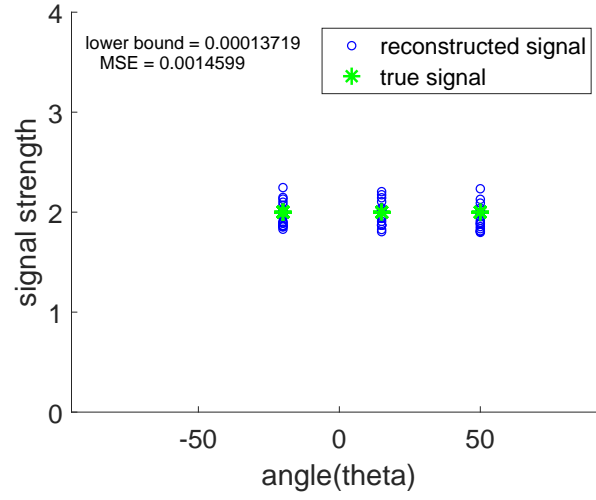
Figs. 6 - 9 show the estimated results with different number of elements in the array, the normal non-linear array is used and all factors are kept same except the number of elements, in each case we conducted 20 experiments to plot the figures. It can be seen that the variance of the estimated sources decreases with the increase of the number of elements in the array.



**Figure 6.**  $\text{SNR} = 20 \text{ dB}$ ,  $N = 3$ ,  $M = 6$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.

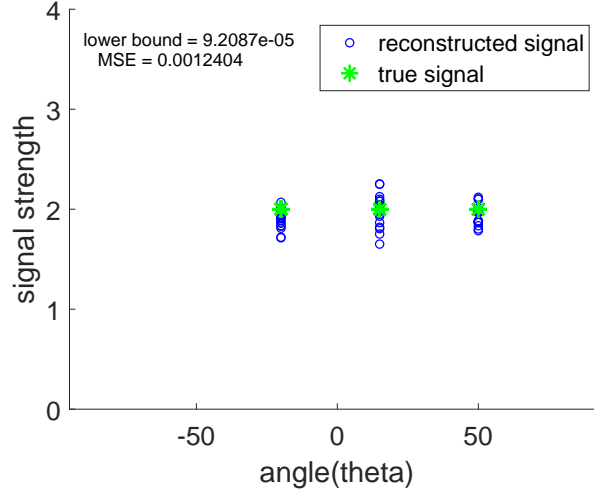


**Figure 7.** SNR = 20 dB,  $N = 3$ ,  $M = 9$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.



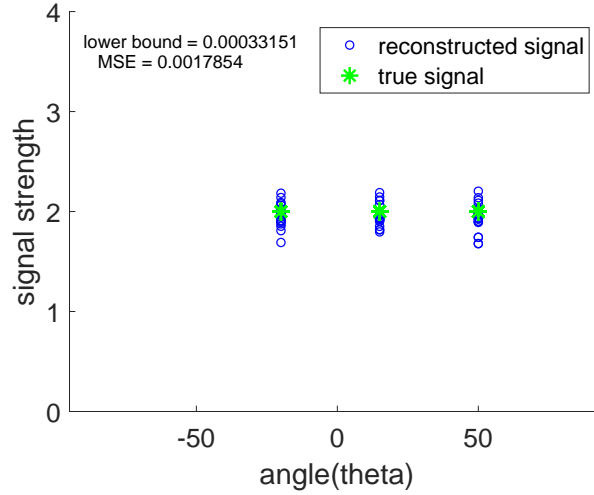
**Figure 8.** SNR = 20 dB,  $N = 3$ ,  $M = 12$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.



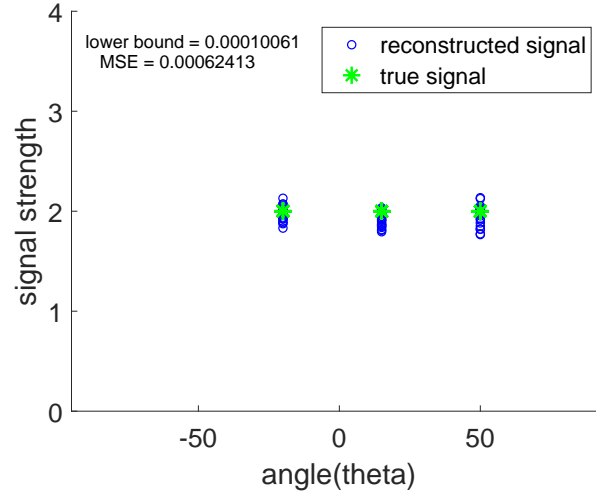


**Figure 9.** SNR = 20 dB,  $N = 3$ ,  $M = 15$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.

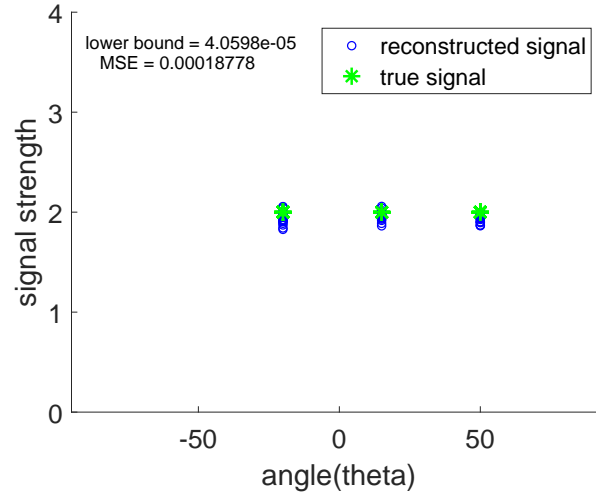
Figs. 10 - 13 show the estimated results with different number of measurements, in each measurement, the position of element could change which corresponds to the second case in the subsection of the effect of the number of measurements to the estimated results. The normal non-linear array is used and all factors are kept same except the number of measurements, in each case we conducted 20 independent experiments to plot the figures. We can see that the variance of the estimated sources decreases with the increase of the number of measurements which means multiple measurements can provide the more robust results.



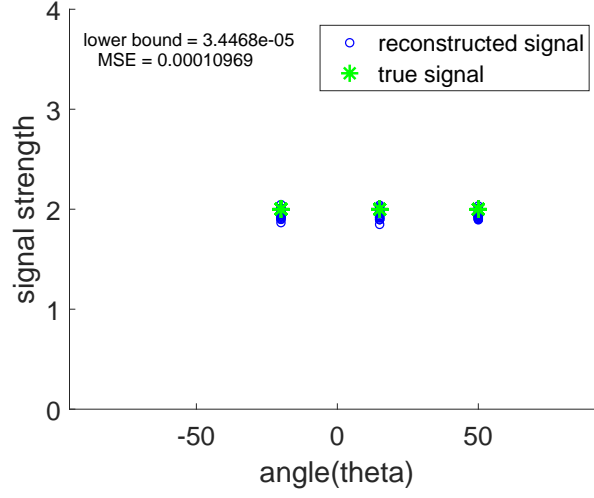
**Figure 10.** SNR = 20 dB,  $N = 3$ ,  $M = 12$ ,  $L = 5$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.



**Figure 11.** SNR = 20 dB,  $N = 3$ ,  $M = 12$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.

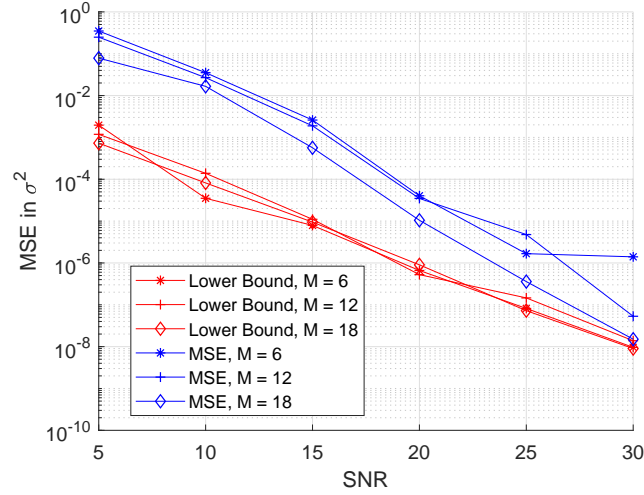


**Figure 12.** SNR = 20 dB,  $N = 3$ ,  $M = 12$ ,  $L = 25$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.

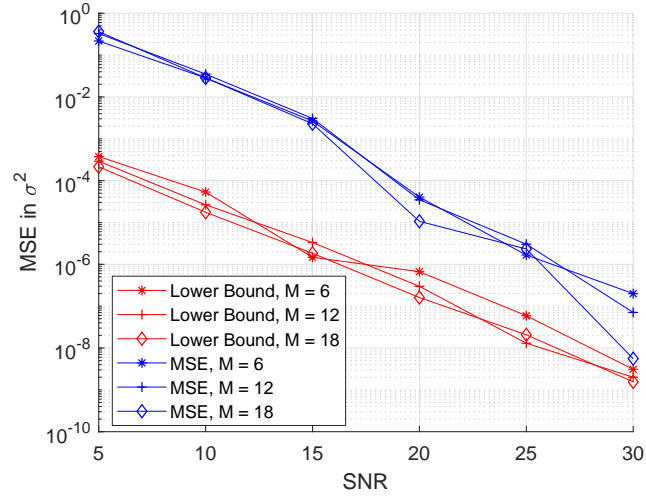


**Figure 13.** SNR = 20 dB,  $N = 3$ ,  $M = 12$ ,  $L = 50$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.

The performance of MSE in  $\sigma^2$  that estimated with 10 and 50 measurements and normal non-linear array are plotted in Figs. 14 and 15, respectively. The lower bound of the variance of estimated  $\sigma^2$  under different SNR and receiving antenna array with different number of elements are presented. When the number of measurement number is fixed, it can be seen the lower bound and actual MSE both drops with the increase of SNR which matches the expression derived in (27) and they also show more receiving elements is helpful to obtain a better estimation for  $\sigma^2$ . Compared Fig. 14 with Fig.15, we can see the MSE performance of  $\sigma^2$  is better with more measurements at most cases.



**Figure 14.**  $N = 3$ ,  $L = 10$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.



**Figure 15.**  $N = 3$ ,  $L = 50$ ,  $\Delta s = 24\lambda$ ,  $r = 6\lambda$ , normal array.