Direction of Arrival Estimation

DSP Lab Course WS 2021/22

Authors:

Heng Lu (heng.lu@fau.de), Mengya Wang (mia.wang@fau.de)

Supervisor: Thomas Haubner

Introduction

The goal of this project is to estimate the direction of arrival (DoA) of the sound source relative to the microphone array. The direction of arrival estimation is achieved by processing the received signal to obtain distance information and orientation information about the target. The physics of this is the time delay of the sound waves impinging on the array, as shown in Figure 1. Therefore, different directions of arrival result in different time delays between microphones.

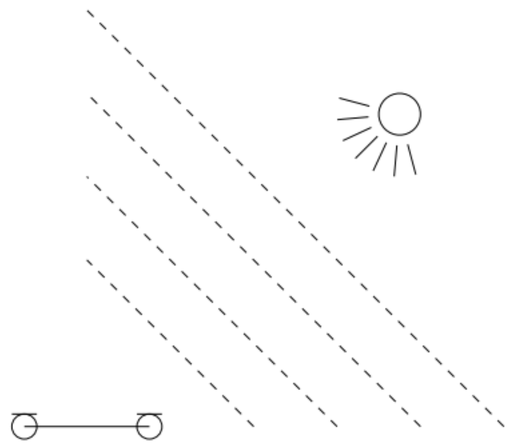

Figure 1: Acoustical waves impinging on microphone array.

The simulated environmental conditions for the project were the presence of one sound source (i.e. signal emitter) and six microphones (i.e. signal receiver) in an enclosed room (room dimensions: length: 7 metres, width: 5 metres, height: 2.75 metres). The project will perform DoA estimation by the MUSIC algorithm in the simulation environment given above.

Algorithm

Introduction to the model. Fundamentals of direction of arrival estimation using the Multiple signal classification (MUSIC) algorithm,Model for linear arrays Assuming a single-input multiple-output transmission model with a total of M uniform linear antenna arrays and a distance d between each receiving antenna real element, let a transmit signal fit the far-field signal model, i.e.: the transmit source is far away from the receiver and the signal received by the receiver can be regarded as parallel, the model schematic is shown in Figure 2.

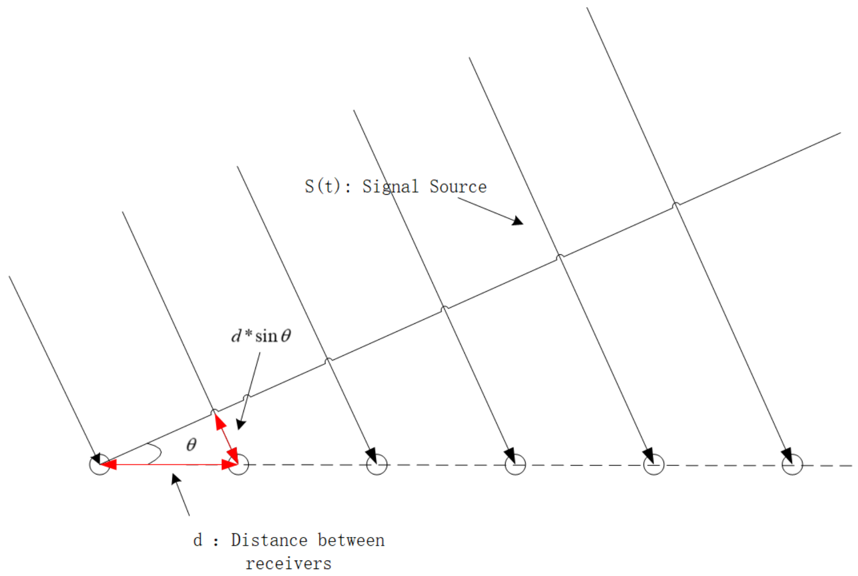

Figure 2: Linear array model.

From the diagram we can see that the signal arriving at the second receiving antenna array element travels $dsin\theta$ more than it does at the first antenna array element, i.e. the signal arriving at the latter array element will travel $dsin\theta$ more than the previous one. Assuming that the speed of the signal is v and the wavelength is λ, the time difference between the arrival of the signal at the two adjacent array elements can be obtained as,

$$\delta t = \frac{dsin\theta}{v} \tag{1}$$

Let the frequency of the signal be f_0, the phase difference between the two arrays can be obtained as,

$$\delta\phi = \frac{2\pi f_0 dsin\theta}{v} = \frac{2\pi dsin\theta}{\lambda} \tag{2}$$

Let the transmitted signal be $s(t)$ and the noise received during transmission be $n(t)$, the signal of the whole received array can be expressed as,

$$x(t) = \begin{bmatrix} 1 & e^{j2\pi\frac{dsin\theta}{\lambda}} & e^{j2\pi\frac{2dsin\theta}{\lambda}} & e^{j2\pi\frac{3dsin\theta}{\lambda}} & \dots & e^{j2\pi\frac{(M-1)dsin\theta}{\lambda}} \end{bmatrix} s(t) + n(t) \tag{3}$$

If defined,

$$A = \begin{bmatrix} 1 & e^{j2\pi\frac{dsin\theta}{\lambda}} & e^{j2\pi\frac{2dsin\theta}{\lambda}} & e^{j2\pi\frac{3dsin\theta}{\lambda}} & \ldots & e^{j2\pi\frac{(M-1)dsin\theta}{\lambda}} \end{bmatrix} \tag{4}$$

Then the signal can be expressed as follows,

$$x(t) = As(t) + n(t) \tag{5}$$

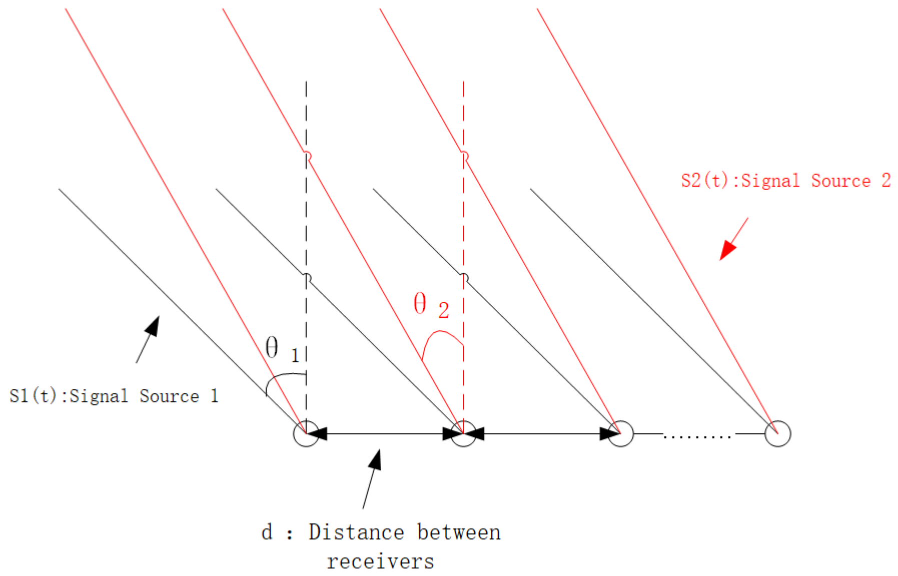

Figure 3: Multi-transmitter, multi-receiver model.

If there are multiple sources, the signal model can be extended for multiple transmits and multiple transmissions as the angles at which the sources reach the receivers do not need to be different.As shown in Figure 3,assuming that there are D signal sources and M receivers, the signal from the receiving array can be expressed as,

$$x(t) = \begin{bmatrix} e^{j2\pi\frac{0dsin\theta_1}{\lambda}} & e^{j2\pi\frac{0dsin\theta_2}{\lambda}} & \ldots & e^{j2\pi\frac{0dsin\theta_D}{\lambda}} \\ e^{j2\pi\frac{dsin\theta_1}{\lambda}} & e^{j2\pi\frac{dsin\theta_2}{\lambda}} & \ldots & e^{j2\pi\frac{dsin\theta_D}{\lambda}} \\ \ldots & \ldots & \ldots & \ldots \\ e^{j2\pi\frac{(M-1)dsin\theta_1}{\lambda}} & e^{j2\pi\frac{(M-1)dsin\theta_1}{\lambda}} & \ldots & e^{j2\pi\frac{(M-1)dsin\theta_1}{\lambda}} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \\ \ldots \\ s_D(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \\ n_3(t) \\ \ldots \\ n_D(t) \end{bmatrix} \tag{6}$$

Similar to equation (4), at this point,

$$A = \begin{bmatrix} a(\theta_1) & a(\theta_2) & a(\theta_3) & \ldots & a(\theta_D) \end{bmatrix} \tag{7}$$

of which,

$$a(\theta) = \begin{bmatrix} 1 & e^{j2\pi\frac{dsin\theta}{\lambda}} & e^{j2\pi\frac{2dsin\theta}{\lambda}} & e^{j2\pi\frac{3dsin\theta}{\lambda}} & ... & e^{j2\pi\frac{(M-1)dsin\theta}{\lambda}} \end{bmatrix}^H \tag{8}$$

$a(\theta)$ is the Spatially oriented vector, which essentially describes the spatial phase and has a structure related to the relative position of the receiver. The steering vector can be obtained by analysing the received signal $x\,(t)$, which in turn allows DoA estimation.

Principle of the MUSIC algorithm. The main principle of the MUSIC algorithm is to perform covariance decomposition on the received signal to obtain two orthogonal sets of eigenvectors, namely the signal subspace and the noise subspace. Using the orthogonal characteristics of these two sets of vectors, the spatial spectrum of the response can be constructed, and then the angle can be determined by the position of the peak of the spectrum according to the relationship between the spatial spectrum function and the angle of incidence.

The received signal is first sampled to obtain its discrete signal,

$$x(k) = As(k) + n(k) \tag{9}$$

Expand i.e.

$$x(k) = \begin{bmatrix} e^{j2\pi\frac{0dsin\theta_1}{\lambda}} & e^{j2\pi\frac{0dsin\theta_2}{\lambda}} & ... & e^{j2\pi\frac{0dsin\theta_D}{\lambda}} \\ e^{j2\pi\frac{dsin\theta_1}{\lambda}} & e^{j2\pi\frac{dsin\theta_2}{\lambda}} & ... & e^{j2\pi\frac{dsin\theta_D}{\lambda}} \\ ... & ... & ... & ... \\ e^{j2\pi\frac{(M-1)dsin\theta_1}{\lambda}} & e^{j2\pi\frac{(M-1)dsin\theta_1}{\lambda}} & ... & e^{j2\pi\frac{(M-1)dsin\theta_1}{\lambda}} \end{bmatrix} \begin{bmatrix} s_1(k) \\ s_2(k) \\ s_3(k) \\ ... \\ s_D(k) \end{bmatrix} + \begin{bmatrix} n_1(k) \\ n_2(k) \\ n_3(k) \\ ... \\ n_D(k) \end{bmatrix} \tag{10}$$

Finding the covariance matrix of the discrete signal $x\,(k)$ yields,

$$R_{xx} = E\left[x(k)x^H(k)\right] \tag{11}$$

Substituting equation (9) into equation (11),we get,

$$R_{xx} = E\left[(As+n)(As+n)^H\right] \tag{12}$$

Similarly, the covariance matrix of the sending signal can be derived as follows,

$$R_{ss} = E\left[s(k)s^H(k)\right] \tag{13}$$

Substituting equation (13) into equation (12),we get,

$$R_{xx} = E\left[(As+n)(As+n)^H\right] = AE\left[ss^H\right]A^H + E\left[nn^H\right] = AR_sA^H + \sigma^2 I \tag{14}$$

where σ^2 is the power of the noise and I is the unit matrix. Assuming that the noise is Gaussian white noise, independent of the transmitting signal, the $x\left(k\right)$ matrix model is a vector matrix of $M \times 1$ since the received signal vector matrix $x\left(k\right)$ consists of M receivers, provided that the individual signals are not correlated with each other. Similarly,$s\left(k\right)$ consisting of D sources is a vector matrix of $D \times 1$.

An eigenvalue decomposition is performed on R_{xx} to analyse the covariance matrix R_{ss} of the transmitted signal and the covariance matrix R_N of the noise, where,

$$R_N = \sigma^2 I \tag{15}$$

Decomposing R_{xx} yields,

$$\hat{R}_{xx} = U \sum U^H = \left[U_s U_N\right] \begin{bmatrix} \sum_x & 0 \\ 0 & \sum_N \end{bmatrix} \begin{bmatrix} U_s^H \\ U_N^H \end{bmatrix} \tag{16}$$

In equation (16) $\sum_x$ is a vector diagonal matrix of rank D, where the elements on the diagonal are $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_D$, and $\sum_N$ is a vector diagonal matrix of $(M-D)X(M-D)$ with eigenvalues σ^2, $\sum_N$ with eigenvalues $\lambda_{D+1} = \lambda_{D+2} = \lambda_3 = ...\lambda_D = \sigma^2$, and it is usually assumed that the noise power $\sigma^2 = 0$, so the eigenvalues of $\sum_x$ should obviously all be greater than $\sum_N$. While U_s is the eigenvector matrix corresponding to the first D larger eigenvalues of R_{xx}, U_s is the matrix of $M \times D$, U_N is the eigenvector matrix of the next $M - D$ smaller eigenvalues of R_{xx}, and U_N is the matrix of $M \times (M - D)$.

To investigate the relationship between the signal subspace U_s and the noise subspace U_N, it is necessary to simplify R_{xx} and let R_{xx} be multiplied by U_N to obtain,

$$R_{xx} U_N = A R_{ss} A^H U_N + \sigma^2 I U_N = \sigma^2 U_N \tag{17}$$

Subtracting $\sigma^2 U_N$ on both sides of the equation to $R_{xx} U_N$ simplifies to give,

$$A R_{ss} A^H U_N = 0 \tag{18}$$

Assuming that the individual signals of the signal source s are independent of each other, then R_{ss} is a non-singular matrix, i.e. there exists R_{xx}^{-1}, and by the properties of complex matrices, AA^H is also a non-singular matrix $(AA^H)^{-1}$. For multiplying both sides of equation (18) by $\left[R_{ss}^{-1}(AA^H)^{-1}A^H\right]$ at the same time gives,

$$\left[R_{ss}^{-1}(AA^H)^{-1}A^H\right] A R_s A^H U_N = 0 \left[R_{ss}^{-1}(AA^H)^{-1}A^H\right] \tag{19}$$

So it follows that,

$$A^H U_N = 0 \tag{20}$$

i.e.

$$a(\theta)^H U_N = 0 \tag{21}$$

It is clear from equation (20) that since the column vector of matrix A corresponds to the direction of the signal source, then the direction of the source can be estimated based on this property. In practical applications, where the length of the data cannot be infinite, and N is the number of snapshots, then there are,

$$\hat{R}_{xx} = \frac{1}{N}\sum_{i}^{N} x_i x_i^H \tag{22}$$

Eigenvalue decomposition of the matrix,

$$\hat{R}_{xx} = \hat{U} \sum \hat{U}^H \tag{23}$$

The estimated vector matrix $\hat{U}$ of the signal subspace can be obtained, and the estimated vector matrix $\hat{U}_N$ of the noise signal subspace can be obtained by rearranging the eigenvectors of the signal from largest to smallest according to the eigenvalue size of the signal.In practice, the signal Spatially oriented vectors $a(\theta)$ cannot be perfectly orthogonal to the noise subspace $\hat{U}_N$ because of the effect of noise, i.e.

$$a(\theta)^H \hat{U}\hat{U}^H a(\theta) \neq 0 \tag{24}$$

Thus defining the MUSIC pseudo spectrum $P_{music}(\theta)$,

$$P_{music}(\theta) = \frac{1}{a(\theta)^H \hat{U}\hat{U}^H a(\theta)} \tag{25}$$

When the denominator of the equation $a(\theta)^H \hat{U}\hat{U}^H a(\theta)$ is a minimum, $P_{music}(\theta)$ reaches a maximum, i.e. the direction of the sound source can be estimated from the peak of $P_{music}(\theta)$.

Implementation

According to the previous assumptions, in a room of 7 metres long, 5 metres wide and 2.75 metres high, there is one sound source and six microphones are used as receivers, which is a single input, multiple output model.

(1) First construct the receiver data $x\,(k)$ from equation (9),

$$x(k) = As(k) + n(k) \tag{26}$$

Consider the actual environmental factors, such as room size, microphone location, sound source location, and set the sampling frequency to 16000, sampling points to 3200, and build the filter, and then simulate the microphone signal $x(k)$, as there are six microphones, this is a 6×3200 matrix. Set the noise to Gaussian white noise and the noise power $\sigma^2 = 0$.

2) According to Nyquist's sampling theorem, the sampling frequency is more than twice the frequency of the original signal, so the frequency of the sound source should be distributed between 0 and 8000. divide different frequency bands and perform the short-time Fourier transform. In turn, the result of the short-time Fourier transform of the microphone is used to obtain the covariance matrix $\hat{R}_{xx}$.

$$\hat{R}_{xx} = mstft(k) * mstft(k)^H \tag{27}$$

Of which,$mstft(k)$ is the short-time Fourier transform of the microphone signal.

3) Eigenvalue decomposition of the covariance matrix, $\hat{R}_{xx}$, according to equation (16),

$$\hat{R}_{xx} = U \sum U^H = [U_s U_N] \begin{bmatrix} \sum_x & 0 \\ 0 & \sum_N \end{bmatrix} \begin{bmatrix} U_s^H \\ U_N^H \end{bmatrix} \tag{28}$$

Sort the eigenvector matrix $\hat{U}_s$ from largest to smallest according to the eigenvalue size, intercept the part with the smaller eigenvalue, the length is the number of microphones minus the number of sound sources to obtain the noise subspace $\hat{U}_N$.

4) According to equation (15),

$$P_{music}(\theta) = \frac{1}{a(\theta)^H \hat{U} \hat{U}^H a(\theta)} \tag{29}$$

Since the signal is divided according to different frequencies, the resulting $P_{music}(\theta)$ is assigned at different frequencies. In each frequency range, all angles are traversed to find the peak of $P_{music}(\theta)$, and the angle at which the peak is located in each frequency range is recorded.

5) The angles where the peaks are located in the range of frequencies greater than 500 and less than 4000 are summed up and the mean value is found, which is the Direction of Arrival.

Problems encountered in the implementation process

The first problem is, we used the microphone samples directly to calculate the covariance, without considering multiple frequency bands. Solution, We do not need to use time domain signal,we need to do covariance matrix for the results of the over short-time Fourier transform of the microphone signal.We use a for-loop over frequency band,we compute the covariance matrix for each frequency bands.

The second problem is, we did not notice that the eigenvectors obtained by decomposing the covariance matrix R_{xx} using the function $eig\,()$ in Matlab are arranged from largest to smallest, leading to errors in the calculation of the noise subspace. Solution, We found this error by looking at the values of the eigenvalue vector and then adjusted the way we calculated the noise subspace.

Evaluation Results

In a room 7 metres long, 5 metres wide and 2.75 metres high, with one sound source and six microphones, with source coordinates (3.5, 3.0 , 1.7) and microphone centre coordinates (1.5, 4.0, 1.7), the following results are produced.

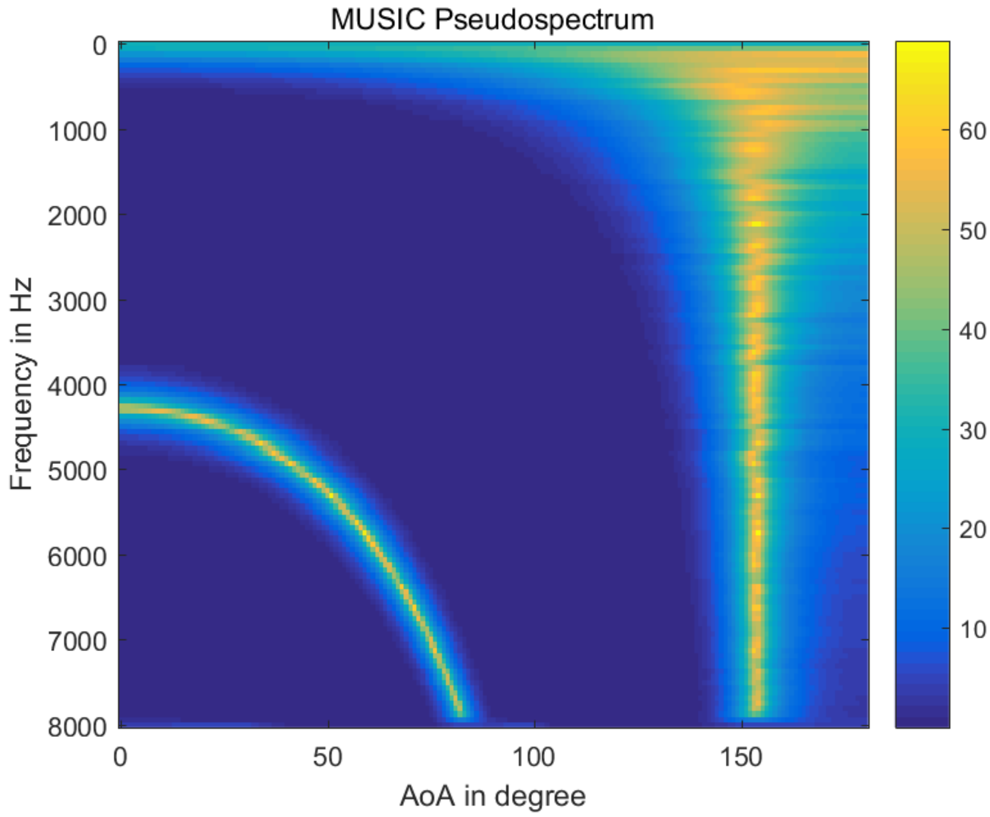

Figure 4: MUSIC Pseudospectrum.

In the MUSIC Pseudospectrum diagram, the horizontal axis represents the scanning angle, ranging from 0 to 180 degrees, the number axis represents the source frequency, ranging from 0 to 8000 Hz, and the shades of colour represent the value of equation (29).

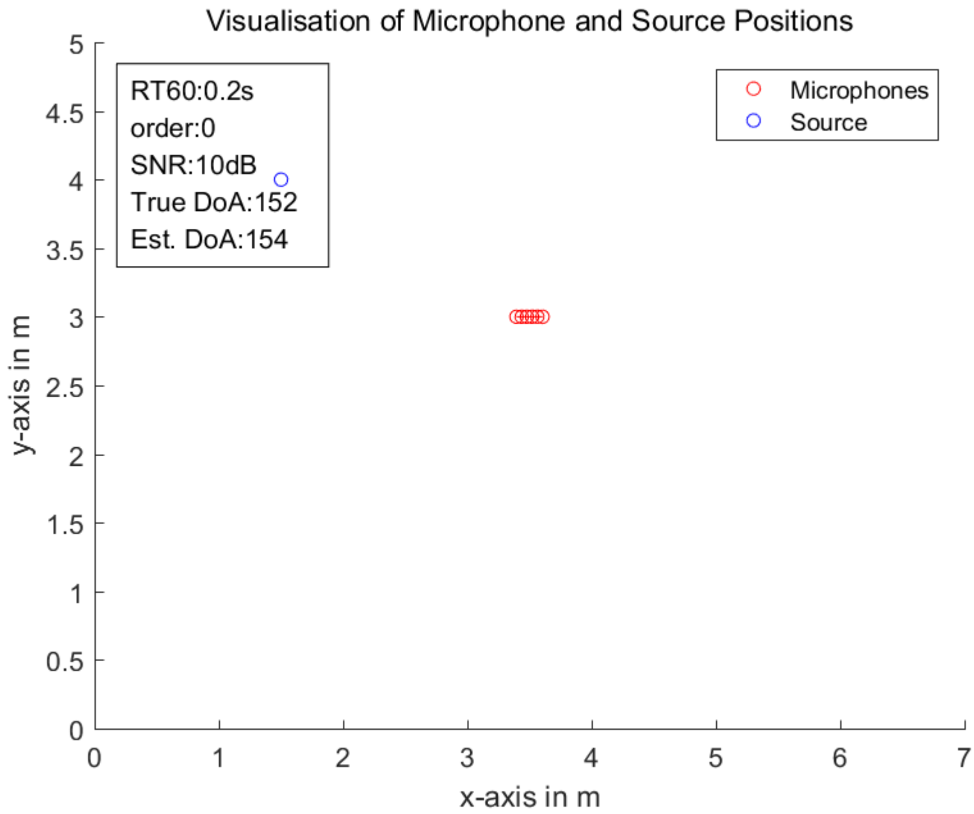

Figure 5: Visualizing room.

Bibliography

[1] Schmidt R. Multiple emitter location and signal parameter estimation[J]. IEEE transactions on antennas and propagation, 1986, 34(3): 276-280.

[2] Stoica P, Nehorai A. MUSIC, maximum likelihood, and Cramer-Rao bound[J]. IEEE Transactions on Acoustics, speech, and signal processing, 1989, 37(5): 720-741.