

Application of Bayesian Decision Theory Based on Prior Information in the Multi-Objective Optimization Problem

Xia Lei

*National Key Laboratory of Communications, University of Electronic Science and Technology of China
Chengdu, 610065, P.R. China
E-mail: leixia@uestc.edu.cn*

Maozhu Jin*

*Business School of Sichuan University
Chengdu, 610065, P.R. China
E-mail: jinmaozhu@scu.edu.cn*

Qiang Wang

*National Key Laboratory of Communications, University of Electronic Science and Technology of China
Chengdu, 610065, P.R. China
E-mail: wangqiang @uestc.edu.cn*

Abstract

General multi-objective optimization methods are hard to obtain prior information, how to utilize prior information has been a challenge. This paper analyzes the characteristics of Bayesian decision-making based on maximum entropy principle and prior information, especially in case that how to effectively improve decision-making reliability in deficiency of reference samples. The paper exhibits effectiveness of the proposed method using the real application of multi-frequency offset estimation in distributed multiple-input multiple-output system. The simulation results demonstrate Bayesian decision-making based on prior information has better global searching capability when sampling data is deficient.

Keywords: Multi-Objective Optimization, Prior Information, Maximum Entropy Principle, Distributed multiple inputs and multiple outputs.

1. Introduction

Bayesian statistics is one of key branches in modern statistics has its origin of the famous paper ‘Discussion on the Problem of Problem of a Chance or an Opportunity’ authored by British scholar Thomas Bayes (1702-1761) which proposed the Bayesian formulation and the corresponding reasoning method.

The basic point of view in Bayesian statistics study is that any unknown parameter should be regarded as a statistics variable which can be described by a

probability distribution, a prior distribution which represents the knowledge of the event before proceeding sampling investigation.¹⁻³ This is what the difference lies between Bayesian statistics and classic statistics. Classic statistic scholars didn’t recognize the prior information and advocates the statistical deduce using sample information, but that situation has changed and prior information has gained acknowledge in classic statistics. The focal argument is how to obtain and utilize the prior information to determine the prior distribution appropriately.

* Corresponding author

Bayesian decision obtains the conditional distribution density (prior distribution) of unknown parameters in condition of known sampling distribution. Since this sampling distribution is obtained after the sampling process is complete, it is called posteriori distribution. The key to Bayesian deduction is that any results must and can only rely on posteriori distribution.

Multi-objective decision-making method focuses on simultaneous optimization of multiple targets, different targets can't use uniform measuring criteria and more usually than not these targets are contradictory.³ So in this sense multi-objective decision-making gives rise to a set of trade-off optimal solutions, instead of a single optimum solution.⁴

From Bayesian statistics perspective, Bayesian multi-objective decision-making deals with the multi-risk problem using multi-objective decision-making principles. Bayesian decision-making which falls into the conventional decision-making field focuses on single objective decision. However, multi-objective decision-making theory has neglected the benefits of Bayesian decision-making to handle multi-objective decision-making problem with uncertainties.^{5,6}

Bayesian decision-making estimates probabilities of partly unknown parameters and rectify the probability using Bayesian formula, and in last step, making decision according to the expectation value and rectified probabilities.⁷

Suppose that θ is the objective parameter, then the optimization proceeds using maximum posteriori distribution based on Bayesian Criteria, which makes

$$\hat{\theta} = \arg \max_{\theta} p(\theta|\mathbf{x}) \quad (1)$$

Where \mathbf{x} is the data vector including the information of θ . To get maximum value of $p(\theta|\mathbf{x})$, we observe

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \quad (2)$$

So maximum value of $p(\mathbf{x}|\theta)p(\theta)$ is equivalent to (1). According to (2), other than the existence of prior probability density function (PDF), $p(\mathbf{x}|\theta)p(\theta)$ is tantamount to maximum likelihood examination.

Therefore,

$$\hat{\theta} = \arg \max_{\theta} p(\mathbf{x}|\theta)p(\theta) \quad (3)$$

Or

$$\hat{\theta} = \arg \max_{\theta} [\ln p(\mathbf{x}|\theta) + \ln p(\theta)] \quad (4)$$

The PDF of prior information is more “abrupt”, more important influence it will put on the estimation. Information is more included on the sampling data, more effects sampling data will exert on the estimation, at the same time the influence of prior information will wane.^{8,9}

Put it in another way, when information provided by sampling data is deficient, the performance of optimization strategy will be compromised. Making appropriate use of prior information to make up for this deficiency is of great importance to improve the effect of strategic optimization.^{10,11}

Previous study based on Bayesian Criteria deals with the situation that Bayesian Criteria is relevant to prior probability and cost function when prior probability and cost function are known. But in a real setting, prior probability and cost function may stay unknown and fuzzy; in that case, how they take effect on the system performance remains a challenge to the academia worthy of further research.¹²

2. System Model

Distributed multiple inputs and multiple outputs (MIMO) system is an important application area of future communication technique. It uses multiple parallel channels transmission method so that it can achieve higher transmission efficiency. After applied the distributed antenna configuration, the original single-objective optimization problem become a typical multi-objective optimization problem because transmitting antennas or receiving antennas located in different geographic locations and using different crystal lead to different transmitting antennas to the same receiving antennas has different frequency offset. Even a general distributed MIMO system that transmitting antenna located in a different location, the mobile service also uses a distributed antenna. At this point, we could consider that the time delay and frequency offset between different receiving antenna and transmitting antenna will different too. Under such assumptions, the time matrix offset and frequency offset matrix could extended to

$$\Theta = [\theta_{i,j}] = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,M_T} \\ \theta_{2,1} & \theta_{2,2} & & \theta_{2,M_T} \\ \vdots & & \ddots & \\ \theta_{M_R,1} & \theta_{M_R,2} & \dots & \theta_{M_R,M_T} \end{bmatrix} \quad (5)$$

$$\mathbf{E} = [\varepsilon_{i,j}] = \begin{bmatrix} \varepsilon_{1,1} & \varepsilon_{1,2} & \dots & \varepsilon_{1,M_T} \\ \varepsilon_{2,1} & \varepsilon_{2,2} & & \varepsilon_{2,M_T} \\ \vdots & & \ddots & \\ \varepsilon_{M_R,1} & \varepsilon_{M_R,2} & \dots & \varepsilon_{M_R,M_T} \end{bmatrix} \quad (6)$$

Where $\theta_{i,j}$ denotes time offset between receiving antenna i and receiving antenna j , $\varepsilon_{i,j}$ denotes normalized frequency offset between receiving antenna i and receiving antenna j .

(6) indicates when both transmitting antennas and receiving antennas employ distributed model, frequency offset estimation is transformed into the problem of multi-parameter estimation.

Analysis on maximum posteriori estimation has inspired us to utilize prior information of frequency offset to estimate multi-frequency offset in distributed MIMO system. Especially in the environment of low signal-noise-ratio (SNR), analysis on maximum posteriori estimation of multi-frequency offset has great applicable value on improving the performance of frequency synchronization.

Considering the case of distributed MIMO system with M_T transmitting antennas and M_R receiving antennas, which is tantamount to M_R distributed multi-input single-output (MISO) systems. Without loss of generality, this paper focuses on the k^{th} MISO system, the receiving signal can be represented by

$$\mathbf{y}_k = \sum_{l=1}^{M_T} h_l \mathbf{E}_l \mathbf{X}_l + \boldsymbol{\omega} \in C^{N \times 1} \quad (7)$$

where $k \in \{1, 2, \dots, M_T\}$, \mathbf{E}_l is diagonal matrix with elements of

$$[\mathbf{E}_l]_{n,n} = e^{j2\pi\varepsilon_l n} \quad (8)$$

ε_l is the frequency offset produced by l^{th} transmitting antenna. $\mathbf{X}_l \in C^{N \times 1}$ is the sequence send by l^{th} transmitting antenna and $h_l, l \in \{1, 2, \dots, M_T\}$ is flat fading channel coefficient. $\boldsymbol{\omega} \in CN(\mathbf{0}, \sigma^2 \mathbf{1})$ is complex Gaussian noise vector.

2.1. Obtaining strategy of prior information based on the principle of maximum entropy

According to the description of the previous section, prior information in Bayesian decision played an important role, especially when the sample collection of event is inadequate, the prior information could reflect its effect on decision-making optimization.

Assuming the probability of discrete random variable x is

$$p(x = a_i) = p_i, i = 1, \dots, N \quad (9)$$

when its value equal to a_1, \dots, a_N ,

$$H(x) = -\sum_{i=1}^N p_i \ln p_i \quad (10)$$

Represents the entropy of x , similarly, if x is a continuous random variable, its probability density function is $p(x)$, and

$$H(x) = -\int p(x) \ln p(x) dx \quad (11)$$

(11) is finite integrable, similarly, it represents the entropy of x .

According to the definition, the entropy only link with random variable's probability density function, which means related to the distribution of random variables.

In 1957, Jaynes ET Jaynes first proposed the maximum entropy principle: among all the compatible distributions, choose some distributions which can make the information entropy reach to its maximum under the situation that suffices some constraint conditions. In case of part of information has gained, therefore, deduction of distribution can be made only if it takes the probability distribution conform to constraints but entropy has the maximum values. Any other options means we add some other constraints or conditions which are unavailable based on the information has already gained. Maximum entropy principle¹³ is that when priori information distribution is not determined the selected priori information should have maximums entropy distribution.

Accordingly, Johnny pointed out that some important real-world probability distributions which comply with the maximums entropy principle under some constraints. When the value of random variable is continuous, meanwhile, if the expectation and variance are known, the maximum entropy distribution is the normal distribution. This paper will prove the results that the

maximum entropy distribution is Gaussian distribution as random variables' mean and variance are known. First, there are two lemmas.¹³

Lemma 1: Suppose $\varphi(t)$ is a convex function that the domain is (a, b) , which is accountable for random variables ξ , $E\varphi(\xi)$, $\varphi(E\xi)$, then

$$E\varphi(\xi) \geq \varphi(E\xi) \tag{12}$$

Obviously, if

$$\varphi(t) = t \ln t \tag{13}$$

We can get

$$\varphi'(t) = 1 + \ln t \tag{14}$$

$$\varphi''(t) = 1/t \tag{15}$$

hence, $\varphi(t)$ is a convex function when $t \in (0, +\infty)$, then,

$$E(\xi \ln \xi) \geq (E\xi) \ln (E\xi) \tag{16}$$

can be deduced by(12).

Lemma 2: For the distribution probability density function

$$f_i(x), i = 1, 2 \tag{17}$$

if both ends of (18) have meanings, then, (19) is available

$$\int f_1(x) \ln f_1(x) dx \geq \int f_1(x) \ln f_2(x) \tag{20}$$

Proof: Let

$$\zeta = f_1(\lambda) / f_2(\lambda), \lambda \in f_2(x) \tag{21}$$

Because

$$\begin{aligned} E\zeta &= E[f_1(\lambda) / f_2(\lambda)] \\ &= \int \frac{f_1(\lambda)}{f_2(\lambda)} f_2(\lambda) d\lambda \\ &= 1 \end{aligned} \tag{22}$$

Then

$$\ln(E\zeta) = 0 \tag{23}$$

According to Lemma 1, we can proof

$$\begin{aligned} 0 &\leq E\zeta \ln \zeta \\ &= \int \left(\frac{f_1(\lambda)}{f_2(\lambda)} \ln \frac{f_1(\lambda)}{f_2(\lambda)} \right) f_2(\lambda) d\lambda \\ &= \int f_1(\lambda) \ln f_1(\lambda) d\lambda - \int f_1(\lambda) \ln f_2(\lambda) d\lambda \end{aligned} \tag{24}$$

Based on Lemma 2, if a random variable's mean and variance are limited, then the maximal entropy distribution is the Gaussian distribution.

Proof: Consider set

$$P = \left\{ \begin{aligned} &p(\varepsilon) : p(\varepsilon) \geq 0, \\ &\int_{-\infty}^{+\infty} \varepsilon p(\varepsilon) d\varepsilon = 0, \\ &\int_{-\infty}^{+\infty} \varepsilon^2 p(\varepsilon) d\varepsilon = \sigma_\varepsilon^2 \end{aligned} \right\} \tag{25}$$

Obviously, the probability density function of $\varepsilon \in N(0, \sigma_\varepsilon^2)$ is

$$p_0(\varepsilon) = e^{-\varepsilon^2/2\sigma_\varepsilon^2} / \sqrt{2\pi\sigma_\varepsilon^2} \in P \tag{26}$$

According to Lemma 2, $\forall p_1(\varepsilon) \in P$, then

$$\int_{-\infty}^{+\infty} p_1(\varepsilon) \ln p_1(\varepsilon) d\varepsilon - \int_{-\infty}^{+\infty} p_1(\varepsilon) \ln p_0(\varepsilon) d\varepsilon \geq 0 \tag{27}$$

Further deduction of (27), then

$$\begin{aligned} &\int_{-\infty}^{+\infty} p_1(\varepsilon) \ln p_1(\varepsilon) d\varepsilon \geq \\ &\int_{-\infty}^{+\infty} p_1(\varepsilon) \left[\ln \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} - \frac{1}{2\sigma_\varepsilon^2} \varepsilon^2 \right] d\varepsilon \\ &= \ln \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} - \frac{1}{2} = \int_{-\infty}^{+\infty} p_0(\varepsilon) \ln p_0(\varepsilon) d\varepsilon \end{aligned} \tag{28}$$

Therefore, the mean 0 and variance of random variables, the maximum entropy distribution is normal distribution. The usual frequency offset estimation is doesn't consider to use the frequency offset prior information, so the maximum likelihood (ML) estimation method is the most optimal detection algorithm. As previously discussed, the actual frequency offset can be changed as the temperature, humidity and other environmental factors' change, the Refs. 14 regards the frequency offset as random variables to study the synchronization problem. If you can get frequency offset prior information reasonably, then, it will play an important

role for improving the frequency synchronization under low SNR environment.

Refs. 15 proposed a synchronization algorithm in cooperative communication system which based on frequency offset prior information. In low SNR region, it can improve the performance of frequency synchronization. Although the literature proposed the application of frequency offset prior information in collaborative communication environment, however, how to choose the distribution of prior information and obtain the deviation of the prior information and practicing it in a multi-frequency offset estimation of distributed MIMO systems is still a valuable researching direction.

Can be realizable in the physical system, frequency offset is often distributed in a limited range, namely $\varepsilon \in (-\eta, \eta)$. The former literature regards frequency offset as random variables which mean is 0 and variance is σ_ε^2 . Based on this and combined with the analysis of prior information above, according to the principle of maximum entropy, furthermore, this paper argue that frequency offset obeys normal distribution, ie: $\varepsilon_k \in N(0, \sigma_{\varepsilon_k}^2)$, $k = 1, 2, \dots, M_T$.

After determining the prior information, the question how to get more reasonable values for the variance parameters is worthy of further study.

According to the Three times standard deviation principle¹⁶

$$P\left\{\left|\xi - \mu\right| < 3\sigma_\xi\right\} = 2\Phi(3) - 1 \approx 0.9974 \quad (29)$$

As a results, the vast majority of Random variable's sample will distribute in $(\mu - 3\sigma_\xi, \mu + 3\sigma_\xi)$, As for the other areas, we can regard as small probability event, therefore, prior information's variance can be determined by the Three times standard deviation principle. If frequency offsets' maximums can be get, $\varepsilon_k \in (-\eta_k, \eta_k)$, $k = 1, 2, \dots, M_T$, then

$$\sigma_{\varepsilon_k}^2 = \eta_k^2 / 9 \quad (30)$$

According to the Three times standard deviation principle, using frequency offsets distribution range to do estimation for frequency offsets prior variance in distribution MIMO system is a reasonable method.

2.2. Bayesian decision algorithm based on prior information

The current frequency offset estimation algorithms basically regarded frequency offset as a fixed parameter, however, the facts it becomes a random variable because of the influence of environmental change factors such as temperature, humidity etc is often overlooked. A few literatures (Refs. 14, 15 and 17) had begun to research multi-frequency offset estimation algorithm as the frequency offset change to variable. If prior information of frequency offset can be gained reasonably, it will improve the frequency synchronization performance, especially when sampling data is deficient. Main research of this paper is distributed MIMO system's maximum posteriori multi-frequency estimation algorithm based on frequency offset priori information.

2.2.1 Optimal Maximum a Posteriori Algorithm

According to Eq.(3), maximum multi-frequency offset estimation of distributed MIMO system is to obtain¹⁸

$$\hat{\varepsilon} = \arg \max_{\varepsilon} p(\mathbf{y}|\varepsilon) p(\varepsilon) \quad (31)$$

Where, $p(\mathbf{y}|\varepsilon)$ is conditional distribution density of received sampling data and $p(\varepsilon)$ is distribution density of prior information, they are respectively

$$p(\mathbf{y}|\varepsilon) = \frac{1}{\pi^N \sigma^{2N}} \exp\left\{-\frac{1}{\sigma^2} \left\|\mathbf{y} - \sum_{l=1}^{M_T} h_l \mathbf{E}_l \mathbf{X}_l\right\|^2\right\} \quad (32)$$

$$p(\varepsilon) = \frac{1}{(2\pi)^N \det^{\frac{1}{2}}(\mathbf{R})} \exp\left\{-\frac{1}{2} \varepsilon^T \mathbf{R}^{-1} \varepsilon\right\} \quad (33)$$

Taking the logarithm of objective function $p(\mathbf{y}|\varepsilon) p(\varepsilon)$

$$L(\mathbf{y}, \varepsilon) = \ln p(\mathbf{y}|\varepsilon) + \ln p(\varepsilon) \propto -\frac{1}{\sigma^2} \left\|\mathbf{y} - \sum_{l=1}^{M_T} h_l \mathbf{E}_l \mathbf{X}_l\right\|^2 - \frac{1}{2} \varepsilon^T \mathbf{R}^{-1} \varepsilon \quad (34)$$

Note

$$\mathbf{Z} = [\mathbf{E}_1 \mathbf{X}_1, \mathbf{E}_2 \mathbf{X}_2, \dots, \mathbf{E}_{M_T} \mathbf{X}_{M_T}] \quad (35)$$

and estimate maximum likelihood (ML) of channel in case of known frequency offset

$$\hat{\mathbf{H}} = \min_{\mathbf{H}} \|\mathbf{y} - \mathbf{Z}\mathbf{H}\|^2 \quad (36)$$

Further simplify and get

$$\hat{\mathbf{H}} = (\mathbf{Z}^H \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{y} \quad (37)$$

Substitute (37) into (34) and simplify the result to get

$$L(\mathbf{y}, \boldsymbol{\varepsilon}) \propto -\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{Z}\hat{\mathbf{H}}\|^2 - \frac{1}{2} \boldsymbol{\varepsilon}^T \mathbf{R}^{-1} \boldsymbol{\varepsilon} \quad (38)$$

With the Eq.(31), then MAP estimation of multi-frequency offset is obtain by

$$\hat{\boldsymbol{\varepsilon}} = \arg \min_{\boldsymbol{\varepsilon}} \left\{ \left\| \left(\mathbf{I} - \mathbf{Z}(\mathbf{Z}^H \mathbf{Z})^{-1} \mathbf{Z}^H \right) \mathbf{y} \right\|^2 + \frac{\sigma^2}{2} \boldsymbol{\varepsilon}^T \mathbf{R}^{-1} \boldsymbol{\varepsilon} \right\} \quad (39)$$

(39) indicates the first term of objective function is maximum likelihood estimation and second term is produced by prior information. When prior information is not considered, objective function is converted to ML estimation

$$\hat{\boldsymbol{\varepsilon}}_{ML} = \arg \min_{\boldsymbol{\varepsilon}} \left\{ \left\| \left(\mathbf{I} - \mathbf{Z}(\mathbf{Z}^H \mathbf{Z})^{-1} \mathbf{Z}^H \right) \mathbf{y} \right\|^2 \right\} \quad (40)$$

In essence (40) is same as ML estimation, that is equivalent to

$$\hat{\boldsymbol{\varepsilon}}_{ML} = \arg \max_{\boldsymbol{\varepsilon}} \mathbf{y}^H \mathbf{Z}(\mathbf{Z}^H \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{y} \quad (41)$$

It is evident that maximum a posteriori (MAP) estimation of multi-frequency offset combines sampling information of receiving data and prior information of frequency offset. The general ML estimation can only be based on data sampling information, so if interference is relative large (for example, low SNR) or training sequence is short, received signal sample can't provide sufficient useful information that results in worse performance of frequency offset estimation. MAP estimation introducing prior information combines information of data sample and prior information of frequency offset to make up the deficiency of insufficient useful information of data sample.

2.2.2 Suboptimum Quasi Maximum a Posteriori algorithm

Analysis in section 2.2.1 obtains objective function of multi-frequency offset MAP estimation and indicates there is internal correlation between MAP estimation and ML estimation. So borrowing from the insights of

ML estimation algorithm (for example, Expectation Conditional Maximum (ECM) algorithm) MAP estimation based on prior information can be obtained, but complexity will increase in direct solving objective function due to introduction of prior information.

Next, the paper will propose a suboptimum MAP estimation algorithm, that is, quasi maximum a posteriori (QMAP) algorithm.

Since ML estimation is an unbiased estimation,

$$E(\hat{\varepsilon}_i) = \varepsilon_i, 1, 2, \dots, M_T \quad (42)$$

The mean square error can be represented by

$$MSE(\hat{\varepsilon}_i) = E\left[(\hat{\varepsilon}_i - \varepsilon_i)^2\right] \quad (43)$$

Eq. (43) is equivalent to

$$MSE(\hat{\varepsilon}_i) = E\left[\left((\hat{\varepsilon}_i - E(\hat{\varepsilon}_i)) + (E(\hat{\varepsilon}_i) - \varepsilon_i)\right)^2\right] \quad (44)$$

Simplify Eq.(44) and get

$$MSE(\hat{\varepsilon}_i) = \text{var}(\hat{\varepsilon}_i) + E\left[(E(\hat{\varepsilon}_i) - \varepsilon_i)^2\right] \quad (45)$$

Where $\text{var}(\hat{\varepsilon}_i)$ is variance of the estimator $\hat{\varepsilon}_i$. According to the analysis, theoretically MAP estimation will get better estimation performance, which is smaller MSE.

Suppose that MAP estimation is¹⁹

$$\hat{\varepsilon}_{i,MAP} = \alpha_i \hat{\varepsilon}_i \quad (46)$$

and substitute this estimator into (45), that will make MSE smaller, that is

$$MSE(\hat{\varepsilon}_{i,MAP}) = \text{var}(\alpha_i \hat{\varepsilon}_i) + E\left[(E(\alpha_i \hat{\varepsilon}_i) - \varepsilon_i)^2\right] \quad (47)$$

Simplify Eq.(47) and get

$$MSE(\hat{\varepsilon}_{i,MAP}) = \alpha_i^2 \text{var}(\hat{\varepsilon}_i) + E\left[(\alpha_i \varepsilon_i - \varepsilon_i)^2\right] \quad (48)$$

Doing the derivation over parameter α_i in (48) and set the results equal to zero

$$\begin{aligned} \frac{\partial}{\partial \alpha_i} MSE(\hat{\varepsilon}_{i,MAP}) &= 2\alpha_i \text{var}(\hat{\varepsilon}_i) + 2(\alpha_i - 1)E(\hat{\varepsilon}_i^2) \\ &= 0 \end{aligned} \quad (49)$$

Then, we can get the α_i that can satisfy the condition

$$\alpha_i = \frac{E(\hat{\epsilon}_i^2)}{E(\hat{\epsilon}_i^2) + \text{var}(\hat{\epsilon}_i)} \quad (50)$$

Combining the statistical characteristics of prior information, when SNR gradually increases, $E(\hat{\epsilon}_i^2) \rightarrow \sigma_{\epsilon_i}^2$ and at the same time $\text{var}(\hat{\epsilon}_i)$ will gradually approach cramer-rao low bound (CRLB). However, exact CRLB of ML estimation is dependent on information channel and frequency offset. If $\text{var}(\hat{\epsilon}_i)$ is substituted by approaching CRLB, this will produce result that it is only related to power of training sequence

$$asCRLB(\hat{\epsilon}_i) = \frac{6\sigma^2}{N^3 \cdot P_l \cdot |h_l|^2} \propto \frac{6}{N^3 \cdot P_l \cdot SNR} \quad (51)$$

Where

$$P_l = E\{|x(l)|^2\}, l=1, 2, \dots, N \quad (52)$$

According to this, the paper gets the approximate corrected value

$$\alpha_i = \frac{\sigma_{\epsilon_i}^2}{\sigma_{\epsilon_i}^2 + asCRLB(\hat{\epsilon}_i)} \quad (53)$$

In summary, ML estimator of frequency offset can be obtained by only using information of the data sample. So when prior information is taken into consideration, approximate estimator of MAP can be obtained by multiplying corrected value calculated by Eq. (53) with ML estimator and the result doesn't introduce extra complexity.

3. Simulation Results and Analysis

Fig.1 provides low limit performance curve of frequency estimation based on maximum likelihood and maximum posterior principle. Assuming that average value of frequency offset is zero, variance of frequency offset is decided by hardware characteristics of crystal oscillator. According to recent study, variance of frequency offset can be estimated by certain algorithms. Suppose that instability of crystal oscillator with exact estimation will cause standard deviation of prior distribution of frequency offset $\sigma_{\epsilon} = 0.01$

Relevant simulation parameters are listed in table1.

10000 Monte-Carlo simulation tests proceed at each SNR situation.

Table 1.Simulation condition for CRLB based on maximum likelihood and maximum posterior principle

System includes 4 transmitting antennas and 2 receiving antennas	
Normalized frequency offset value between Transmitting antenna 1 and Receiving antennas 1	$2\pi \cdot 0.01$
Normalized frequency offset value between Transmitting antenna 2 and Receiving antenna 1	$2\pi \cdot 0.015$
Normalized frequency offset value between Transmitting antenna 3 and Receiving antenna 1	$2\pi \cdot 0.015$
Normalized frequency offset value between Transmitting antenna 4 and Receiving antenna 1	$2\pi \cdot 0.015$
System includes 4 transmitting antennas and 2 receiving antennas	
Normalized frequency offset value between Transmitting antenna 1 and Receiving antennas 1	$2\pi \cdot 0.01$
Normalized frequency offset value between Transmitting antenna 2 and Receiving antenna 1	$2\pi \cdot 0.015$

Fig1 indicates when the SNR is low and with known prior information, CRLB of multi-frequency estimation based on maximum posterior estimation is better than that based on maximum likelihood estimation, which provides theoretical basis for frequency estimation using prior information.

Then, the paper will provide the simulation results and analysis and makes comparison between two maximum posterior multi-frequency offset algorithms in distributed MIMO systems in the following.

Specifically, the paper will explore mean square error performance of optimum MAP estimation and suboptimum QMAP estimation and improvement of frequency synchronization performance in the low SNR environment through Monte-Carlo simulation tests and theoretical analysis. Prior distribution of frequency offset is Gaussian distribution with average value of zero and variance of frequency offset is obtained by estimation of distribution range $(-\eta, \eta)$.

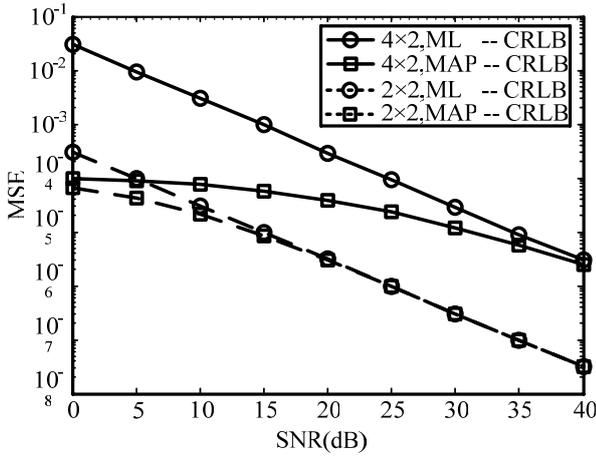


Fig.1. CRLB based on ML and MAP

3.1. Optimal Maximum a Posteriori Algorithm

Here, the paper will make comparison between MAP estimation algorithm and ML estimation algorithm in 2x1 and 4x1 distributed MISO systems and analyze relevant CRLB. To ML estimation, the paper still choose expectation maximum algorithm to get the solution²⁰, specific simulation condition is listed in table2.

Table 2. Simulation condition of MAP estimation algorithm

Parameters	Distributed MISO	Distributed MISO
	2x1	4x1
Length of orthogonal training sequence	$N = 32$	$N = 32$
Range of normalized frequency offset (1)	$2\pi \cdot (-0.05, 0.05)$	$2\pi \cdot (-0.05, 0.05)$
Range of normalized frequency offset (2)	$2\pi \cdot (-0.1, 0.1)$	$2\pi \cdot (-0.1, 0.1)$
Information channel	flat Rayleigh fading	flat Rayleigh fading
Number of simulation	10000	10000

Due to introduction of prior information of frequency offset, the performance of ML estimation is worse below 0dB. There are two reasons for this:

- Algorithm of ML estimation is sensitive at low SNR and induces the problem of convergence.

- Considering prior information, the frequency offset is no longer set value in conventional estimation, so performance of ML estimation will become worse in strong noise environment.

Fig2 and Fig3 indicates the pperformance comparison in 2x1 distributed MISO system

Fig4 and Fig5 indicates the pperformance comparison in 4x1 distributed MISO system

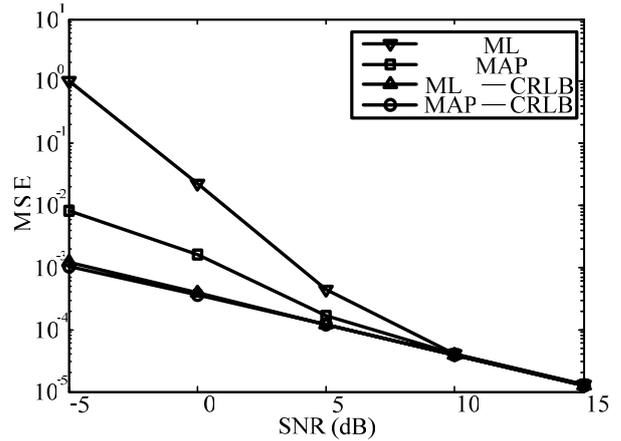


Fig.2. Performance comparison between MAP and ML in 2x1 system ($\eta = 0.05$)

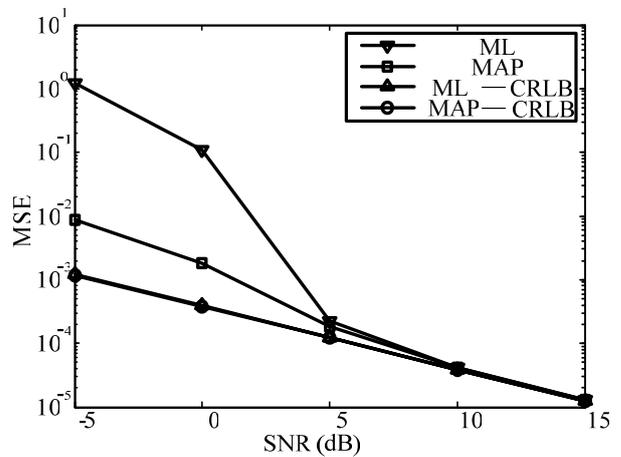


Fig.3. Performance comparison between MAP and ML in 2x1 system ($\eta = 0.1$)

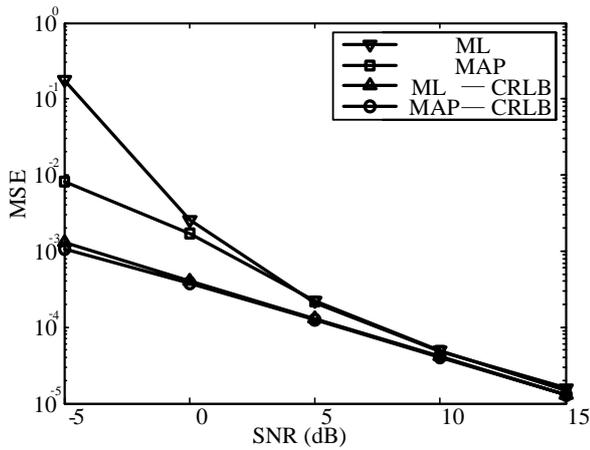


Fig.4. Performance comparison between MAP and ML in 4x1 system ($\eta = 0.05$)

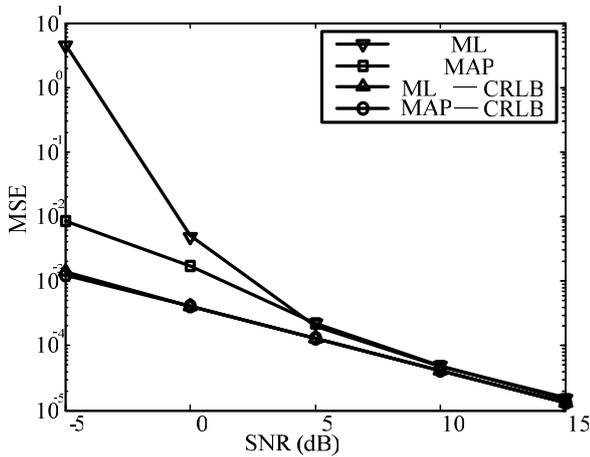


Fig.5. Performance comparison between MAP and ML in 4x1 system ($\eta = 0.1$)

To MAP estimation, in different frequency offset range and 2x1 and 4x1 distributed MISO systems, the performance stays stable.

At $SNR = -5dB$, in 2x1 distributed MISO system, MAP estimation obtains performance gain 5dB more than that of ML estimation; in 4x1 distributed MISO system, MAP estimation obtains performance gain 5dB smaller than that of ML estimation. This difference results from multi-antenna interference.

When entering the area of $SNR > 5dB$, performance of ML and MAP is approaching the same. This is mainly because diminish of noise interference makes enhanced the effective information of sampling data, the effect of prior information is waning, PDF function of sampling

data becomes acute and influence of local convergence is decreasing.

Simulation results indicate MAP estimation significantly improves performance frequency synchronization in low environment. Besides, from Fig.2 to Fig.5, it can be observed that CRLB of MAP estimation based on prior information is a little bit lower than that of ML estimation, and they are approaching the same. This observation is consistent with the rules of real estimation performance.

Since length of training sequence is 32, ideally information provided by sample data is dominant which results insignificant difference between two methods. It is worth mentioning that range of frequency offset is smaller, so is variance. So prior information play bigger role that cause better performance, otherwise, performance becomes worse.

3.2. Suboptimum Quasi Maximum a Posteriori algorithm

In previous simulation, relatively longer sequence is employed. Here, the paper will use shorter sequence to make comparison between optimum MAP estimation algorithm and suboptimum QMAP estimation algorithm and meanwhile further demonstrate the influence of number of sample data to CRLB. Due to previous mentioned different range of frequency offset, here, only one range of frequency offset is adopted. Relevant simulation parameters are listed in table3.

Table 3. Simulation condition of MAP estimation algorithm

Parameters	Distributed MISO	Distributed MISO
	2x1	4x1
Length of orthogonal training sequence	$N=16$	$N=16$
Range of normalized	$(-0.05, 0.05)$	$(-0.05, 0.05)$
Information channel	flat Rayleigh fading	flat Rayleigh fading
Number of simulation	10000	10000

Fig.6 and Fig.7 indicates the performance comparison results in 2x1 and 4x1 distributed systems.

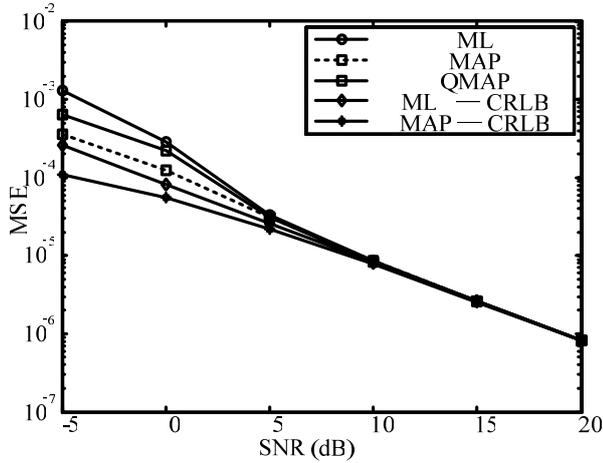


Fig.6. Performance comparison between optimum and suboptimum in 2×1 system ($\eta = 0.05$)

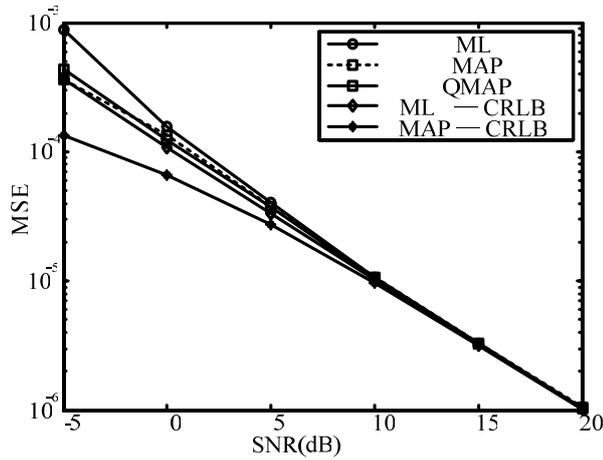


Fig.7. Performance comparison between optimum and suboptimum in 4×1 system ($\eta = 0.05$)

According to Fig.6 and Fig.7, even if suboptimum QMAP estimation can reach the performance of optimal MAP estimation, but it still can improve performance frequency synchronization in low environment. It is real advantage of suboptimum QMAP estimation that without introduction of extra complexity, suboptimum QMAP estimation can approach performance of optimal MAP estimation.

As for comparison with ML estimation, it is the same situation with optimum MAP, so the discussion is omitted here. Besides, with previous simulation results, it can be observed that when the length of sample data becomes shorter, prior information will play a bigger

role in low SNR environment. At the same time, influence of prior information in 4×1 system is larger than that in 2×1 system. At $SNR = -5dB$, performance gain provided by prior information approach 5dB because enhanced multi-antenna interference increase its influence over sample data that cause weaker information effectiveness. In high SNR environment, influence of sample data is enhanced and performance is approaching CRLB.

4. Conclusion

Bayesian decision-making estimates parameters utilizing existing data and prior information. This paper mainly discussed the priori information based distribution MIMO system's system multi-frequency offset MAP estimation. Firstly, it discussed how to choose frequency offset prior information. Secondly, the paper deduced CRLB of MAP algorithm when prior information was taken into consideration. Meanwhile, it also deduced a suboptimal QMAP estimation algorithm which only needs a related corrected value. Finally, this paper compares the performance of ML, MAP and QMAP by using simulation. The simulation results demonstrated that maximum a posteriori multi-frequency estimation based on priori information can significantly improve frequency synchronization performance especially when sampling data is deficient.

Appendix A. CRLB of the performance of Bayesian Decision Theory Based on Prior Information in the Multi-Objective Optimization Problem

Firstly, defining object parameter

$$\theta^T = [\text{Re}(\mathbf{H})^T, \text{Im}(\mathbf{H})^T, \boldsymbol{\varepsilon}^T] \quad (\text{A.1})$$

where

$$\mathbf{H}^T = [h_1, h_2, \dots, h_{M_T}] \quad (\text{A.2})$$

$$\boldsymbol{\varepsilon}^T = [\varepsilon_1, \varepsilon_1, \dots, \varepsilon_{M_T}] \quad (\text{A.3})$$

$\text{Re}(\mathbf{H})$, $\text{Im}(\mathbf{H})$ represent the real component and empty component of channel matrix, parameter's Fisher information matrix is

$$FIM_{\theta} = E_{\boldsymbol{\varepsilon}}(FIM_{\theta|e}) + FIM_e \quad (\text{A.4})$$

The noise variance and the target parameter are statistically independent, therefore

$$\left[FIM_{\theta_{\varepsilon}} \right]_{i,j} = 2 \operatorname{Re} \left[\frac{\partial \mathbf{u}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{C}_{\omega}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\partial \theta_j} \right] \quad (\text{A.5})$$

$\mathbf{C}_{\omega}(\boldsymbol{\theta})$ Represents covariance matrix of noise,

$$\mathbf{u}(\boldsymbol{\theta}) = \mathbf{y} - \boldsymbol{\omega} \quad (\text{A.6})$$

furthermore

$$E_{\varepsilon} \left(FIM_{\theta_{\varepsilon}} \right) = \frac{2}{\sigma^2} \begin{bmatrix} \operatorname{Re}(\mathbf{U}) & -\operatorname{Im}(\mathbf{U}) & -\operatorname{Im}(\mathbf{T}) \\ \operatorname{Im}(\mathbf{U}) & \operatorname{Re}(\mathbf{U}) & \operatorname{Re}(\mathbf{T}) \\ -\operatorname{Im}(\mathbf{T})^T & \operatorname{Re}(\mathbf{T})^T & \operatorname{Re}(\mathbf{V}) \end{bmatrix} \quad (\text{A.7})$$

Where,

$$\mathbf{U} = \mathbf{Z}_{\varepsilon}^H \mathbf{Z}_{\varepsilon} \quad (\text{A.8})$$

$$\mathbf{T} = 2\pi \mathbf{Z}_{\varepsilon}^H \mathbf{D} \mathbf{Z}_{\varepsilon} \mathbf{H}_0 \quad (\text{A.9})$$

$$\mathbf{V} = 4\pi^2 \mathbf{H}_0^H \mathbf{Z}_{\varepsilon}^H \mathbf{D}^2 \mathbf{Z}_{\varepsilon} \mathbf{H}_0 \quad (\text{A.10})$$

$$\mathbf{H}_0 = \operatorname{diag}(\mathbf{H}) \quad (\text{A.11})$$

$$\mathbf{D} = \operatorname{diag}\{1, 2, \dots, N\} \quad (\text{A.12})$$

and

$$\mathbf{Z}_{\varepsilon} = \begin{bmatrix} x_1[1]e^{j2\pi\varepsilon_1} & x_2[1]e^{j2\pi\varepsilon_2} & \dots & x_{M_T}[1]e^{j2\pi\varepsilon_{M_T}} \\ x_1[2]e^{j4\pi\varepsilon_1} & x_2[2]e^{j4\pi\varepsilon_2} & \dots & x_{M_T}[2]e^{j4\pi\varepsilon_{M_T}} \\ \vdots & \vdots & \dots & \vdots \\ x_1[N]e^{j2N\pi\varepsilon_1} & x_2[N]e^{j2N\pi\varepsilon_2} & \dots & x_{M_T}[N]e^{j2N\pi\varepsilon_{M_T}} \end{bmatrix} \quad (\text{A.13})$$

Additionally, the second part of (A.4) can

$$FIM_{\varepsilon} = -E_{\varepsilon} \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} L(\boldsymbol{\varepsilon}) \right) \quad (\text{A.14})$$

Where,

$$L(\boldsymbol{\varepsilon}) = \ln p(\boldsymbol{\varepsilon}) \quad (\text{A.15})$$

is logarithm probability density function of frequency offset prior information, there are the last M_T non-zero elements in the diagonal of matrix FIM_{ε} . As

$$CRB(\boldsymbol{\theta}) = FIM_{\theta}^{-1} \quad (\text{A.16})$$

$CRB(\boldsymbol{\varepsilon})$ of frequency offset can be obtained by matrix decomposition of FIM_{θ}^{-1}

$$CRB(\boldsymbol{\varepsilon}) = \left\{ \frac{2}{\sigma^2} \operatorname{Re}(\mathbf{V} - \mathbf{T}^H \mathbf{U}^{-1} \mathbf{T}) + \mathbf{R} \right\}^{-1} \quad (\text{A.17})$$

where,

$$\begin{aligned} \mathbf{R} &= E(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}) \\ &= \operatorname{diag} \left\{ 1/\sigma_{\varepsilon_1}^2, 1/\sigma_{\varepsilon_2}^2, \dots, 1/\sigma_{\varepsilon_{M_T}}^2 \right\}_{M_T \times M_T} \end{aligned} \quad (\text{A.18})$$

According to (A.16), CRLB with prior information adds one term \mathbf{R} and because \mathbf{R} is diagonal positive definite matrix, theoretically MAP estimation based on prior information can be proved better performance, since Eq. (A.18) always holds

$$\left\{ \frac{2}{\sigma^2} \operatorname{Re}(\mathbf{V} - \mathbf{T}^H \mathbf{U}^{-1} \mathbf{T}) + \mathbf{R}^{-1} \right\}^{-1} < \left\{ \frac{2}{\sigma^2} \operatorname{Re}(\mathbf{V} - \mathbf{T}^H \mathbf{U}^{-1} \mathbf{T}) \right\}^{-1} \quad (\text{A.19})$$

Acknowledgment

This work was supported by the National Science Foundation of China under Grant number 60602009 61032002 and 71001075, “863” Project under Grant No. 2009AA01Z236 and No.2008AA04A107, Chinese Important National Science & Technology Specific Projects under Grant 2009ZX03005-003.

References

1. R. Etxeberria, P. Larranaga., Global optimization using Bayesian networks[C], The 2nd Symposium on Artificial Intelligence, Habana: CIMAF Press, 1999:332-339.
2. M. Pelikan, D. E. Goldberg and Cantu Paz E. BOA, The Bayesian optimization algorithm[C], Proceedings of the Genetic and Evolutionary Computation Conference. San Mateo: Morgan Kaufmann Publishers, 1999:525-532.
3. H. Mühlenbein, T. Mahnig and A. O. Rodriguez, Schemata, Distributions and graphical models in evolutionary optimization[J], Journal of Heuristics, 1999,(5):215-247.
4. Q Zhang, H Muhlenbein., On the convergence of a class of estimation of distribution algorithms[J], IEEE Trans. on evolutionary computation, 2004, 8(2):127-135.
5. D Heckerman, D Geiger and D M. Chickering, Learning Bayesian Networks: The combination of knowledge and statistical data[J], Machine Learning, 1995, 20(3):197-243.
6. M. Pelikan, K. Sastry, Fitness inheritance in the Bayesian Optimization Algorithm[C], Genetic and Evolutionary

- Computation Conference(GECCO), Heidelberg: Springer Berlin, 2004:48-59.
7. R. Srinivasan., Distributed radar detection theory[C], Proceedings of the Institute of Electronical Engineers(PI F), 1986, 133(1):55~60.
 8. I. Y. Hoballah, and P. K. Varshney., Distributed Bayesian signal detection[J]. IEEE Trans. on Information Theory, 1989, 35(5):995~1000.
 9. A. Assenza, M. Valle and M. Verleysen, A comparative study of various probability density estimation methods for data analysis[J], International Journal of Computational Intelligence Systems, 2008, 1-2:188- 201.
 10. G. Kong, D. Xu and J. Yang, Clinical decision support systems: a review of knowledge representation and inference under uncertainties[J], International Journal of Computational Intelligence Systems, 2008, 1-2:159 – 167.
 11. Y. Liu, J. Huang, A novel fast multi-objective evolutionary algorithm for QoS multicast routing in MANET[J], International Journal of Computational Intelligence Systems, 2009, 2-3: 1875-6883.
 12. W.A. Hashlarmoun, and P.K. Varshnery., An approach to the design of distributed Bayesian decision structures[J]. IEEE Trans. on Systems, Manm and Cybernetics, 1991, 21(5):1206-1210.
 13. Y. Zhang, H. Chen., Bayesian Statistics Conclusion [M]. Publishing House of Science, 1991
 14. Z. Zhang, W. Zhang, C. Tellambura., Robust OFDMA uplink synchronization by exploiting the variance of carrier frequency offsets[J]. IEEE Trans. on Vehicular Technology, 2008, 57(5):3028-3039.
 15. T. Bayes, An essay towards solving a problem in the doctrine of chances[J], Philosophical Trans. of the Royal Society of London, 1763, 53:370-418.; reprinted in Biometrike, 1958, 45:293-315.
 16. Q. Xu, S. Lv, Probability and Mathematical Statistics[M], Higher Education Press, July 2004.
 17. P. A. Parker, P. Mitran and D. W. Bliss, et al. On bounds and algorithms for frequency synchronization for collaborative communication systems[J]. IEEE Trans. on Signal Processing, 2008, 56(8):3742-3752.
 18. Q. Wang, X. Lei and J. Du, et al. A novel MAP estimation algorithm for multiple frequency offsets in distributed MIMO system[C], IEEE Wireless Communication, Networking and Mobile Computing, 2009.05,page:1-4.
 19. Q. Wang, X. Lei and Y. Xiao et al. A Quasi-MAP estimation algorithm for multiple frequency offsets in distributed MIMO system[C], the First International Conference on Wireless Communications and Signal Processing (WCSP), 2009.11,page:1-5.
 20. The-Hanh Pham, Nallanathan A. and Y. Liang. Joint channel and frequency offset estimation in distributed MIMO flat-fading channels[J]. IEEE Trans. Wireless Commun., Feb. 2008, 7(2): 648 – 656.