Importance Sampling Based Direct Maximum Likelihood Position Determination of multiple emitters using finite measurements

Kegang Hao, Qun Wan*

School of Information and Communication Engineering, University of Electronic Science and Technology of China

Abstract

The exact Direct Maximum Likelihood Position Determination (ML-DPD) of multiple emitters requires a multidimensional searching with exponential complexity. The subspace-based DPD technique [1] and the filtering-based DPD technique [2, 3] are recently proposed to reduce the complexity by transforming the original high-dimensional optimization problem into several low-dimensional ones. These techniques, however, approximate the exact ML solution only when the measurements are assumed infinite. In this paper, we model the received signals in the *nonlinear regression* form given by [4] and then resort to the multidimensional optimization framework, consisting of the Pincus' theorem and the Importance Sampling (IS) technique, to approximate the exact ML solution using finite measurements. The proposed IS based ML DPD approach also reduces the complexity by decoupling the multidimensional optimization problem into several three-dimensional ones. Moreover, the new non-iterative ML DPD approach guarantees global optimality and does not suffer from the off-grid problems which are inherent to most DPD technique. The numerical simulations show that the proposed ML DPD estimator also acquires the exact multidimensional ML solution using a small bandwidth or at low Signal to Noise Ratio (SNR).

*Corresponding author

Email addresses: haokegang@std.uestc.edu.cn (Kegang Hao), wanqun@uestc.edu.cn (Qun Wan)

Keywords: Direct Position Determination, Mean Maximum Likelihood, Importance Sampling, Multiple Emitters, Finite Measurements

1. Introduction

The source localization technique has been studied for several decades and applied in many practical scenarios to provide the location information, for instance, the Location-Based Services (LBS) in communication systems, objects tracking in Internet of Things (IoT) and targets detection in radar system. The classical localization methods are two-step processing. Firstly, intermediate parameters that rely on the locations of the sources, like the angle of arrival (AOA), the time difference of arrival (TDOA), the doppler frequency shift (DFS) or the received signal strength (RSS), are estimated from the measurements of different receivers respectively. The locations of sources are, then, estimated based on these intermediate parameters. The two-step methods are sub-optimal since they measure the intermediate parameters at each receiver separately and independently, which ignores the constraint that the measurements from different receivers have to correspond to the same source position. The DPD approaches have recently been proposed as a single-step localization technique where the locations of sources are estimated by minimizing a single cost function, into which all received data enter jointly. It has been verified that the DPD methods outperform the two-step methods especially at low SNR and the two methods' performance converge together as SNR increases. Furthermore, an additional data association step (to partition the intermediate parameters corresponding to the same source location into one set) is required in multiple sources scenario, which is avoided inherently in DPD methods.

In multiple sources scenario, the exact multidimensional ML estimator can be derived but it requires a multi-dimensional searching which is usually impractical. To sidestep the multi-dimensional searching in multiple sources scenario, several existing DPD methods abandon the exact ML estimator and resort to the subspace-based technique or the filtering technique such as the

well-known Multiple Signal Classification (MUSIC) [1], the Match Filter (MF) [2] and the Minimum Variance Distortionless Response (MVDR) filter [3]. These techniques transform the multiple sources localization problem into several single ones which requires only two- or three-dimensional space searching. These methods, however, require enough observation samples to achieve the asymptotically optimal performance at high SNR. To be specific, the MUSIC-DPD [1] and MVDR-DPD [3] both need a lots of observation samples to approximate the covariance matrix by the sample covariance matrix and the MF-DPD [2] also need enough observation samples to decouple the problem based on the independence among transmitted signals. While the number of observation samples is often limited for some practical reasons such as the request for real-time processing or the assumption that sources are stationary during the observation. The performance of these estimators will deteriorate when there are only a few observation samples available. Besides, the iterative ML methods [5] is proposed to achieve the exact ML solution efficiently. As in many iterative algorithms, the convergence to the global minimum of the cost function, however, depends on the initial estimate.

In this paper, we resort to the optimization framework [4], consisting of the Pincus' theorem [6] and the IS technique, to efficiently approximate the global solution of the multi-dimensional ML problem, using a few observation samples. The major difficulty to use IS technique consists in generating the effective parameter realizations (i.e., sampling in the vicinity of the exact ML solutions) according to a given (multi-dimensional) pdf. Much like all the IS-based works mentioned in section 4, we design a factorable pdf on all involved parameters, which allows a very easy generation of the required parameter realizations. Numerical simulations show that the proposed IS-based multidimensional ML estimator outperforms the subspace-based DPD techniques and the filtering-based DPD techniques in multiple sources scenario.

2. Problem Formulation

Given Q adjacent narrow-band emitters at $\boldsymbol{p}_q \in \mathbb{R}^{D\times 1}, q = 1, ..., Q$ (In general, $D = 2$ for plane geometry or $D = 3$ for solid geometry) and N_r space separated receivers which intercept the transmitted signals. Every receiver is equipped with an antenna array consisting of M elements. To avoid the angle ambiguity, the carrier frequency is small than the inverse of the propagation time over the array aperture. The complex envelopes of the signals observed by the jth receiver are given by

$$\boldsymbol{y}_j(t) \triangleq \sum_{q=1}^{Q} \alpha_{q,j}\boldsymbol{a}_j(\boldsymbol{p}_q)s_q(t-\tau_j(\boldsymbol{p}_q)-\mathring{t}_q)+\boldsymbol{w}_j(t), \quad 0 \le t \le T \tag{1}$$

where $\boldsymbol{y}_j(t) \in \mathbb{C}^{M\times 1}$ is the jth received signal, $\alpha_{q,j}$ is an unknown complex scalar representing the channel attenuation between the qth emitter and the jth receiver, $\boldsymbol{a}_j(\boldsymbol{p}_q) \in \mathbb{C}^{M\times 1}$ is the jth array response to the signal transmitted from position $\boldsymbol{p}_q$ and $s_q(t-\tau_j(\boldsymbol{p}_q)-\mathring{t}_q)$ is the copy of the qth transmitted signal waveform, which is transmitted at time $\mathring{t}_q$ and is received by the jth receiver after being delayed by $\tau_j(\boldsymbol{p}_q)$. The vector $\boldsymbol{w}_j(t) \in \mathbb{C}^{M\times 1}$ represents noise and interference observed by the jth array. The received signal can be partitioned into L sections and the length of each section satisfies $T/L \gg \max_{j,q}\tau_j(\boldsymbol{p}_q)$ to avoid the delay ambiguity. The emitters are assumed stationary during the total observation time T. Each section can be Discrete Fourier Transformed (DFT) into

$$\boldsymbol{y}_j(n,l) = \sum_{q=1}^{Q} \alpha_{q,j}\boldsymbol{a}_j(\boldsymbol{p}_q)e^{-j2\pi f_n(\tau_j(\boldsymbol{p}_q)+\mathring{t}_q)}s_q(n,l)+\boldsymbol{w}_j(n,l) \tag{2}$$

$$j = 1, ..., N_r; n = 1, ..., N; l = 1, ..., L$$

where $\boldsymbol{y}_j(n,l) \in \mathbb{C}^{M\times 1}$ is the nth Fourier coefficient vector of the lth section of the jth received signal, $s_q(n,l)$ is the nth Fourier coefficient of the lth section of the qth transmitted signal waveform, and $\boldsymbol{w}_j(n,l) \in \mathbb{C}^{M\times 1}$ represents the nth Fourier coefficient vector of the lth section of the jth noise waveform.

For compact representation, we define the following vectors:

$$\boldsymbol{\gamma}_{j,q,n,l}(\boldsymbol{p}_q, \mathring{t}_q | s_q(n,l)) \triangleq \boldsymbol{a}_j(\boldsymbol{p}_q) e^{-j2\pi f_n(\tau_j(\boldsymbol{p}_q)+\mathring{t}_q)} s_q(n,l) \tag{3}$$

Here we assume that the spectrum of transmitted signals, like the synchronization and training sequences in cellular systems, are known to location system but the transmitted time $\mathring{t}_q$, which is assumed known in some DPD-related works [1, 5], is assumed unknown here (for the precise clock-synchronization between emitter and receiver is a challenging task). We observe that all information about the emitters' positions is embedded in the manifold vector $\boldsymbol{\gamma}_{j,q,n,l}$ (for the sake of brevity, the dependence on $\boldsymbol{p}_q$, $\mathring{t}_q$ and $s_q(n,l)$ is omitted). The harmonic-superposition form of (2) is given by:

$$\boldsymbol{y}_j(n,l) = \sum_{q=1}^{Q} \alpha_{q,j} \boldsymbol{\gamma}_{j,q,n,l} + \boldsymbol{w}_j(n,l) \tag{4}$$

which can be written in compact matrix form as follows,

$$\boldsymbol{y}_j(n,l) = \boldsymbol{\Gamma}_{j,n,l}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) \boldsymbol{\alpha}_j + \boldsymbol{w}_j(n,l) \tag{5}$$

where

$$\begin{aligned} \boldsymbol{\Gamma}_{j,n,l}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) &\triangleq [\boldsymbol{\gamma}_{j,1,n,l}, ..., \boldsymbol{\gamma}_{j,Q,n,l}] \in \mathbb{C}^{M \times Q} \\ \boldsymbol{\alpha}_j &\triangleq [\alpha_{1,j}, ..., \alpha_{Q,j}]^T \in \mathbb{C}^{Q \times 1} \end{aligned} \tag{6}$$

$\boldsymbol{P}, \mathring{\boldsymbol{t}}$ are the sets of emitters' positions and transmitted times respectively.

Collecting and stacking all the N coefficient vectors of the lth section of the jth received signal,

$$\boldsymbol{y}_j(l) = \boldsymbol{\Gamma}_{j,l}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) \boldsymbol{\alpha}_j + \boldsymbol{w}_j(l) \tag{7}$$

where

$$\begin{aligned} \boldsymbol{y}_j(l) &\triangleq [\boldsymbol{y}_j^T(1,l), ..., \boldsymbol{y}_j^T(N,l)]^T \in \mathbb{C}^{NM \times 1} \\ \boldsymbol{\Gamma}_{j,l}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) &\triangleq [\boldsymbol{\Gamma}_{j,1,l}^T(\boldsymbol{P}, \mathring{\boldsymbol{t}}), ..., \boldsymbol{\Gamma}_{j,N,l}^T(\boldsymbol{P}, \mathring{\boldsymbol{t}})]^T \in \mathbb{C}^{NM \times Q} \\ \boldsymbol{w}_j(l) &\triangleq [\boldsymbol{w}_j^T(1,l), ..., \boldsymbol{w}_j^T(N,l)]^T \in \mathbb{C}^{NM \times 1} \end{aligned} \tag{8}$$

The received data model (7) is exactly the *nonlinear regression* model given by [4], which is suitable to the IS based global optimization framework.

The problem that we address now is stated as follows: Given the measurements $\{\boldsymbol{y}_j(l)\}_{j,l=1}^{N_r,L}$, how to efficiently and accurately estimate the locations of the emitters $\boldsymbol{P}$.

3. the Concentrated Likelihood Function

In this subsection, we will derive the *Concentrated* Likelihood Function (CLF) that depends on the parameters of interest $\boldsymbol{P}$ and the auxiliary parameters $\mathring{\boldsymbol{t}}$ (which are conducive to acquiring the location information from the propagation delay $\tau_j(\boldsymbol{p})$ since the receivers are assumed synchronous). In fact, since $\boldsymbol{w}_j(l)$ in (7) are independent among different sections of different received signals and follow the identity distribution $\boldsymbol{w}_j(l) \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{NM})$, it can be shown that the Log Likelihood Function (LLF), of which the constant terms are dropped, is given by,

$$\mathcal{L}(\boldsymbol{P}, \mathring{\boldsymbol{t}}, \boldsymbol{\alpha}) \triangleq -\sum_{j=1}^{N_r} \sum_{l=1}^{L} \|\boldsymbol{y}_j(l) - \boldsymbol{\Gamma}_{j,l}(\boldsymbol{P}, \mathring{\boldsymbol{t}})\boldsymbol{\alpha}_j\|^2 \tag{9}$$

To get a compact CLF, we define the manifold matrix for the jth received signal by collecting and stacking all L signal sections,

$$\boldsymbol{\Gamma}_j(\boldsymbol{P}, \mathring{\boldsymbol{t}}) \triangleq [\boldsymbol{\Gamma}_{j,1}^T, ..., \boldsymbol{\Gamma}_{j,L}^T]^T \in \mathbb{C}^{LNM \times Q} \tag{10}$$

The CLF (9) can be rewritten as follows,

$$\mathcal{L}(\boldsymbol{P}, \mathring{\boldsymbol{t}}, \boldsymbol{\alpha}) = -\sum_{j=1}^{N_r} \|\boldsymbol{y}_j - \boldsymbol{\Gamma}_j(\boldsymbol{P}, \mathring{\boldsymbol{t}})\boldsymbol{\alpha}_j\|^2 \tag{11}$$

where $\boldsymbol{y}_j \triangleq [\boldsymbol{y}_j^T(1), ..., \boldsymbol{y}_j^T(L)]^T$. It is extremely challenging to maximize (11) with respect to all parameters jointly. Fortunately, for any given $\boldsymbol{P}, \mathring{\boldsymbol{t}}$, the problem of finding the optimal $\{\boldsymbol{\alpha}_j\}_{j=1}^{N_r}$ becomes a linear least squares (LS) problem [7] whose solution is given by

$$\hat{\boldsymbol{\alpha}}_j = \boldsymbol{\Gamma}_j^{\dagger}(\boldsymbol{P}, \mathring{\boldsymbol{t}})\boldsymbol{y}_j \tag{12}$$

where $\boldsymbol{\Gamma}_j^{\dagger}$ is the Moore-Penrose pseudo-inverse of $\boldsymbol{\Gamma}_j$ given by $(\boldsymbol{\Gamma}_j^H\boldsymbol{\Gamma}_j)^{-1}\boldsymbol{\Gamma}_j^H$. Note that $(\boldsymbol{\Gamma}_j^H\boldsymbol{\Gamma}_j)^{-1}$ always exists since $\boldsymbol{\Gamma}_j$ has full column rank.

Substituting the estimations $\{\hat{\boldsymbol{\alpha}}_j\}_{j=1}^{N_r}$ back into (11), we obtain the so-called *concentrated* likelihood function after performing some straightforward algebraic operations,

$$\mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}}) = \sum_{j=1}^{N_r} \boldsymbol{y}_j^H \boldsymbol{\Gamma}_j (\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j)^{-1} \boldsymbol{\Gamma}_j^H \boldsymbol{y}_j \tag{13}$$

The joint ML estimates of $\boldsymbol{P}$ and $\mathring{\boldsymbol{t}}$ are then obtained as,

$$[\hat{\boldsymbol{P}}, \hat{\mathring{\boldsymbol{t}}}] = arg \max_{\boldsymbol{P}, \mathring{\boldsymbol{t}}} \mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}}) \tag{14}$$

The reduced-dimension optimization problem (14) still has $Q(D+1)$ unknown parameters (including Q D-dimensional location vectors $\{\boldsymbol{p}_q\}_{q=1}^{Q}$ and Q transmitted times scalar $\{\mathring{t}_q\}_{q=1}^{Q}$). Thus, the exact multi-dimensional ML estimation will require an impractical multi-dimensional search over the parameter space, which takes exponential complexity.

To solve the multidimensional ML problem (14) efficiently, the approach proposed in [1] is to transform the problem into several D-dimensional optimization problems with assuming the number of sections $L \to \infty$. Considering the length of section is limited by $T/L \gg \max \tau_j(\boldsymbol{p}_q)$, the number of sections L is then proportional to the total observation time T. The observation time T is, however, limited in some practical reasons such as the request for real-time processing or the assumption that emitters are stationary during the observation. The MF based DPD approach proposed in [1] may, therefore, be failure for the finite signal sections in some practical cases, especially when the emitters are adjacent to one another.

4. Global Maximization of The CLF

Like the previous works within other estimation problems ([4, 8, 9, 10, 11]), we resort to the optimization framework [4], consisting of the Pincus' theorem [6] and the IS technique, to efficiently approximate the global solution of the

multi-dimensional ML problem, using finite observation resources such as the observation time, the signal bandwidth and the array aperture. First of all, we introduce the theorem of Pincus [6] as follows.

Theorem 1. Let $F(x_1, ..., x_n) = F(\boldsymbol{x})$ be a continuous function on the closure of a bounded domain S in n-dimensional Euclidean space of $\mathbb{R}^n$. Assume that F attains a global maximum at exactly on point $\hat{\boldsymbol{x}} = [\hat{x}_1, ..., \hat{x}_n]^T$ of S. Then for $i = 1, ..., n$,

$$\hat{x}_i = \lim_{\rho \to \infty} \frac{\int \cdots \int x_i e^{\rho F(\boldsymbol{x})} dx_1 \cdots dx_n}{\int \cdots \int e^{\rho F(\boldsymbol{x})} dx_1 \cdots dx_n} \tag{15}$$

Applying this general result to our multidimensional optimization problem (14) with $\boldsymbol{x} \triangleq [\boldsymbol{P}, \mathring{\boldsymbol{t}}]$ and $F(\boldsymbol{x}) \triangleq \mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ leads to the following ML estimation for the emitter locations and the transmitted times respectively,

$$\hat{\boldsymbol{p}}_q(d) = \int \cdots \int \boldsymbol{p}_q(d) \mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) d\boldsymbol{p}_1(1) \cdots d\mathring{t}_Q \tag{16}$$

$$\hat{\mathring{t}}_q = \int \cdots \int \mathring{t}_q \mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) d\boldsymbol{p}_1(1) \cdots d\mathring{t}_Q \tag{17}$$

$$q = 1, ..., Q; d = 1, ..., D$$

where

$$\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) \triangleq \lim_{\rho \to \infty} \frac{e^{\rho \mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}})}}{\int \cdots \int e^{\rho \mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}})} d\boldsymbol{p}_1(1) \cdots d\mathring{t}_Q} \tag{18}$$

The Pincus' theorem can be interpreted intuitively as follows: as ρ tends to infinity (ρ taking a sufficiently large value in practice), $\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ becomes a Dirac-delta function centered at the global maximum of $\mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}})$, which eliminates the integrals in (16) and (17) and chooses its center (i.e., the global ML solution) as the result directly.

Nevertheless, the multidimensional grid searching we attempt to avoid is transformed into several multidimensional integrations in (16) and (17), which bear the same computational load. By closely analyzing the properties of function (18), it turns out that $\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ can be seen as a pdf since it is nonnegative

and integrates to one. As a pdf, $\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ indicates the probability of the candidate parameter values being true. Furthermore, the estimators in (16) and (17) can be regarded as statistical expectations, i.e., we have,

$$\hat{\boldsymbol{p}}_q(d) \triangleq \mathbb{E}_{\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})}\{\boldsymbol{p}_q(d)\} \quad \hat{\mathring{t}}_q \triangleq \mathbb{E}_{\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})}\{\mathring{t}_q\} \tag{19}$$

Consequently, the multidimensional integrations are transformed into the statistical expectations, which can be approximated by the sample mean,

$$\hat{\boldsymbol{p}}_q(d) = \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{p}_q^{(r)}(d) \quad \hat{\mathring{t}}_q = \frac{1}{R}\sum_{r=1}^{R} \mathring{t}_q^{(r)} \tag{20}$$

Clearly, as the number of realizations R increases, the two sample mean in (20) approach the statistical expectations in (19). Unfortunately, the extremely nonlinear property of the pdf $\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ makes it impractical to generate realizations $\{\boldsymbol{P}^{(r)}\}_{r=1}^{R}$ and $\{\mathring{\boldsymbol{t}}^{(r)}\}_{r=1}^{R}$. To sidestep this problem, we can resort to the IS method [4, 8] and rewrite (16) and (17) in the following equivalent forms,

$$\hat{\boldsymbol{p}}_q(d) = \int \cdots \int \boldsymbol{p}_q(d) \frac{\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})}{\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})} \mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) d\boldsymbol{p}_1(1) \cdots d\mathring{t}_Q \tag{21}$$

$$\hat{\mathring{t}}_q = \int \cdots \int \mathring{t}_q \frac{\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})}{\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})} \mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) d\boldsymbol{p}_1(1) \cdots d\mathring{t}_Q \tag{22}$$

where $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ is another pdf called *importance function* (IF). By introducing the IF, the multidimensional integrations in (21) and (22) are interpreted as statistical expectations of two transformed random variates (RVs), i.e.,

$$\hat{\boldsymbol{p}}_q(d) \triangleq \mathbb{E}_{\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})}\{\eta(\boldsymbol{P}, \mathring{\boldsymbol{t}})\boldsymbol{p}_q(d)\} \quad \hat{\mathring{t}}_q \triangleq \mathbb{E}_{\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})}\{\eta(\boldsymbol{P}, \mathring{\boldsymbol{t}})\mathring{t}_q\} \tag{23}$$

where $\eta(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ is called *importance weight* (IW) and defined as,

$$\eta(\boldsymbol{P}, \mathring{\boldsymbol{t}}) \triangleq \frac{\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})}{\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})} \tag{24}$$

If IF $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ is carefully designed, the realizations $\{\boldsymbol{P}^{(r)}\}_{r=1}^{R}$ and $\{\mathring{\boldsymbol{t}}^{(r)}\}_{r=1}^{R}$ corresponding to large IWs can be generated from the IF $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ efficiently.

the expectations in (23) can be approximated at any desired degree of accuracy (by increasing R) using the corresponding sample mean estimates. In fact, the choice of $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ is a tradeoff between the ease and efficiency of the sampling (i.e., the generation of the required realizations). The appropriate choice of the IF will be discussed in the following section.

5. Appropriate Choice of Importance Function

To generate the required realizations efficiently, the appropriate $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ must be designed as close as possible to $\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ while it should also take the ease of generating the realizations into account.

In order to find the appropriate $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$, we start with constraining IF's form to be separable in terms of the Q position-transmitted time pairs $\{(\boldsymbol{p}_q, \mathring{t}_q)\}_{q=1}^{Q}$ for it is relative easy to generate realizations from pdf with the factorized form as follows,

$$\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}}) \triangleq \prod_{q=1}^{Q} g_q(\boldsymbol{p}_q, \mathring{t}_q) \tag{25}$$

According to the general results from probability theory, the independence among Q groups of RV pairs $\{(\boldsymbol{p}_q, \mathring{t}_q)\}_{q=1}^{Q}$ is introduced by the assumption for the factorized form in (25). Hence, we can independently generate Q groups of realizations for $(D+1)$-dimensional random vectors $\{\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(r)}\}_{r=1}^{R}, q = 1, ..., Q$ using $g_q(\boldsymbol{p}_q, \mathring{t}_q), q = 1, ..., Q$ respectively, instead of using $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ to generate realizations for $Q(D+1)$-dimensional random vector $\{\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}\}_{r=1}^{R}$ with difficulty.

According to the definition of $\mathcal{F}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ (18), the assumption for the factorized form in (25) requests that the CLF $\mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ in (18) has summation form with respect to q. Note that the product $\boldsymbol{y}_j^H \boldsymbol{\Gamma}_j$ in CLF (13) is a Q-dimensional vector. Diagonalizing the inverse matrix $(\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j)^{-1}$ is, therefore, the key step to the separable form (25). According to the definition of $\boldsymbol{\Gamma}_j$ in (10), $\boldsymbol{\Gamma}_{j,l}$ in (8), $\boldsymbol{\Gamma}_{j,n,l}$ in (6) and $\boldsymbol{\gamma}_{j,q,n,l}$ in (3), the (u, v)th element of the normalized matrix

product $\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j$ is given by,

$$[\frac{1}{MNL}\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j]_{u,v} \tag{26}$$

$$= \frac{1}{MNL}\sum_{l=1}^{L}\sum_{n=1}^{N}\boldsymbol{\gamma}_{j,u,n,l}^H \boldsymbol{\gamma}_{j,v,n,l} \tag{27}$$

$$= \underbrace{\frac{1}{M}\boldsymbol{a}_j^H(\boldsymbol{p}_u)\boldsymbol{a}_j(\boldsymbol{p}_v)}_{r_{doa,j}^{(u,v)}} \cdot \underbrace{\frac{1}{N}\sum_{n=1}^{N} e^{j2\pi f_n(\tau_j(\boldsymbol{p}_u)+\mathring{t}_u-\tau_j(\boldsymbol{p}_v)-\mathring{t}_v)}}_{r_{toa,j}^{(u,v)}} \cdot \underbrace{\frac{1}{L}\sum_{l=1}^{L} s_u^*(n,l)s_v(n,l)}_{r_{sig,n}^{(u,v)}} \tag{28}$$

where $r_{doa,j}^{(u,v)}, r_{toa,j}^{(u,v)}, r_{sig,n}^{(u,v)}$ denote the (u,v)th normalized correlation coefficients in aspects of the jth AOA parameter, the jth TOA parameter and the nth Fourier coefficient of the transmitted signal, respectively. The $Q \times Q$ matrix $\frac{1}{MNL}\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j$ is, therefore, a normalized correlation matrix among all emitters. In order to diagonalize the normalized correlation matrix $\frac{1}{MNL}\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j$, we set its off-diagonal elements to zeros at the cost of losing the correlation information. As a consequence, the expected diagonal matrix, for $j = 1, ..., N_r$, is given by,

$$\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j = MNL \cdot \boldsymbol{I} \tag{29}$$

where $\boldsymbol{I}$ is the identity matrix, the coefficient MNL is the diagonal element of $\boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j$.

It is worthy to note that the lost correlation information can be ignored when there are enough observation resources. To be specific, large array aperture (with large M), large signal bandwidth (with large N) or many signal sections (with large L) all make the off-diagonal correlation coefficients such small that the corresponding correlation information is negligible, which is the essential assumption in the MF-DPD technique [2]. While the same correlation information is embedded in the IWs (24) in our IS based optimization framework so the IF can be designed without the correlation information as long as the realizations in the vicinity of the exact ML solutions can be generated.

Substituting (29) into (13) and ignoring the constant coefficient, we get the

summation form of CLF as follows,

$$\tilde{\mathcal{L}}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}}) = \sum_{j=1}^{N_r} \|\boldsymbol{\Gamma}_j^H \boldsymbol{y}_j\|^2 \tag{30}$$

$$= \sum_{q=1}^{Q} \tilde{\ell}_c(\boldsymbol{p}_q, \mathring{t}_q) \tag{31}$$

where

$$\tilde{\ell}_c(\boldsymbol{p}_q, \mathring{t}_q) \triangleq \sum_{j,n,l=1}^{N_r,N,L} |\boldsymbol{\gamma}_{j,q,n,l}^H \boldsymbol{y}_j(n,l)|^2 \tag{32}$$

Replacing $\mathcal{L}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ and ρ in (18) with $\tilde{\mathcal{L}}_c(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ and an appropriate ρ_2 respectively, we get the factorized form of $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$, of which the qth factor is given by,

$$g_q(\boldsymbol{p}_q, \mathring{t}_q) \triangleq \frac{e^{\rho_2 \tilde{\ell}_c(\boldsymbol{p}_q, \mathring{t}_q)}}{\int\int e^{\rho_2 \tilde{\ell}_c(\boldsymbol{p}_q, \mathring{t}_q)} d\boldsymbol{p}_q d\mathring{t}_q} \tag{33}$$

The generation of qth group of realizations $\{(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(r)})\}_{r=1}^R$ follows, furthermore, the well-known general result from probability theory that the joint pdf $g_q(\boldsymbol{p}_q, \mathring{t}_q)$ can be factorized as the product of marginal and conditional pdfs,

$$g_q(\boldsymbol{p}_q, \mathring{t}_q) = g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q) g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q | \boldsymbol{p}_q) \tag{34}$$

The usual method to obtain the marginal pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q)$ is integrating the joint pdf $g_q(\boldsymbol{p}_q, \mathring{t}_q)$ with respect to the transmitted time $\mathring{t}_q$. The analytical integration is, however, difficult since $g_q(\boldsymbol{p}_q, \mathring{t}_q)$ is extremely non-linear about $\mathring{t}_q$. The discrete approximation also involves $(D+1)$-dimensional grid search with exponential complexity. Fortunately, we find another alternative method to construct $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q)$ directly and efficiently,

$$g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q) \triangleq \frac{e^{\rho_1 \ell_q(\boldsymbol{p}_q)}}{\int e^{\rho_1 \ell_q(\boldsymbol{p}_q)} d\boldsymbol{p}_q} \tag{35}$$

where $\ell_q(\boldsymbol{p}_q)$ is the *single* Likelihood Function (SLF) of the position $\boldsymbol{p}_q$, which is derived from (32) by eliminating the transmitted time $\mathring{t}_q$. The detailed derivation of SLF $\ell_q(\boldsymbol{p}_q)$ is put in Appendix A.

Given position realization $\boldsymbol{p}_q^{(r)}$ generated from $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q)$, the conditional transmitted time pdf is defined as,

$$g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\boldsymbol{p}_q^{(r)}) = \frac{g_q(\boldsymbol{p}_q^{(r)}, \mathring{t}_q)}{\int g_q(\boldsymbol{p}_q^{(r)}, \mathring{t}_q) d\mathring{t}_q} \tag{36}$$

Substituting (33) into (36), we have,

$$g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\boldsymbol{p}_q^{(r)}) = \frac{e^{\rho_2 \tilde{\ell}_c(\boldsymbol{p}_q^{(r)}, \mathring{t}_q)}}{\int e^{\rho_2 \tilde{\ell}_c(\boldsymbol{p}_q^{(r)}, \mathring{t}_q)} d\mathring{t}_q} \tag{37}$$

Note that we use the scalar integration $\int e^{\rho_2 \tilde{\ell}_c(\boldsymbol{p}_q^{(r)}, \mathring{t}_q)} d\mathring{t}_q$, instead of $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q^{(r)})$, to normalize the conditional pdf (36) correctly.

The process of generating Q groups of the required realizations $\{(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(r)})\}_{r=1}^R$ are partitioned into two phases: for qth group of realizations, firstly, generate $\{\boldsymbol{p}_q^{(r)}\}_{r=1}^R$ using the marginal position pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q)$ (A.8) and then use the conditional pdf $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\boldsymbol{p}_q^{(r)})$ (37) to generate $\{\mathring{t}_q\}_{r=1}^R$.

6. Generations Of The Required Realizations

We recall the following lemma [12] that will be used to generate the required realizations:

Lemma 1. Let $X \in \mathcal{X}$ be any RV with pdf $f_X(x)$ and the Cumulative Distribution Function (CDF) $F_X(x)$ and denote the inverse CDF as $F_X^{-1}(\cdot) : [0,1] \to \mathcal{X}, u \to x \ s.t. F_X(x) = u$. Then, for any uniform RV, $U \in [0,1]$, the RV $\tilde{X} = F_X^{-1}(U)$ is distributed according to $f_X(\cdot)$.

According to the generating process described before, firstly, we should use the marginal position pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q)$ (A.8) to generate $\{\boldsymbol{p}_q^{(r)}\}_{r=1}^R$. $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q)$ is, however, unsuitable to be used along with the result of Lemma 1 since $\boldsymbol{p}_q$ is a D-dimensional vector (defining $\boldsymbol{p}_q = (x_q, y_q)$ for $D = 2$) instead of a scalar. To acquire the $D = 2$ scalar pdfs, consequently, $g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q)$ should be factorized as follows,

$$g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q) \triangleq g_x^{\mathrm{mar}}(x_q) g_{y|x}^{\mathrm{con}}(y_q|x_q) \tag{38}$$

where

$$g_x^{\mathrm{mar}}(x_q) \triangleq \int g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q) dy_q \tag{39}$$

$$g_{y|x}^{\mathrm{con}}(y_q|x_q) \triangleq \frac{g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q)}{g_x^{\mathrm{mar}}(x_q)} \tag{40}$$

In conclusion, the approach of generating the qth group of realizations $\{(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(r)})\}_{r=1}^R$ can be stated as follows: generate R realizations $\{u_x^{(r)}\}_{r=1}^R \sim U[0,1]$ and obtain $x_q^{(r)} = G_{x_q}^{-1}(u_x^{(r)})$ where $G_{x_q}(\cdot)$ is the CDF associated to $g_x^{\mathrm{mar}}(x_q)$ in (39); For $r = 1, ..., R$, then, generate two realizations $u_y^{(r)}, u_{\mathring{t}}^{(r)} \sim U[0,1]$ and obtain $y_q^{(r)} = G_{y_q|x_q^{(r)}}^{-1}(u_y^{(r)})$ and $\mathring{t}_q^{(r)} = G_{\mathring{t}_q|\boldsymbol{p}_q^{(r)}}^{(-1)}(u_{\mathring{t}}^{(r)})$ where $G_{y_q|x_q}(\cdot)$ and $G_{\mathring{t}_q|\boldsymbol{p}_q}(\cdot)$ are the CDF associated to $g_{y|x}^{\mathrm{con}}(y_q|x_q)$ in (39) and $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\boldsymbol{p}_q^{(r)})$ in (36) respectively.

7. Implementation Details

In this section, we give all the necessary details for an efficient implementation of our proposed IS-based ML DPD algorithm. The plane area of interest, where emitters may be put in, is confined within $x \in [x_{\min}, x_{\max}]$ and $y \in [y_{\min}, y_{\max}]$ and the range of transmitted time is $\mathring{t} \in [0, \mathring{t}_{\max}]$, where $\mathring{t}_{\max}$ can be freely chosen as high as desired. The process of generating the required Q groups of realizations amounts to performing the following steps for every $q = 1, ..., Q$:

STEP 1: Evaluate the qth SLF (A.7) at $I \times J$ two-dimensional grid points $\{(x_q^{(i)}, y_q^{(j)})\}_{i,j=1}^{I,J}$ in the area of interest $[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$, with relatively large step Δx and Δy (a rough search). Then, substitute SLF values into the qth marginal position pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q)$ in (35) and approximate integrals with discrete summations evaluate as follows,

$$g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q^{(i)}, y_q^{(j)}) \simeq \frac{e^{\rho_1 \ell_q(x_q^{(i)}, y_q^{(j)})}}{\sum_{i=1}^{I}\sum_{j=1}^{J} e^{\rho_1 \ell_q(x_q^{(i)}, y_q^{(j)})} \Delta x \Delta y} \tag{41}$$

STEP 2: Initialize the estimation of the qth emitter position as the grid point corresponding to the maximum peak of $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q^{(i)}, y_q^{(j)})$,

$$[\mathring{x}_q, \mathring{y}_q] = arg \max_{x^{(i)}, y^{(j)}} g_{\boldsymbol{p}}^{\mathrm{mar}}(x^{(i)}, y^{(j)}) \tag{42}$$

STEP 3: Substitute the initial position estimation $\mathring{\boldsymbol{p}}_q = (\mathring{x}_q, \mathring{y}_q)$ into the conditional transmitted time pdf $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\mathring{\boldsymbol{p}}_q)$ in (36), and then evaluate it at S grid points $\{\mathring{t}_q^{(s)}\}_{s=1}^{S}$ in the interval $[0, \mathring{t}_{\max}]$, with relatively large step $\Delta\mathring{t}$ (a rough search) as follows,

$$g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q^{(s)}|\mathring{\boldsymbol{p}}_q) = \frac{g_q(\mathring{\boldsymbol{p}}_q, \mathring{t}_q^{(s)})}{\sum_{s=1}^{S} g_q(\mathring{\boldsymbol{p}}_q, \mathring{t}_q^{(s)})\Delta\mathring{t}} \tag{43}$$

STEP 4: Initialize the estimation of the qth transmitted time $\mathring{t}_q$ as the grid point corresponding to the maximum peak of the conditional transmitted time pdf $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q^{(s)}|\mathring{\boldsymbol{p}}_q)$.

To force the qth group of realizations $\{(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(r)})\}_{r=1}^{R}$ to be generated in the vicinity of the ML solutions $(\hat{\boldsymbol{p}}_q, \hat{\mathring{t}}_q)$, we fix the local parameter space to include the ML solutions as follows,

$$\hat{x}_q \in D_{\mathring{x}_q} \triangleq [\mathring{x}_q - \Delta_x, \mathring{x}_q + \Delta_x] \tag{44}$$

$$\hat{y}_q \in D_{\mathring{y}_q} \triangleq [\mathring{y}_q - \Delta_y, \mathring{y}_q + \Delta_y] \tag{45}$$

$$\hat{\mathring{t}}_q \in D_{\mathring{t}_q} \triangleq [\mathring{t}_q - \Delta_{\mathring{t}}, \mathring{t}_q + \Delta_{\mathring{t}}] \tag{46}$$

STEP 5: Evaluate the qth marginal position pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q)$ at $I' \times J'$ new grid points in local area $D_{\mathring{x}_q} \times D_{\mathring{y}_q}$, with relatively small step $\delta x < \Delta x$ and $\delta y < \Delta y$, as follows,

$$g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q^{(i')}, y_q^{(j')}) \simeq \frac{e^{\rho_1 \ell_q(x_q^{(i')}, y_q^{(j')})}}{\sum_{i'=1}^{I'}\sum_{j'=1}^{J'} e^{\rho_1 \ell_q(x_q^{(i')}, y_q^{(j')})}\delta x \delta y} \tag{47}$$

STEP 6: Compute the marginal coordinate x_q pdf by approximating integrals with discrete summations in (39),

$$g_x^{\mathrm{mar}}(x_q^{(i')}) \simeq \sum_{j'=1}^{J'} g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q^{(i')}, y_q^{(j')})\delta y, \quad \forall x_q^{(i')} \in D_{\mathring{x}_q} \tag{48}$$

STEP 7: Compute the coordinate x_q's CDF as follows:

$$G_{x_q}(x_q^{(r)}) = \sum_{i' \le r} g_x^{\mathrm{mar}}(x_q^{(i')})\delta x, \quad \forall x_q^{(r)} \in D_{\mathring{x}_q} \tag{49}$$

The following steps are designed to generate the rth realization corresponding to the qth emitter $(x_q^{(r)}, y_q^{(r)}, \mathring{t}_q^{(r)})$, i.e., for $r = 1, ..., R$:

STEP 8: Generate the rth RV $u_{x_q}^{(r)} \sim U[0,1]$ and invert $G_{x_q}(\cdot)$ via linear interpolation in order to obtain the rth coordinate x_q realization $x_q^{(r)} = G_{x_q}^{-1}(u_{x_q}^{(r)})$.

STEP 9: Given the rth realization $x_q^{(r)}$, obtain the conditional coordinate y_q pdf from the marginal position pdf (already evaluated in "STEP 5") as follows:

$$g_{y|x}^{\text{con}}(y_q^{(j')}|x_q^{(r)}) \simeq \frac{g_{\boldsymbol{p}}^{\text{mar}}(x_q^{(r)}, y_q^{(j')})}{g_x^{\text{mar}}(x_q^{(r)})} \tag{50}$$

STEP 10: Evaluate the coordinate y_q's CDF $G_{y_q}(y_q^{(r)})$, similarly to $G_{x_q}(x_q^{(r)})$ in (49), and generate the rth coordinate y_q realization $y_q^{(r)} = G_{y_q}^{-1}(u_{y_q}^{(r)})$ using linear interpolation as well.

STEP 11: Given the rth position realization $\boldsymbol{p}_q^{(r)} = (x_q^{(r)}, y_q^{(r)})$, evaluate the qth conditional transmitted time pdf (36) at S' new grid points in the local interval $D_{\mathring{t}_q}$, with relatively small step $\delta\mathring{t} < \Delta\mathring{t}$, as follows,

$$g_{\mathring{t}}^{\text{con}}(\mathring{t}_q^{(s')}|\boldsymbol{p}_q^{(r)}) = \frac{g_q(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(s')})}{\sum_{s'=1}^{S'} g_q(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(s')})\delta\mathring{t}} \tag{51}$$

STEP 12: Evaluate the transmitted time $\mathring{t}_q$'s CDF $G_{\mathring{t}_q|\boldsymbol{p}_q}(\cdot)$, similarly to $G_{x_q}(x_q^{(r)})$ in (49), and generate the rth transmitted time realization $\mathring{t}_q^{(r)} = G_{\mathring{t}_q|\boldsymbol{p}_q}^{-1}(u_{\mathring{t}_q}^{(r)})$ using linear interpolation as well.

All the Q groups of required realizations $\{(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(r)})\}_{r,}^{R}, q = 1, ..., Q$ can be generated according to the above steps . The qth expectations in (23), which is equal to the ML solutions, can be approximated based on these realizations as follows,

$$\hat{\boldsymbol{p}}_q = \frac{1}{R}\sum_{r=1}^{R} \eta(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)})\boldsymbol{p}_q^{(r)}, \quad \hat{\mathring{t}}_q = \frac{1}{R}\sum_{r=1}^{R} \eta(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)})\mathring{t}_q^{(r)} \tag{52}$$

According to the definitions (24), (18), (25), (33), (34), (35) and (37), the rth IW can be written as,

$$\eta(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) \triangleq A e^{\phi(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)})} \tag{53}$$

$$\phi(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) \triangleq \rho_0 \mathcal{L}_c(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) - \sum_{q=1}^{Q}[\rho_1 \ell_q(\boldsymbol{p}_q^{(r)}) + \rho_2 \tilde{\ell}_c(\boldsymbol{p}_q^{(r)}, \mathring{t}_q^{(r)})] \tag{54}$$

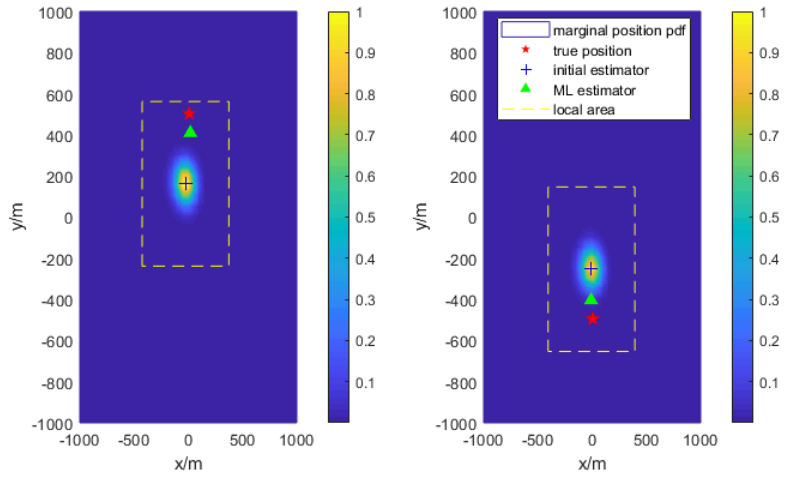

Figure 1: the marginal position pdfs ($\rho_1 = 500$) of two adjacent emitters (the red solid pentacle) at $\boldsymbol{p}_1 = (0, 500)$m and $\boldsymbol{p}_2 = (0, -500)$m respectively, the main simulation conditions are SNR=15dB, the number of sections $L = 10$, the number of elements $M = 3$, the bandwidth $B_w = 20$kHz and the range $\Delta_x = \Delta_y = 400$m (the yellow dotted rectangle).

where ρ_0, introduced by the Pincus' theorem (15), is selected as a sufficiently high value in practice, which is able to make $\phi(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)})$ too large to be stored in memory. On the other hand, the scale factor A, which has absorbed the coefficient R^{-1} in (52), involves the multidimensional integration. To greatly reduce the computational load and the required memory size without changing the final results, we can use the following *normalized* IW,

$$\bar{\eta}(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) = \frac{e^{\phi(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) - \max_r \phi(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)})}}{\sum_{r=1}^{R} e^{\phi(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) - \max_r \phi(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)})}} \tag{55}$$

The final IS based ML estimator is, therefore, given by

$$\hat{\boldsymbol{p}}_q = \sum_{r=1}^{R} \bar{\eta}(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) \boldsymbol{p}_q^{(r)}, \quad \hat{\mathring{t}}_q = \sum_{r=1}^{R} \bar{\eta}(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}) \mathring{t}_q^{(r)} \tag{56}$$

There is another detail about the super-parameters ρ_1 and ρ_2 to be stated. The distribution of probability mass in the local area $D_{\mathring{x}_q} \times D_{\mathring{y}_q}$ (or $D_{\mathring{t}_q}$) is governed by the super-parameter ρ_1 in the marginal position pdf $g_{\boldsymbol{p}}^{\text{mar}}(x_q, y_q)$ (or ρ_2 in the conditional transmitted time pdf $g_{\mathring{t}}^{\text{con}}(\mathring{t}_q | \boldsymbol{p}_q^{(r)})$) . Similarly to (18),

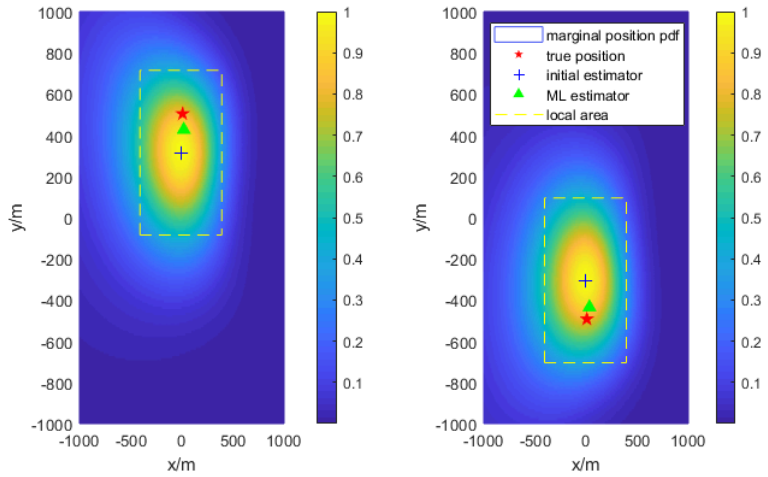

Figure 2: the marginal position pdfs with appropriate $\rho_1 = 20$. Other conditions are the same.

the probability mass of $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q)$ (or $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\boldsymbol{p}_q^{(r)})$) would concentrate at the location of the maximum peak $(\mathring{x}_q, \mathring{y}_q)$ (or $\mathring{t}_q$) as ρ_1 (or ρ_2) increases. The too large ρ_1 (or ρ_2) would result in that the points, included in local area but far away from the maximum peak point, are hardly to be sampled (as seen in Fig.1). The range parameter Δ_x, Δ_y (or $\Delta_{\mathring{t}}$) and ρ_1 (or ρ_2) should, therefore, be well matched, in order to generate the required realizations efficiently (as seen in Fig.2).

8. Complexity analysis

The main computation of IS-based ML DPD algorithm concentrates on the generation of the required realizations and the computation of the IWs. The generation process involves evaluating the marginal position pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q)$ and the conditional transmitted time pdf $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\boldsymbol{p}_q)$ at grids with different ranges and steps. The computational complexity, therefore, relies on the complexity of the two pdfs and the number of grid points. According to the definition of $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q)$ (A.8), its complexity is $\mathcal{O}(N^3)$ for the operation $\lambda_{\max}$

involves the eigen-decomposition. The marginal position pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q, y_q)$ is evaluated in (41), (47) and (50), at IJ, $I'J'$ and RJ' grid points respectively. The complexity of generating Q groups of $\{(\boldsymbol{p}_q^{(r)})\}_{r=1}^{R}, q = 1, ..., Q$ is, therefore, about $\mathcal{O}(Q(IJ + I'J' + RJ')N^3)$. Considering the (q, r)th matrix $\boldsymbol{\Pi}_q(\boldsymbol{p}_q^{(r)})$ in the marginal position pdf $g_{\boldsymbol{p}}^{\mathrm{mar}}(x_q^{(r)}, y_q^{(r)})$ is used in the conditional transmitted time pdf $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\boldsymbol{p}_q^{(r)})$ again. The residual complexity of $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\mathring{\boldsymbol{p}}_q^{(r)})$ is $\mathcal{O}(N^2)$. $g_{\mathring{t}}^{\mathrm{con}}(\mathring{t}_q|\mathring{\boldsymbol{p}}_q^{(r)})$ is evaluated in (43) and (51), at S and RS' grid points respectively. Therefore, the complexity of generating Q groups of $\{(\mathring{t}_q^{(r)})\}_{r=1}^{R}, q = 1, ..., Q$ is about $\mathcal{O}(Q(S + RS')N^2)$. According to the definition (25) and (34), the values of $\mathcal{G}(\boldsymbol{P}, \mathring{\boldsymbol{t}})$ at all R groups of realizations $\{\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}\}_{r=1}^{R}$ have already been computed during the generation of realizations. Thus, the computation of I-Ws only involves $\mathcal{F}(\boldsymbol{P}^{(r)}, \mathring{\boldsymbol{t}}^{(r)}), r = 1, ..., R$, of which the complexity is about $\mathcal{O}(RQ(LNMN_r + Q^2))$. In conclusion, the total complexity of IS-based ML D-PD is about $\mathcal{O}(Q(G_{\boldsymbol{p}}N^3 + (G_{\mathring{t}} + RN_r)N^2 + RQ^2))$ where $G_{\boldsymbol{p}} = IJ + I'J' + RJ'$ and $G_{\mathring{t}} = S + RS'$.

9. Numerical Simulation and Discussion

In this section, we design several numerical simulations to evaluate the performance of the proposed IS based ML DPD estimator, denoted by IS-ML-DPD, and compare it with the decoupled ML DPD estimator proposed in [1] and the MVDR based DPD estimator proposed in [3], denoted by single-ML-DPD and MVDR-DPD respectively. To give an explicit comparison with the exact ML estimator (14), we also give its performance curve that are obtained by an iterative local optimization method using the true emitters' positions as the initial point, denoted by optimal-ML-DPD.

The single-ML-DPD in [1] decouples the $Q(D+1)$-dimensional optimization problem into Q independent $(D+1)$-dimensional optimization problems by approximating the emitters' correlation matrix using a diagonal matrix (similar to the equation (29)) under the assumption $L \to \infty$. The off-diagonal elements of the normalized correlation matrix (26) tend to zero as the number of sec-

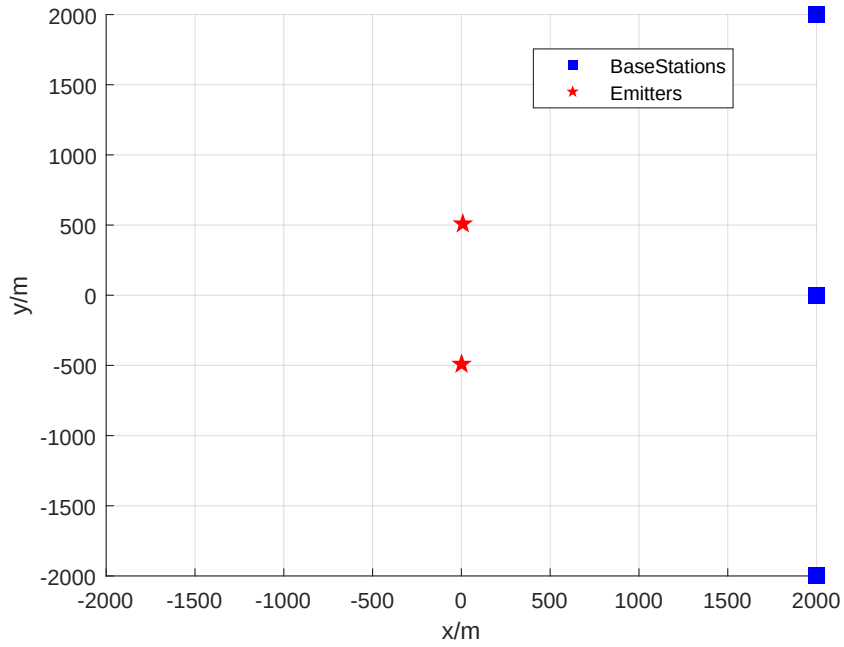

Figure 3: the simulation scenario

tions $L \to \infty$ since the transmitted signals among emitters are uncorrelated. The approximation in [1], consequently, may become inaccurate when the number of sections is limited in practice, which badly worsen the performance of single-ML-DPD estimator.

According to the multiplier effect, The off-diagonal elements (26) can also tend to zero as the DOA or TOA correlation coefficient $r_{doa,j}^{(u,v)}, r_{toa,j}^{(u,v)} \to 0$, even though the number of sections is limited in practice. It means that some estimators with high resolution, like the MVDR filter, may acquire better performance. The MVDR-DPD in [3] constructs the MVDR filter on the location domain directly. It is well known that the MVDR principle requires the Distortionless Response (DR) vector (3) keeps the same among data samples (i.e., the signal sections), which means that the prior waveform information $s_q(n,l)$ has to be eliminated from the DR vector (3). The MVDR-DPD estimator can, therefore, only use the DOA and TOA information to distinguish emitters.

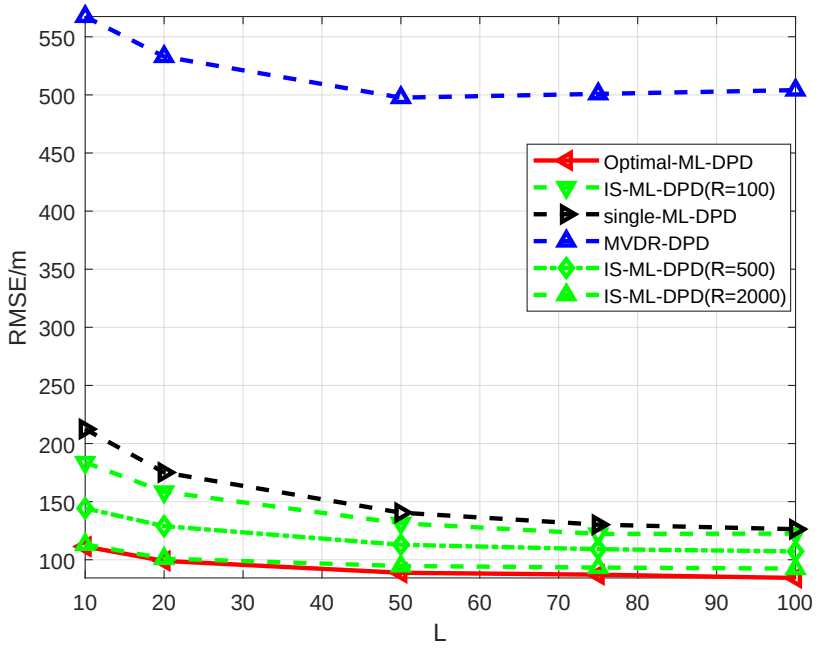

Figure 4: the RMSE of each estimator varies with the number of sections L. The signal bandwidth is 20kHz and the $SNR = 5$dB.

We use the Root Mean Square Error (RMSE) of the emitter's location estimation to evaluate the performance of each algorithm,

$$\text{RMSE} = \sqrt{\frac{\sum_{w=1}^{W} \|\boldsymbol{p}_1^{(w)} - \boldsymbol{p}_1\|^2}{W}} \tag{57}$$

where the W is the number of Monte Carlo simulations and we use $W = 300$ to obtain the statistic results. As illustrated in Fig.3, we consider the scenario where there are $N_r = 3$ Base Stations (BS) at $(2, 2)$km, $(2, 0)$km and $(2, -2)$km respectively and $Q = 2$ adjacent emitters at $\boldsymbol{p}_1 = (0, 500)$m and $\boldsymbol{p}_2 = (0, -500)$m. Each BS is equipped with an Uniform Linear Array (ULA) of $M = 3$ elements. The signals transmitted by the two emitters are two independent Gaussian random process. The Fourier coefficients of the transmitted signals are, consequently, complex Gaussian RVs, $s_q(n, l) \sim \mathcal{N}(0, \frac{10^{(SNR/10)}}{\sigma^2})$. The signal bandwidth is 20kHz. The number of sections $L = 10$ and the num-

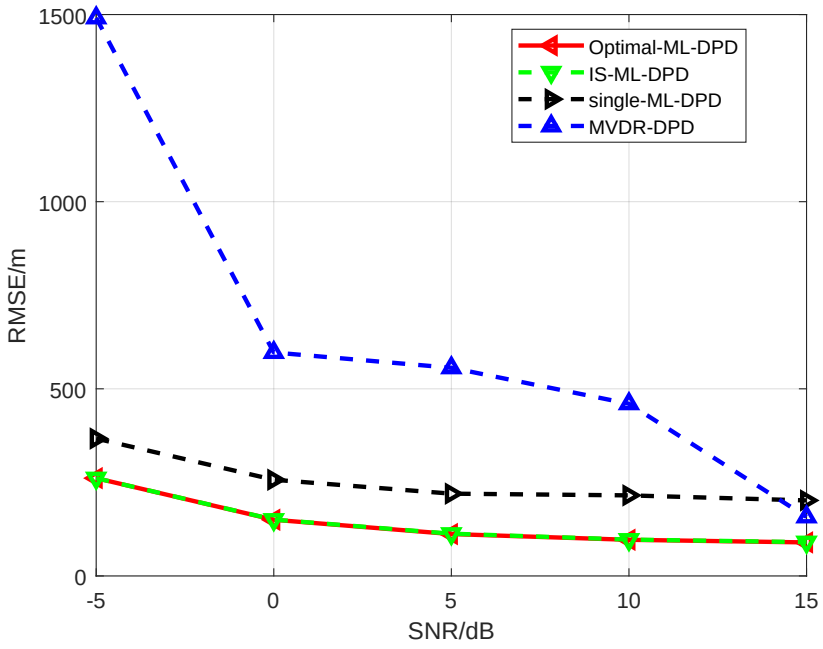

Figure 5: the RMSE of each estimator varies with the SNR. The signal bandwidth is 20kHz and the number of signal sections is $L = 10$

ber of samples in each section is $N = 5$. The transmitted time $\mathring{t}_q$ is selected from a Uniform Distribution $U[0, 100\mu s]$.

The super-parameter $\rho_0 = 666$ in IW expressions (53) is set as large as possible while $\rho_1 = 20$ and $\rho_2 = 8$ are designed appropriately to make the ML solutions $\hat{\boldsymbol{p}}_q$ and $\hat{\mathring{t}}_q$ included in the main lobe of $g^{\text{mar}}_{\boldsymbol{p}}(x_q, y_q)$ and $g^{\text{con}}_{\mathring{t}}(\mathring{t}_q|\mathring{\boldsymbol{p}}_q)$ respectively. Corresponding to the size of the main lobe, the range parameters are set as $\Delta_x = \Delta_y = 400\text{m}$ and $\Delta_{\mathring{t}} = 10\mu s$. These super-parameter values can be,statistically, acquired by plenty of simulations.

Note that the correlation coefficient r_{sig} in (26) would decrease until zero (the correlation information can be ignored) as the number of sections L increases sufficiently, for the independence among Q transmitted signals. Under the condition that L is sufficiently large, the single-ML-DPD estimator would, therefore, achieve the performance of the optimal-ML-DPD estimator approxi-

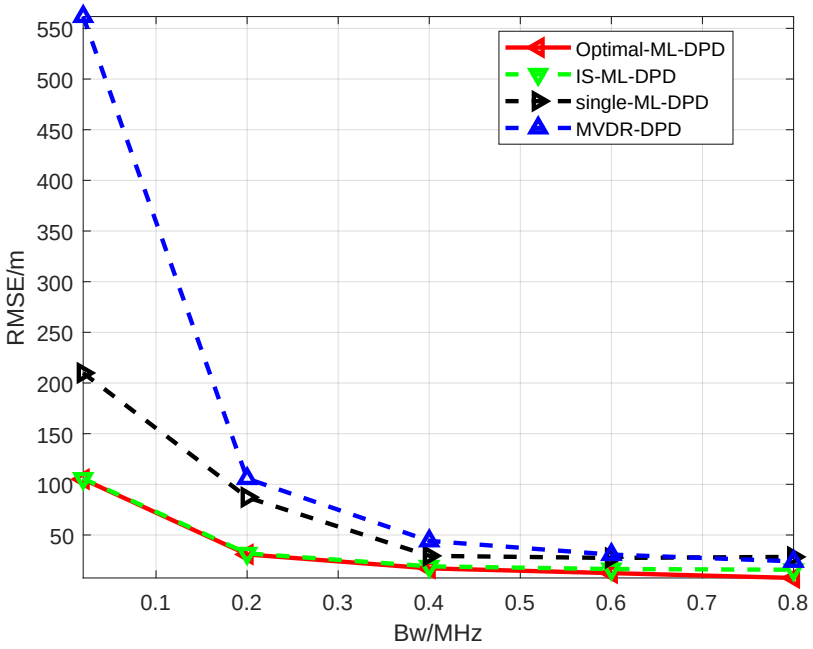

Figure 6: the RMSE of each estimator varies with the bandwidth.

mately. To indicate the IS-ML-DPD estimator's advantage when there are only a few signal sections, we start with varying the number of sections L between 10 and 100 in the first simulation. Fig.4 shows that all three IS-ML-DPD performance curves (corresponding to the number of realizations $R = 100, 500$ and 2000 respectively) locate in the gap between the performance curves of single-ML-DPD and optimal-ML-DPD. The localization accuracy of IS-ML-DPD improves as the number of realizations R increases and approximates the accuracy of optimal-ML-DPD at $R = 2000$. The reason is that enough realizations can guarantee that the realizations in vicinity of the ML solutions, with large IWs, can be sampled. It is worthy to note that $R = 2000$ realizations is just a few in the $Q(D+1) = 6$-dimensional parameter space. In another words, the complexity of brute searching for exact ML solution in $Q(D+1) = 6$-dimensional parameter space is about $\mathcal{O}(N_g^{Q(D+1)})$ (Ng is the number of grid points for each dimension), which is significantly larger than that of IS-ML-DPD. The MVDR-

DPD estimator's RMSE is around half of the distance between the two emitters at all L candidate values, which indicates that it fails to distinguish the two adjacent emitters at $SNR = 5dB, Bw = 20$kHz since the beforehand waveform information is not used at all.

Fig.5 shows that the RMSE of each algorithm varies with SNR in the second simulation. The SNR varies between -5dB and 15dB. The IS-ML-DPD (the number of realizations $R = 2000$) also approximates the performance of the optimal-ML-DPD at all SNR. The reason why there is a performance gap between single-ML-DPD and the IS-ML-DPD (or the optimal-ML-DPD) is that there exists the correlation among emitters for the limited observation resources ($B_w = 20$kHz,$L = 10$), which results in the interferences with one another. In such case, assuming the independence among emitters to decouple the multidimensional optimization problem is sub-optimal. On the condition of limited observation resources, the MVDR-DPD estimator is also unable to distinguish the two adjacent emitters until $SNR = 15$dB. The MVDR-DPD shows the higher resolution than single-ML-DPD and approximates the accuracy of IS-ML-DPD at relatively high SNR.

The first two simulations indicate that the single-ML-DPD and MVDR-DPD are sub-optimal while the IS-ML-DPD can approximate the performance of the optimal-ML-DPD in the case of limited observation resources. To further validate this result, we gradually increase the signal bandwidth B_w from 20kHz to 800kHz in the third simulation. As expected, the IS-ML-DPD outperforms the single-ML-DPD and MVDR-DPD, approximating the performance of the optimal-ML-DPD, when $B_w \leq 400$kHz and all estimators' performances converge together as the B_w increases.

10. Conclusion

In this paper, we research for direct ML position determination of dense emitters' positions using finite observation resources. We resort to a multidimensional optimization framework, consisting of the Pincus' theorem [6] and

the IS technique, to transform the multidimensional ML problem into several single ones using finite observation resources, Instead of assuming infinite observation resources. We design an appropriate IF to efficiently generate the required realizations which are used to approximate the exact multidimensional ML solution. The numerical simulations show that the proposed IS based ML DPD approach acquires the exact multidimensional ML solution approximately using finite observation resources (low SNR, a few signal sections or small signal bandwidth). Compared with the exponential complexity of the brute multidimensional searching, the complexity of the new approach is acceptable. due to space limitation, the work to reduce the computation of pdfs at grid points, based on the two-dimensional Fast Fourier Transformation (2D-FFT), will be presented soon.

Appendix A. The derivation of SLF

Our derivation starts with a vector definition as follows,

$$\boldsymbol{\gamma}_{q,j,l} \triangleq \boldsymbol{\Phi}_{q,j,l}(\boldsymbol{p}_q)\boldsymbol{b}(\mathring{t}_q) \tag{A.1}$$

where

$$\boldsymbol{\Phi}_{q,j,l}(\boldsymbol{p}_q) \triangleq \boldsymbol{\Lambda}_j(\boldsymbol{p}_q, s_q(l)) \otimes \boldsymbol{a}_j(\boldsymbol{p}_q) \tag{A.2}$$

$$\boldsymbol{b}(\mathring{t}_q) \triangleq [e^{-j2\pi f_1 \mathring{t}_q}, ..., e^{-j2\pi f_N \mathring{t}_q}]^T \tag{A.3}$$

$$\boldsymbol{\Lambda}_j(\boldsymbol{p}_q, s_q(l)) \triangleq \begin{bmatrix} s_q(1,l)e^{-j2\pi f_1 \tau_j(\boldsymbol{p}_q)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_q(N,l)e^{-j2\pi f_N \tau_j(\boldsymbol{p}_q)} \end{bmatrix} \tag{A.4}$$

substituting (A.1) and (A.2) into (32), we get

$$\tilde{\ell}_c(\boldsymbol{p}_q, \mathring{t}_q) = \sum_{j=1}^{N_r}\sum_{l=1}^{L} \boldsymbol{\gamma}_{q,j,l}^H \boldsymbol{y}_j(l)\boldsymbol{y}_j^H(l)\boldsymbol{\gamma}_{q,j,l} \tag{A.5}$$

$$= \boldsymbol{b}^H(\mathring{t}_q) \underbrace{\left(\sum_{j=1}^{N_r}\sum_{l=1}^{L} \boldsymbol{\Phi}_{q,j,l}^H(\boldsymbol{p}_q)\boldsymbol{y}_j(l)\boldsymbol{y}_j^H(l)\boldsymbol{\Phi}_{q,j,l}(\boldsymbol{p}_q)\right)}_{\triangleq \boldsymbol{\Pi}_q(\boldsymbol{p}_q)} \boldsymbol{b}(\mathring{t}_q) \tag{A.6}$$

The right hand of the second equation is a non-normalized Rayleigh Quotient with respect to $\mathring{t}_q$. Consequently, maximizing $\tilde{\ell}_c(\boldsymbol{p}_q, \mathring{t}_q)$ with respect to $\mathring{t}_q$, we get,

$$\ell_q(\boldsymbol{p}_q) \triangleq \lambda_{\max}(\boldsymbol{\Pi}_q(\boldsymbol{p}_q)) \tag{A.7}$$

where the notation $\lambda_{\max}(\boldsymbol{X})$ denotes the maximum eigenvalue of the matrix $\boldsymbol{X}$. Substituting (A.7) into (35), we get the marginal pdf of the qth emitter's position,

$$g_{\boldsymbol{p}}^{\mathrm{mar}}(\boldsymbol{p}_q) = \frac{e^{\rho_1 \lambda_{\max}(\boldsymbol{\Pi}_q(\boldsymbol{p}_q))}}{\int e^{\rho_1 \lambda_{\max}(\boldsymbol{\Pi}_q(\boldsymbol{p}_q))} d\boldsymbol{p}_q} \tag{A.8}$$

References

[1] A. Weiss, A.J., Direct position determination of multiple radio signals, EURASIP J. Adv. Signal Process 1 (2005) 37–49. doi:10.1155/ASP.2005.37.

[2] A. J. Weiss, Direct position determination of narrowband radio transmitters, 2004, pp. 513–516. doi:10.1109/LSP.2004.826501.

[3] T. Tirer, A. J. Weiss, High resolution direct position determination of radio frequency sources, IEEE Signal Processing Letters 23 (2) (2015) 192–196. doi:10.1109/LSP.2015.2503921.

[4] S. Kay, S. Saha, Mean likelihood frequency estimation, Signal Processing IEEE Transactions on 48 (7) (2000) 1937–1946. doi:10.1109/78.847780.

[5] A. Amar, A. J. Weiss, Efficient position determination of multiple emitterssdoi:10.1109/EEEI.2006.321149.

[6] M. Pincus, A closed form solution of certain programming problems, Operations Research 16 (3) (1968) 690–694. doi:10.1287/opre.16.3.690.

[7] G. H. Golub, V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, Siam Journal on Numerical Analysis 10 (2) (1973) 413–432. doi:10.1137/0710036.

[8] S. S. . Wang H, Kay S, An importance sampling maximum likelihood direction of arrival estimator, IEEE Transactions on Signal Processing 56 (10) (2008) 5082–5092. doi:10.1109/TSP.2008.928504.

[9] S. Saha, S. M. Kay, Maximum likelihood parameter estimation of superimposed chirps using monte carlo importance sampling, IEEE Transactions on Signal Processing 50 (2) (2002) 224–230. doi:10.1109/78.978378.

[10] H. Wang, S. Kay, Maximum likelihood angle-doppler estimator using importance sampling, IEEE Transactions on Aerospace and Electronic Systems 46 (2) (2010) p.610–622. doi:10.1109/TAES.2010.5461644.

[11] J. Chen, Y. C. Wu, S. C. Chan, T. S. Ng, Joint maximum-likelihood cfo and channel estimation for ofdma uplink using importance sampling, IEEE Transactions on Vehicular Technology 57 (6) (2008) 3462–3470. doi:10.1109/TVT.2008.920473.

[12] S. M. Kay, Intuitive Probability and Random Processes Using MATLAB, Springer Science,, 2006. doi:10.1007/0-387-24158-2_20.