

# Elsevier L<sup>A</sup>T<sub>E</sub>X template<sup>☆</sup>

Elsevier<sup>1</sup>

*Radarweg 29, Amsterdam*

*Elsevier Inc<sup>a,b</sup>, Global Customer Service<sup>b,\*</sup>*

*<sup>a</sup>1600 John F Kennedy Boulevard, Philadelphia*

*<sup>b</sup>360 Park Avenue South, New York*

---

## Abstract

*Keywords:*

---

## 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 location of the sources, like the angle of arrival (AOA), time difference of arrival (TDOA), Doppler frequency shift (DFS) or received signal strength (RSS), are estimated from the measurements of different receivers. 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 ignore the constraint that the measurements from different receivers have to correspond to the same source position. The Direct Position Determination (DPD) approaches have recently been proposed as a single-step localization

---

<sup>☆</sup>Fully documented templates are available in the elsarticle package on CTAN.

<sup>\*</sup>Corresponding author

*Email address:* [support@elsevier.com](mailto:support@elsevier.com) (Global Customer Service)

*URL:* [www.elsevier.com](http://www.elsevier.com) (Elsevier Inc)

<sup>1</sup>Since 1880.

technique where the source locations 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 Signal to Noise Ratio (SNR) and the two methods' performance converge together as  
20 SNR increases. Furthermore, an additional data association step (to partition the intermediate parameters corresponding to the same source into one set) is required in multiple sources scenario, which is avoided inherently in DPD methods.

In multiple sources scenario, the exact Maximum Likelihood (ML) estimator  
25 can be derived but it requires a multi-dimensional search which is usually impractical. To sidestep the multi-dimensional search 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 (M-  
30 F) [2] and the Minimum Variance Distortionless Response (MVDR) filter [3]. These techniques transform the multiple sources localization problem into single source localization problem which requires only two- or three-dimensional space search. These methods, however, require enough observation samples to achieve the asymptotically optimal performance at high SNR. To be specific,  
35 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  
40 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 [4] is proposed to achieve the exact ML solution efficiently. As in many iterative algorithms, the convergence to the global minimum of the cost  
45 function, however, depends on the initial estimate.

In this paper, we resort to the optimization framework [5], consisting of the

Pincus' theorem [6] and the Importance Sampling (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  
50 generating the alternative parameter realizations (i.e., sampling) 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  
55 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  $\mathbf{p}_q \in \mathbb{R}^{D \times 1}, q = 1, \dots, Q$  (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  $j$ th receiver are given by

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

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

the total observation time  $T$ . Each section can be Discrete Fourier Transformed (DFT) into

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

$$j = 1, \dots, N_r; n = 1, \dots, N; l = 1, \dots, L$$

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

For compact representation, we define the following vectors:

$$\gamma_{j,q,n,l}(\mathbf{p}_q, \hat{t}_q | s_q(n, l)) \triangleq \mathbf{a}_j(\mathbf{p}_q) e^{-j2\pi f_n(\tau_j(\mathbf{p}_q) + \hat{t}_q)} s_q(n, l) \quad (3)$$

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  $\hat{t}_q$ , which is assumed known in some DPD-related works [1, 4], is assumed unknown here (for the precise clock-synchronization between emitter and receiver is a challenging task). We observe that all information about the emitter's position is embedded in the manifold vector  $\gamma_{j,q,n,l}$  (for clarity, the dependence on  $\mathbf{p}_q$ ,  $\hat{t}_q$  and  $s_q(n, l)$  is omitted). The harmonic-superposition form of (2) is given by:

$$\mathbf{y}_j(n, l) = \sum_{q=1}^Q \alpha_{q,j} \gamma_{j,q,n,l} + \mathbf{w}_j(n, l) \quad (4)$$

which can be written in compact matrix form as follows,

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

where

$$\begin{aligned} \mathbf{\Gamma}_{j,n,l}(\mathbf{P}, \hat{\mathbf{t}}) &\triangleq [\gamma_{j,1,n,l}, \dots, \gamma_{j,Q,n,l}] \in \mathbb{C}^{M \times Q} \\ \boldsymbol{\alpha}_j &\triangleq [\alpha_{1,j}, \dots, \alpha_{Q,j}]^T \in \mathbb{C}^{Q \times 1} \end{aligned} \quad (6)$$

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

Collecting and stacking all the  $N$  coefficient vectors of the  $l$ th section of the  $j$ th received signal,

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

where

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

The received data model (7) is exactly the *nonlinear regression* model described in [5], which is suitable to the IS based global optimization framework.

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

### 3. the Concentrated Likelihood Function

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

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

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

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

The CLF (9) can be rewritten as the compact one as follows,

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

where  $\mathbf{y}_j \triangleq [\mathbf{y}_j^T(1), \dots, \mathbf{y}_j^T(L)]^T$ . It is extremely challenging to maximize (11) with respect to all parameters jointly. Fortunately, for any given  $\mathbf{P}, \mathring{\mathbf{t}}$ , 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

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

where  $\mathbf{\Gamma}_j^\dagger$  is the Moore-Penrose pseudo-inverse of  $\mathbf{\Gamma}_j$  given by  $(\mathbf{\Gamma}_j^H \mathbf{\Gamma}_j)^{-1} \mathbf{\Gamma}_j^H$ .

70 Note that  $(\mathbf{\Gamma}_j^H \mathbf{\Gamma}_j)^{-1}$  always exists since  $\mathbf{\Gamma}_j$  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,

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

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

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

The reduced-dimension optimization problem (14) still has  $Q(D+1)$  unknown parameters (including  $Q$   $D$ -dimensional location vectors  $\{\mathbf{p}_q\}_{q=1}^Q$  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  
75 space, which takes exponential complexity.

To solve the multidimensional ML problem (14) efficiently, the approach proposed in [1] is to decouple the problem into several  $D$ -dimensional optimization problems, under the assumption that the number of sections  $L \rightarrow \infty$  or  $L$  is sufficiently large in practice. Considering the length of section has low  
80 limit, i.e.,  $T/L \gg \max \tau_j(\mathbf{p}_q)$ , the number of sections  $L$  is 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 decoupling approach proposed in [1] may, therefore, be failure for the assumption is difficult to be  
 85 satisfied 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 ([8, 5, 9, 10, 11]), we resort to the optimization framework, which consists of the Pincus' theorem  
 90 and the powerful IS method, to achieve the ML solution of the adjacent emitters' positions efficiently and accurately, 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  $F(x_1, \dots, x_n) = F(\mathbf{x})$  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  $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_n]^T$  of  $S$ . Then for  $i = 1, \dots, n$ ,

$$\hat{x}_i = \lim_{\rho \rightarrow \infty} \frac{\int \dots \int x_i e^{\rho F(\mathbf{x})} dx_1 \dots dx_n}{\int \dots \int e^{\rho F(\mathbf{x})} dx_1 \dots dx_n} \quad (15)$$

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

$$\hat{\mathbf{p}}_q(d) = \int \dots \int \mathbf{p}_q(d) \mathcal{F}(\mathbf{P}, \mathbf{t}) d\mathbf{p}_1(1) \dots dt_Q \quad (16)$$

$$\hat{t}_q = \int \dots \int t_q \mathcal{F}(\mathbf{P}, \mathbf{t}) d\mathbf{p}_1(1) \dots dt_Q \quad (17)$$

$$q = 1, \dots, Q; d = 1, \dots, D$$

where

$$\mathcal{F}(\mathbf{P}, \mathbf{t}) \triangleq \lim_{\rho \rightarrow \infty} \frac{e^{\rho \mathcal{L}_c(\mathbf{P}, \mathbf{t})}}{\int \dots \int e^{\rho \mathcal{L}_c(\mathbf{P}, \mathbf{t})} d\mathbf{p}_1(1) \dots dt_Q} \quad (18)$$

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

Nevertheless, the multidimensional grid search 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, fortunately, turns out that  $\mathcal{F}(\mathbf{P}, \mathring{\mathbf{t}})$  can be seen as a Probability Density Function (pdf) since it is nonnegative and integrates to one. As a pdf,  $\mathcal{F}(\mathbf{P}, \mathring{\mathbf{t}})$  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,

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

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

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

Clearly, as the number of realizations  $R$  increases, the two sample mean in (20) approach the statistical expectations in (19). Unfortunately, the extremely non-linear property of the pdf  $\mathcal{F}(\mathbf{P}, \mathring{\mathbf{t}})$  makes it impractical to generate  $\{\mathbf{P}^{(r)}\}_{r=1}^R$  and  $\{\mathring{\mathbf{t}}^{(r)}\}_{r=1}^R$  using itself. To sidestep this problem, we can resort to the IS method [8, 5] and rewrite (16) and (17) in the following equivalent forms,

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

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



where  $\mathcal{G}(\mathbf{P}, \mathbf{t})$  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.,

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

where  $\eta(\mathbf{P}, \mathbf{t})$  is called *importance weight* (IW) corresponding to the realizations of  $\mathbf{P}, \mathbf{t}$  and defined as,

$$\eta(\mathbf{P}, \mathbf{t}) \triangleq \frac{\mathcal{F}(\mathbf{P}, \mathbf{t})}{\mathcal{G}(\mathbf{P}, \mathbf{t})} \quad (24)$$

If  $\mathcal{G}(\mathbf{P}, \mathbf{t})$  is carefully designed, the realizations  $\{\mathbf{P}^{(r)}\}_{r=1}^R$  and  $\{\mathbf{t}^{(r)}\}_{r=1}^R$ , which correspond to large IW  $\eta(\mathbf{P}, \mathbf{t})$ , can be generated from the IF  $\mathcal{G}(\mathbf{P}, \mathbf{t})$  efficiently. the expectations in (23) can be approximated at any desired degree of accuracy (by increasing  $R$ ) using the corresponding sample mean estimates.

105 In fact, the choice of  $\mathcal{G}(\mathbf{P}, \mathbf{t})$  is a tradeoff between the ease and efficiency of 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  $\mathcal{G}(\mathbf{P}, \mathbf{t})$  must  
110 be designed as close as possible to  $\mathcal{F}(\mathbf{P}, \mathbf{t})$  while it should also take the ease of generating the realizations into account.

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

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

According to the general results from probability theory, the independence among  $Q$  groups of RV pairs  $\{(\mathbf{p}_q, \mathbf{t}_q)\}_{q=1}^Q$  is introduced by the assumption for the factorized form in (25). Hence, we can independently generate  $Q$  groups of

115 realizations for  $(D + 1)$ -dimensional random vectors  $\{\mathbf{p}_q^{(r)}, \mathbf{t}_q^{(r)}\}_{r=1}^R, q = 1, \dots, Q$  using  $g_q(\mathbf{p}_q, \mathbf{t}_q), q = 1, \dots, Q$  respectively, instead of using  $\mathcal{G}(\mathbf{P}, \mathbf{t})$  to generate realizations for  $Q(D + 1)$ -dimensional random vector  $\{\mathbf{P}^{(r)}, \mathbf{t}^{(r)}\}_{r=1}^R$  with difficulty.

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

$$[\frac{1}{MNL} \mathbf{\Gamma}_j^H \mathbf{\Gamma}_j]_{u,v} \quad (26)$$

$$= \frac{1}{MNL} \sum_{l=1}^L \sum_{n=1}^N \gamma_{j,u,n,l}^H \gamma_{j,v,n,l} \quad (27)$$

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

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  $j$ th AOA parameter, the  $j$ th TOA parameter and the  $n$ th Fourier coefficient of the transmitted signal, respectively. The  $Q \times Q$  matrix  $\frac{1}{MNL} \mathbf{\Gamma}_j^H \mathbf{\Gamma}_j$  is, therefore, a normalized correlation matrix among all emitters. In order to diagonalize the normalized correlation matrix  $\frac{1}{MNL} \mathbf{\Gamma}_j^H \mathbf{\Gamma}_j$ , we set its off-diagonal elements to zeros at the cost of losing the correlation information. As a consequence, the expected diagonal matrix, for  $j = 1, \dots, N_r$ , is given by,

$$\mathbf{\Gamma}_j^H \mathbf{\Gamma}_j = MNL \cdot \mathbf{I} \quad (29)$$

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

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  
125 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,

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

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

where

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

Replacing  $\mathcal{L}_c(\mathbf{P}, \mathring{\mathbf{t}})$  and  $\rho$  in (18) with  $\tilde{\mathcal{L}}_c(\mathbf{P}, \mathring{\mathbf{t}})$  and an appropriate  $\rho_2$  respectively, we get the factorized form of  $\mathcal{G}(\mathbf{P}, \mathring{\mathbf{t}})$ , of which the  $q$ th factor is given by,

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

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

$$g_q(\mathbf{p}_q, \mathring{t}_q) = g_{\mathbf{p}}^{\text{mar}}(\mathbf{p}_q) g_{\mathring{t}}^{\text{con}}(\mathring{t}_q | \mathbf{p}_q) \quad (34)$$

The usual method to obtain the marginal pdf  $g_{\mathbf{p}}^{\text{mar}}(\mathbf{p}_q)$  is integrating the joint pdf  $g_q(\mathbf{p}_q, \mathring{t}_q)$  with respect to the transmitted time  $\mathring{t}_q$ . The analytical

integration is, however, difficult since  $g_q(\mathbf{p}_q, \mathring{t}_q)$  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  $g_{\mathbf{p}}^{\text{mar}}(\mathbf{p}_q)$  directly and efficiently,

$$g_{\mathbf{p}}^{\text{mar}}(\mathbf{p}_q) \triangleq \frac{e^{\rho_1 \ell_q(\mathbf{p}_q)}}{\int e^{\rho_1 \ell_q(\mathbf{p}_q)} d\mathbf{p}_q} \quad (35)$$

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

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

$$g_{\mathring{t}}^{\text{con}}(\mathring{t}_q | \mathbf{p}_q^{(r)}) = \frac{g_q(\mathbf{p}_q^{(r)}, \mathring{t}_q)}{\int g_q(\mathbf{p}_q^{(r)}, \mathring{t}_q) d\mathring{t}_q} \quad (36)$$

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

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

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

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

## 6. Generations Of The Required Realizations

140 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  $F_X^{-1}(\cdot) : [0, 1] \rightarrow \mathcal{X}, u \rightarrow x$  s.t.  $F_X(x) = u$ . Then, for any uniform RV,  $U \in [0, 1]$ , the RV  
145  $\tilde{X} = F_X^{-1}(U)$  is distributed according to  $f_X(\cdot)$ .

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

$$g_{\mathbf{p}}^{\text{mar}}(x_q, y_q) \triangleq g_x^{\text{mar}}(x_q)g_{y|x}^{\text{con}}(y_q|x_q) \quad (38)$$

where

$$g_x^{\text{mar}}(x_q) \triangleq \int g_{\mathbf{p}}^{\text{mar}}(x_q, y_q) dy_q \quad (39)$$

$$g_{y|x}^{\text{con}}(y_q|x_q) \triangleq \frac{g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)}{g_x^{\text{mar}}(x_q)} \quad (40)$$

In conclusion, the approach of generating the  $q$ th group of realizations  $\{(\mathbf{p}_q^{(r)}, \dot{t}_q^{(r)})\}_{r=1}^R$  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  $g_x^{\text{mar}}(x_q)$  in (39); For  $r = 1, \dots, R$ , then, generate two realizations  $u_y^{(r)}, u_{\dot{t}}^{(r)} \sim U[0, 1]$  and obtain  $y_q^{(r)} = G_{y_q|x_q}^{-1}(u_y^{(r)})$  and  $\dot{t}_q^{(r)} = G_{\dot{t}_q|\mathbf{p}_q^{(r)}}^{(-1)}(u_{\dot{t}}^{(r)})$  where  $G_{y_q|x_q}(\cdot)$  and  $G_{\dot{t}_q|\mathbf{p}_q}(\cdot)$  are the CDF associated to  $g_{y|x}^{\text{con}}(y_q|x_q)$  in (39) and  $g_{\dot{t}}^{\text{con}}(\dot{t}_q|\mathbf{p}_q^{(r)})$  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  $\dot{t} \in [0, \dot{t}_{\max}]$ , where  $\dot{t}_{\max}$  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  $q = 1, \dots, Q$ :

STEP 1: Evaluate the  $q$ th 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  $\Delta x$  and  $\Delta y$  (a rough search). Then, substitute SLF values into the

$q$ th marginal position pdf  $g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)$  in (35) and approximate integrals with discrete summations evaluate as follows,

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

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

$$[\hat{x}_q, \hat{y}_q] = \arg \max_{x^{(i)}, y^{(j)}} g_{\mathbf{p}}^{\text{mar}}(x^{(i)}, y^{(j)}) \quad (42)$$

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

$$g_t^{\text{con}}(\hat{t}_q^{(s)} | \hat{\mathbf{p}}_q) = \frac{g_q(\hat{\mathbf{p}}_q, \hat{t}_q^{(s)})}{\sum_{s=1}^S g_q(\hat{\mathbf{p}}_q, \hat{t}_q^{(s)}) \Delta \hat{t}} \quad (43)$$

160 STEP 4: Initialize the estimation of the  $q$ th transmitted time  $\hat{t}_q$  as the grid point corresponding to the maximum peak of the conditional transmitted time pdf  $g_t^{\text{con}}(\hat{t}_q^{(s)} | \hat{\mathbf{p}}_q)$ .

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

$$\hat{x}_q \in D_{\hat{x}_q} \triangleq [\hat{x}_q - \Delta_x, \hat{x}_q + \Delta_x] \quad (44)$$

$$\hat{y}_q \in D_{\hat{y}_q} \triangleq [\hat{y}_q - \Delta_y, \hat{y}_q + \Delta_y] \quad (45)$$

$$\hat{t}_q \in D_{\hat{t}_q} \triangleq [\hat{t}_q - \Delta_{\hat{t}}, \hat{t}_q + \Delta_{\hat{t}}] \quad (46)$$

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

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

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

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

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

$$G_{x_q}(x_q^{(r)}) = \sum_{i' \leq r} g_x^{\text{mar}}(x_q^{(i')}) \delta x, \quad \forall x_q^{(r)} \in D_{\hat{x}_q} \quad (49)$$

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

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

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

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

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  $r$ th coordinate  $y_q$  realization  $y_q^{(r)} = G_{y_q}^{-1}(u_{y_q}^{(r)})$  using linear interpolation as well.

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

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

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

All the  $Q$  groups of required realizations  $\{(\mathbf{p}_q^{(r)}, \hat{t}_q^{(r)})\}_{r, q}^R$ ,  $q = 1, \dots, Q$  can be generated according to the above steps. The  $q$ th expectations in (23), which is

equal to the ML solutions, can be approximated based on these realizations as follows,

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

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

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

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

where  $\rho_0$ , introduced by the Pincus' theorem (15), is selected as a sufficiently high value in practice, which is able to make  $\phi(\mathbf{P}^{(r)}, \mathring{\mathbf{t}}^{(r)})$  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,

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

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

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

There is another detail about the super-parameters  $\rho_1$  and  $\rho_2$  to be stated.

The distribution of probability mass in the local area  $D_{\hat{x}_q} \times D_{\hat{y}_q}$  (or  $D_{\hat{t}_q}$ ) is  
 175 governed by the super-parameter  $\rho_1$  in the marginal position pdf  $g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)$   
 (or  $\rho_2$  in the conditional transmitted time pdf  $g_{\mathbf{t}}^{\text{con}}(\mathring{t}_q | \mathbf{p}_q^{(r)})$ ). Similarly to (18),  
 the probability mass of  $g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)$  (or  $g_{\mathbf{t}}^{\text{con}}(\mathring{t}_q | \mathbf{p}_q^{(r)})$ ) would concentrate at the  
 location of the maximum peak  $(\hat{x}_q, \hat{y}_q)$  (or  $\hat{t}_q$ ) as  $\rho_1$  (or  $\rho_2$ ) increases. The too  
 large  $\rho_1$  (or  $\rho_2$ ) would result in that the points, included in local area but far  
 180 away from the maximum peak point, are hardly to be sampled (as seen in Fig.1).  
 The range parameter  $\Delta_x, \Delta_y$  (or  $\Delta_{\hat{t}}$ ) and  $\rho_1$  (or  $\rho_2$ ) should, therefore, be well



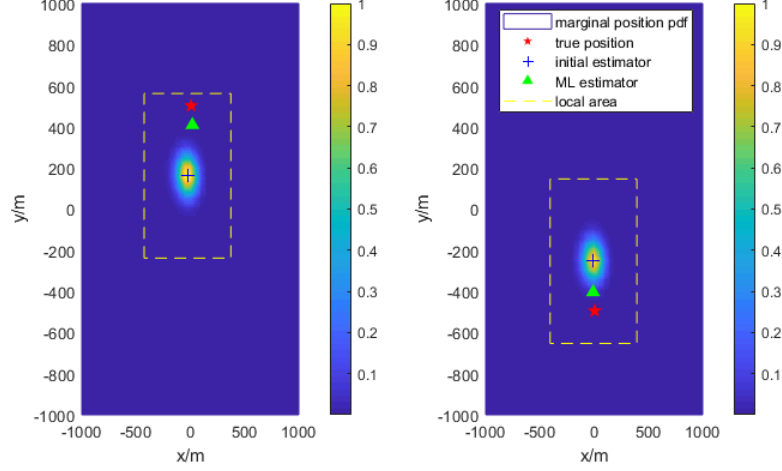


Figure 1: the marginal position pdfs ( $\rho_1 = 500$ ) of two adjacent emitters (the red solid pentacle) at  $\mathbf{p}_1 = (0, 500)\text{m}$  and  $\mathbf{p}_2 = (0, -500)\text{m}$  respectively, the main simulation conditions are SNR=15dB, the number of sections  $L = 10$ , the number of elements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

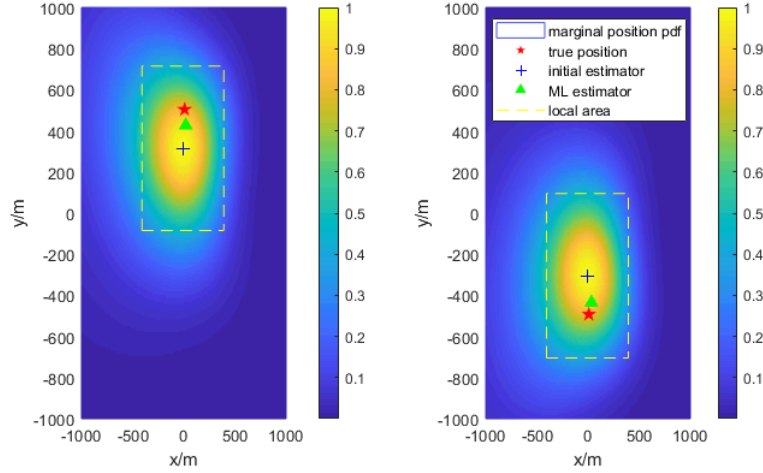


Figure 2: the marginal position pdfs with appropriate  $\rho_1 = 20$ . Other conditions are the same.

matched, in order to generate the required realizations efficiently (as seen in Fig.2).

## 8. Complexity analysis

185 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  $g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)$  and the conditional transmitted time pdf  $g_t^{\text{con}}(\hat{t}_q | \mathbf{p}_q)$  at grids with different ranges and steps. The computational complexity, therefore, relies on the

190 complexity of the two pdfs and the number of grid points. According to the definition of  $g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)$  (A.8), its complexity is  $\mathcal{O}(N^3)$  for the operation  $\lambda_{\max}$  involves the eigen-decomposition. The marginal position pdf  $g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)$  is evaluated in (41), (47) and (50), at  $IJ$ ,  $I'J'$  and  $RJ'$  grid points respectively. The complexity of generating  $Q$  groups of  $\{(\mathbf{p}_q^{(r)})\}_{r=1}^R, q = 1, \dots, Q$  is, therefore,

195 about  $\mathcal{O}(Q(IJ + I'J' + RJ')N^3)$ . Considering the  $(q, r)$ th matrix  $\mathbf{\Pi}_q(\mathbf{p}_q^{(r)})$  in the marginal position pdf  $g_{\mathbf{p}}^{\text{mar}}(x_q^{(r)}, y_q^{(r)})$  is used in the conditional transmitted time pdf  $g_t^{\text{con}}(\hat{t}_q | \mathbf{p}_q^{(r)})$  again. The residual complexity of  $g_t^{\text{con}}(\hat{t}_q | \mathbf{p}_q^{(r)})$  is  $\mathcal{O}(N^2)$ .  $g_t^{\text{con}}(\hat{t}_q | \hat{\mathbf{p}}_q^{(r)})$  is evaluated in (43) and (51), at  $S$  and  $RS'$  grid points respectively. Therefore, the complexity of generating  $Q$  groups of  $\{(\hat{t}_q^{(r)})\}_{r=1}^R, q = 1, \dots, Q$  is

200 about  $\mathcal{O}(Q(S + RS')N^2)$ . According to the definition (25) and (34), the values of  $\mathcal{G}(\mathbf{P}, \hat{\mathbf{t}})$  at all  $R$  groups of realizations  $\{\mathbf{P}^{(r)}, \hat{\mathbf{t}}^{(r)}\}_{r=1}^R$  have already been computed during the generation of realizations. Thus, the computation of IWs only involves  $\mathcal{F}(\mathbf{P}^{(r)}, \hat{\mathbf{t}}^{(r)}), r = 1, \dots, R$ , of which the complexity is about  $\mathcal{O}(RQ(LNMN_r + Q^2))$ . In conclusion, the total complexity of IS-based ML D-

205 PD is about  $\mathcal{O}(Q(G_{\mathbf{p}}N^3 + (G_{\hat{t}} + RN_r)N^2 + RQ^2))$  where  $G_{\mathbf{p}} = IJ + I'J' + RJ'$  and  $G_{\hat{t}} = S + RS'$ .

## 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,

210 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

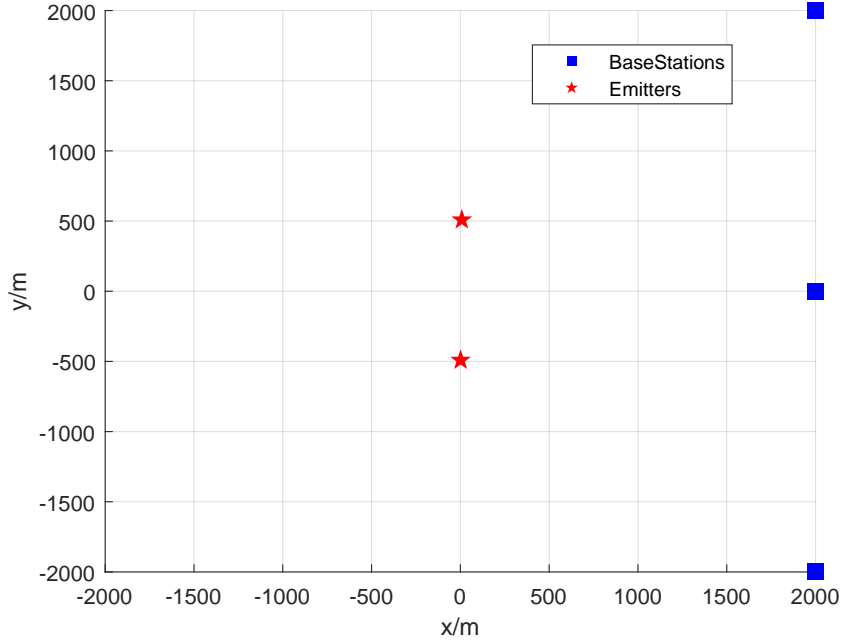


Figure 3: the simulation scenario

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  $L \rightarrow \infty$ . The off-diagonal elements of the normalized correlation matrix (26) tend to zero as the number of sections  $L \rightarrow \infty$  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  $r_{doa,j}^{(u,v)}, r_{toa,j}^{(u,v)} \rightarrow 0$ , even

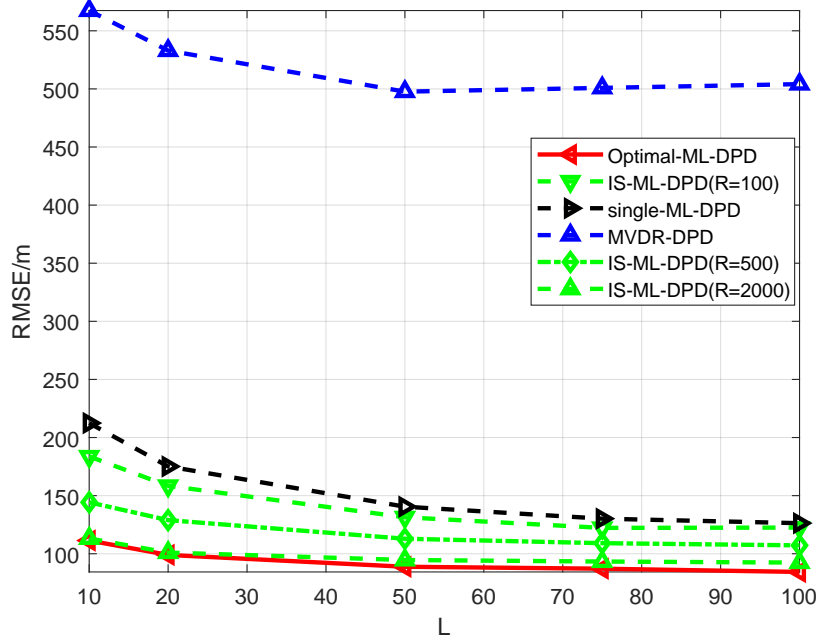


Figure 4: the RMSE of each estimator varies with the number of sections  $L$ . The signal bandwidth is 20kHz and the  $SNR = 5\text{dB}$ .

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.

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

$$\text{RMSE} = \sqrt{\frac{\sum_{s=1}^S \|\mathbf{p}_1^{(s)} - \mathbf{p}_1\|^2}{S}} \quad (57)$$

where the  $S$  is the number of Monte Carlo simulations and we use  $S = 300$  to

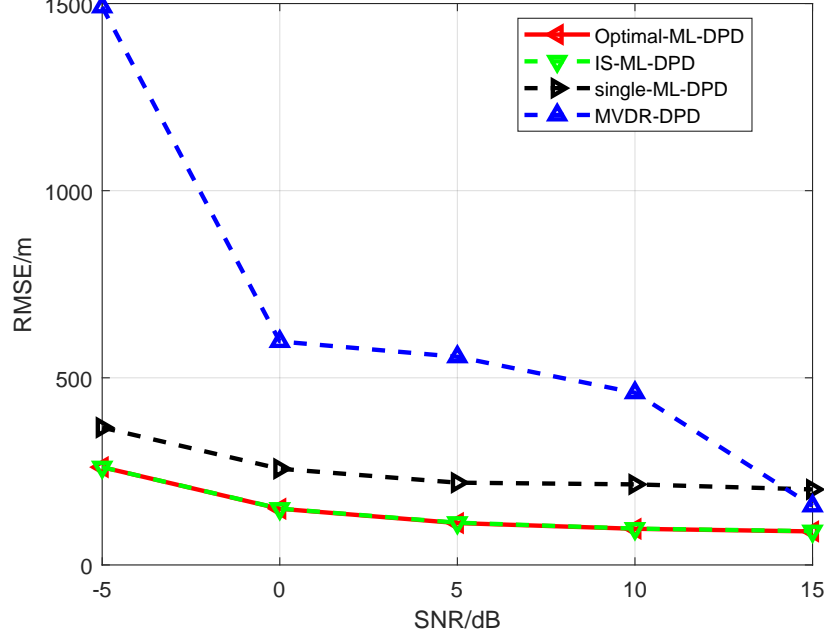


Figure 5: the RMSE of each estimator varies with the SNR. The signal bandwidth is 20kHz and the number of signal sections is  $L = 10$

obtain the statistic results. As illustrated in Fig.3, we consider the scenario in which There are three Base Stations (BS) at  $(2, 2)$ km,  $(2, 0)$ km and  $(2, -2)$ km and  $Q = 2$  adjacent (relative to system's resolution) emitters at  $\mathbf{p}_1 = (0, 500)$ m and  $\mathbf{p}_2 = (0, -500)$ m. Each BS is equipped with an Uniform Linear Array (U-  
240 LA) 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 number of samples in each section is  $N = 5$ . The transmitted time is selected from a  
245 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  $\hat{\mathbf{p}}_q$  and  $\hat{t}_q$  included in the main lobe of  $g_{\mathbf{p}}^{\text{mar}}(x_q, y_q)$  and  $g_t^{\text{con}}(\hat{t}_q | \mathbf{p}_q^{(r)})$

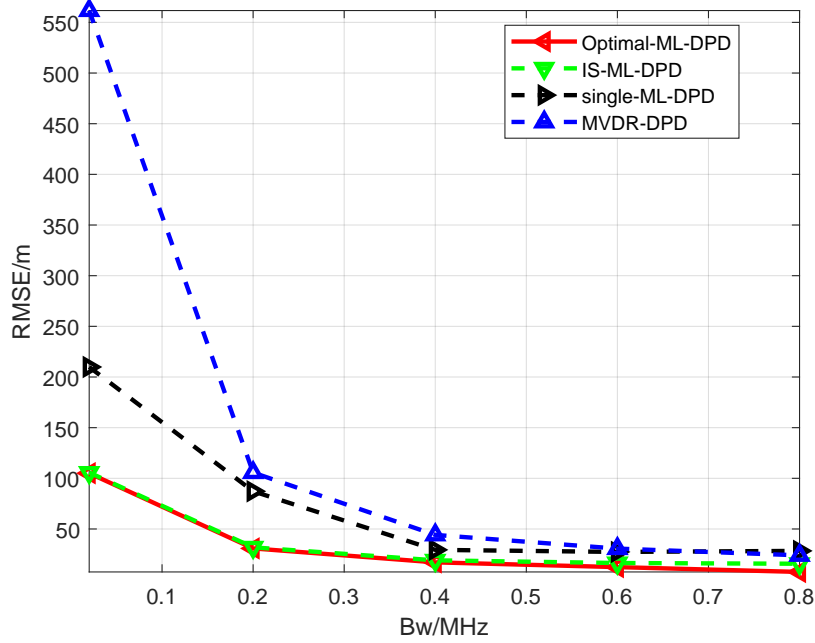


Figure 6: the RMSE of each estimator varies with the bandwidth.

respectively. To match the size of the main lobe, the range parameters are  
 250 set as  $\Delta_x = \Delta_y = 400\text{m}$  and  $\Delta_t = 10\mu\text{s}$ . These super-parameter values can  
 be, statistically, acquired by plenty of simulations.

Note that the transmitted signal correlation coefficient  $r_s$  in (26) would de-  
 crease until zero (the correlation information can be ignored) as the number  
 of sections  $L$  increases sufficiently, for the independence among  $Q$  transmitted  
 255 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 approximately. To indicate the IS-ML-DPD estimator's advantage  
 in the case of few signal sections, we start with varying the number of sec-  
 tions  $L$  between 10 and 100 in the first simulation. Fig.4 shows that all three  
 260 IS-ML-DPD performance curves (corresponding to the number of realization-  
 s  $R = 100, 500, 2000$  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  $R$  increase and achieves the accuracy of optimal-ML-DPD approximately at  $R = 2000$ . The reason is that  
265 generating sufficient realizations can guarantee that the realizations in vicinity of the ML solutions 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 in  $Q(D+1) = 6$ -dimensional parameter space is about  $\mathcal{O}(N_g^{Q(D+1)})$  ( $N_g$  is the number of grid points for each  
270 dimension), which is significantly larger than the complexity of IS-ML-DPD. The MVDR-DPD estimator's localization 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  $SNR = 5dB, B_w = 20kHz$  since the beforehand waveform information is not used sufficiently.

275 Fig.5 shows that the RMSE of each algorithm varies with SNR in the second simulation. The SNR varies between  $-5dB$  and  $15dB$ . It can be seen that the IS-ML-DPD (The number of realizations  $R = 2000$ ) also achieves, approximately, 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 is that  
280 there exist correlation among emitters for the limited observation resources ( $B_w = 20kHz, L = 10$ ), 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  
285 the two adjacent emitters until  $SNR = 15dB$ . The MVDR-DPD shows the higher resolution than single-ML-DPD and approximates the IS-ML-DPD at relatively high SNR.

The first two simulations indicates that the single-ML-DPD and MVDR-DPD are sub-optimal in the case of limited observation resources while the  
290 IS-ML-DPD is able to achieve the performance of the optimal-ML-DPD approximately in the same case. 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,

295 achieving the performance of the optimal-ML-DPD, at  $B_w \leq 400\text{kHz}$  and all estimators' performance converge as the  $B_w$  increases.

## Appendix A. The derivation of SLF

Our derivation starts with a vector definition as follows,

$$\gamma_{q,j,l} \triangleq \Phi_{q,j,l}(\mathbf{p}_q) \mathbf{b}(\hat{t}_q) \quad (\text{A.1})$$

where

$$\Phi_{q,j,l}(\mathbf{p}_q) \triangleq \Lambda_j(\mathbf{p}_q, s_q(l)) \otimes \mathbf{a}_j(\mathbf{p}_q) \quad (\text{A.2})$$

$$\mathbf{b}(\hat{t}_q) \triangleq [e^{-j2\pi f_1 \hat{t}_q}, \dots, e^{-j2\pi f_N \hat{t}_q}]^T \quad (\text{A.3})$$

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

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

$$\tilde{\ell}_c(\mathbf{p}_q, \hat{t}_q) = \sum_{j=1}^{N_r} \sum_{l=1}^L \gamma_{q,j,l}^H \mathbf{y}_j(l) \mathbf{y}_j^H(l) \gamma_{q,j,l} \quad (\text{A.5})$$

$$= \mathbf{b}^H(\hat{t}_q) \underbrace{\left( \sum_{j=1}^{N_r} \sum_{l=1}^L \Phi_{q,j,l}^H(\mathbf{p}_q) \mathbf{y}_j(l) \mathbf{y}_j^H(l) \Phi_{q,j,l}(\mathbf{p}_q) \right)}_{\triangleq \mathbf{\Pi}_q(\mathbf{p}_q)} \mathbf{b}(\hat{t}_q) \quad (\text{A.6})$$

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

$$\ell_q(\mathbf{p}_q) \triangleq \lambda_{\max}(\mathbf{\Pi}_q(\mathbf{p}_q)) \quad (\text{A.7})$$

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

$$g_{\mathbf{p}}^{\text{mar}}(\mathbf{p}_q) = \frac{e^{\rho_1 \lambda_{\max}(\mathbf{\Pi}_q(\mathbf{p}_q))}}{\int e^{\rho_1 \lambda_{\max}(\mathbf{\Pi}_q(\mathbf{p}_q))} d\mathbf{p}_q} \quad (\text{A.8})$$



## 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.  
300
- [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.  
305
- [4] A. Amar, A. J. Weiss, Efficient position determination of multiple emitter-  
sdoi:10.1109/EEEI.2006.321149.
- [5] S. Kay, S. Saha, Mean likelihood frequency estimation, Signal Processing IEEE Transactions on 48 (7) (2000) 1937–1946. doi:10.1109/78.847780.
- [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.  
310
- [7] G. H. Golub, V. Pereyra, The differentiation of pseudo-inverses and non-linear 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.  
315
- [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.  
320
- [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.

- 325 [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.