Joint Data-Aided Carrier Frequency Offset, Phase Offset, Amplitude and SNR Estimation for Millimeter-Wave MIMO Systems

Javier Rodríguez-Fernández†, Nuria González-Prelcic†, and Robert W. Heath Jr.‡

† Universidade de Vigo, Email: {jrodriguez,nuria}@gts.uvigo.es
‡ The University of Texas at Austin, Email: {kiranj,rheath}@utexas.edu

Abstract

This work is devoted to solve the problem of estimating the carrier frequency offset, phase offset, amplitude, and SNR between two mmWave transceivers.

The Cramér-Rao Lower Bound (CRLB) for the different parameters is provided first, as well as the condition for the CRLB to exist, known as Regularity Condition. Thereafter, the problem of finding suitable estimators for the parameters is addressed, for which the proposed solution is the Maximum Likelihood estimator (ML).

Notation: Here we describe the notation to be used throughout this document: bold uppercase $\mathbf{A}$ is used to denote matrices, bold lowercase $\mathbf{a}$ denotes a column vector and non-bold lowercase $a$ denotes a scalar value. We use $\mathcal{A}$ to denote a set. Further, $\| \mathbf{A} \|_F$ is the Frobenius norm and $\mathbf{A}^*$, $\mathbf{A}^C$, $\mathbf{A}^T$ and $\mathbf{A}^\dagger$ denote the conjugate transpose, conjugate, transpose and Moore-Penrose pseudoinverse of a matrix $\mathbf{A}$, respectively. The $(i,j)$-th entry of a matrix $\mathbf{A}$ is denoted using $[\mathbf{A}]_{i,j}$. Similarly, the $i$-th entry of a column vector $\mathbf{a}$ is denoted as $[\mathbf{a}]_i$. The identity matrix of order $N$ is denoted as $\mathbf{I}_N$. If $\mathbf{A}$ and $\mathbf{B}$ are two matrices, $\mathbf{A} \circ \mathbf{B}$ is the Khatri-Rao product of $\mathbf{A}$ and $\mathbf{B}$ and $\mathbf{A} \otimes \mathbf{B}$ is their Kronecker product. We use $\mathcal{N}(\mu, \mathbf{C})$ to denote a circularly symmetric complex Gaussian random vector with mean $\mu$ and covariance matrix $\mathbf{C}$. We use $\mathbb{E}$ to denote expectation. Discrete-time signals are represented as $x[n]$. 

arXiv:1712.05989v1 [cs.IT] 16 Dec 2017
I. SYSTEM MODEL

We consider a mmWave MIMO link to send $N_s$ data streams using a transmitter with $N_t$ antennas and a receiver with $N_r$ antennas. Both the transmitter and the receiver use a fully-connected hybrid MIMO architecture as shown in Fig. 1 with $L_t$ and $L_t$ RF chains. A hybrid precoder is used, with $F = F_{RF}F_{BB} \in \mathbb{C}^{N_t \times N_s}$, where $F_{RF} \in \mathbb{C}^{N_t \times L_t}$ is the analog precoder and $F_{BB} \in \mathbb{C}^{L_t \times N_s}$ the digital one. The RF precoder and combiner are implemented using a fully connected network of phase shifters, as described in [1].

The MIMO channel between the transmitter and the receiver is modeled as a $N_r \times N_t$ matrix denoted as $H$, which is assumed to be a sum of the contributions of $C$ spatial clusters, each contributing with $R_c$ rays. This matrix is given by [1]

$$H = \sqrt{\frac{N_rN_t}{\rho_L}} \sum_{c=1}^{C} \sum_{r=1}^{R_c} \alpha_{c,r} a_R(\phi_{c,r}) a_T^*(\theta_{c,r}),$$

(1)

where $\rho_L$ denotes the path loss, $C$ is the number of scattering clusters, $R_c$ is the number of rays for $c$-th cluster, $\alpha_{c,r} \in \mathbb{C}$ is the complex gain of the $r$-th ray within $c$-th cluster, $\phi_{c,r} \in [0, 2\pi)$ and $\theta_{c,r} \in [0, 2\pi)$ are the angles of arrival and departure (AoA/AoD), respectively of the $r$-th ray within $c$-th cluster, and $a_R(\phi_{c,r}) \in \mathbb{C}^{N_r \times 1}$ and $a_T(\theta_{c,r}) \in \mathbb{C}^{N_t \times 1}$ are the array steering vectors for the receive and transmit antennas respectively. This matrix can be written in a more compact way as

$$H = A_R G A_T^*,$$

(2)
where \( \mathbf{G} \in \mathbb{C}^{\sum_{c=1}^{C} R_c \times \sum_{c=1}^{C} R_c} \) is diagonal with non-zero complex entries, and \( \mathbf{A}_R \in \mathbb{C}^{\sum_{s=1}^{S} R_s \times \sum_{s=1}^{S} R_s} \) and \( \mathbf{A}_T \in \mathbb{C}^{\sum_{s=1}^{S} R_s \times \sum_{s=1}^{S} R_s} \) contain the receive and transmit array steering vectors \( \mathbf{a}_R(\phi_{c,r}) \) and \( \mathbf{a}_T(\theta_{c,r}) \), respectively. \( \mathbf{H}_d \) can be written as

\[
\mathbf{H} \approx \tilde{\mathbf{A}}_R \tilde{\mathbf{G}}^\dagger \tilde{\mathbf{A}}_T^* \tag{3}
\]

where \( \tilde{\mathbf{G}}^\dagger \in \mathbb{C}^{G_t \times G_t} \) is a sparse matrix which contains the path gains of the quantized spatial frequencies in the non zero elements. The dictionary matrices \( \tilde{\mathbf{A}}_T \) and \( \tilde{\mathbf{A}}_R \) contain the transmitter and receiver array response vectors evaluated on grids of sizes \( G_t \) and \( G_r \). Assuming that the receiver applies a hybrid combiner \( \mathbf{W} = \mathbf{W}_{RF} \mathbf{W}\_BB \in \mathbb{C}^{\sum_{s=1}^{S} R_s \times L_t} \), with \( \mathbf{W}_{RF} \in \mathbb{C}^{\sum_{s=1}^{S} R_s \times L_t} \) the analog combiner, and \( \mathbf{W}\_BB \in \mathbb{C}^{L_t \times \sum_{s=1}^{S} R_s} \) the baseband combiner, the received signal at discrete time instant \( n \) can be written as

\[
\mathbf{r}[n] = \mathbf{W}^* \mathbf{H} \mathbf{f}[n] e^{j2\pi \Delta f n} + \mathbf{v}[n], \tag{4}
\]

for \( n = 0, \ldots, N - 1 \). The signal \( \mathbf{s}[n] \in \mathbb{C}^{L_t \times 1} \), \( 0 \leq n \leq N - 1 \) is a training sequence known to the receiver, \( \Delta f \) is the unknown carrier frequency offset (CFO), and \( \mathbf{v}[n] \sim \mathcal{N}(0, \sigma^2 \mathbf{W}^* \mathbf{W}) \) is the circularly symmetric complex Gaussian distributed additive noise vector.

We consider that the training sequence can be expressed as \( \mathbf{s}[n] = \mathbf{q}[n] \), with \( \mathbf{q} \in \mathbb{C}^{L_t \times 1} \) a spatial filter consisting of normalized QPSK symbols. Let us define \( \mathbf{\alpha} \triangleq \mathbf{D}_w^{-1} \mathbf{W}^* \mathbf{H} \mathbf{f}, \mathbf{\alpha} = \begin{bmatrix} \alpha_1 e^{j \beta_1} & \alpha_2 e^{j \beta_2} & \ldots & \alpha_{L_t} e^{j \beta_{L_t}} \end{bmatrix}^T \) and \( \mathbf{\Omega}_n = e^{j2\pi \Delta f n} \mathbf{I}_{L_t} \), where \( \mathbf{D}_w \in \mathbb{C}^{L_t \times L_t} \) is the Cholesky factor of \( \mathbf{C}_w = \mathbf{W}^* \mathbf{W} \), i.e., \( \mathbf{C}_w = \mathbf{D}_w^* \mathbf{D}_w \). Then, if we stack the \( N \) samples of the received signal in (4) we obtain the signal model

\[
\begin{bmatrix} \mathbf{r}[0] \\
\mathbf{r}[1] \\
\vdots \\
\mathbf{r}[N-1] \end{bmatrix} = \Omega_0 \mathbf{\alpha} = \begin{bmatrix} \Omega_0 \mathbf{\alpha} \\
\Omega_1 \mathbf{\alpha} \\
\vdots \\
\Omega_{N-1} \mathbf{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{t}[0] \\
\mathbf{t}[1] \\
\vdots \\
\mathbf{t}[N-1] \end{bmatrix} + \begin{bmatrix} \mathbf{v}[0] \\
\mathbf{v}[1] \\
\vdots \\
\mathbf{v}[N-1] \end{bmatrix}, \tag{5}
\]

such that \( \mathbf{r} \sim \mathcal{N}(\mathbf{M}(\mathbf{\alpha}, \Delta f) \mathbf{t}, \sigma^2 \mathbf{I}_{L_t}) \). Then, the vector of parameters to be estimated is \( \mathbf{\xi} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_{L_t} & \beta_1 & \ldots & \beta_{L_t} & \Delta f & \sigma^2 \end{bmatrix}^T \).

We define the Signal to Noise Ratio (SNR) for the \( i \)-th signal \( \mathbf{r}[n] \), \( i = 1, \ldots, L_t \) as

\[
\gamma_i = \frac{\alpha_i^2}{\sigma^2}, \tag{6}
\]
Furthermore, the average Signal to Noise Ratio at digital level can be defined as well. Let us define $\Lambda \in \mathbb{C}^{L_r \times L_r}$ as $\Lambda = \text{diag}\{\alpha_1 e^{j\beta_1}, \alpha_2 e^{j\beta_2}, \ldots, \alpha_{L_r} e^{j\beta_{L_r}}\}$. Therefore, if the SNR at baseband level is denoted as $\text{SNR}$, it can be written as

$$\text{SNR} = \frac{1}{L_r} \sum_{i=1}^{L_r} \gamma_i = \frac{\text{trace}\{\Lambda \Lambda^*\}}{L_r \sigma^2},$$

which is just the average of the post-combining $\gamma_i$ at each RF chain.

## II. Regularity Condition

For the CRLB to exist, the regularity condition must be fulfilled by the probability density function (pdf) of the data. This condition states \cite{2}

$$\mathbb{E}\left\{ \frac{\partial \ln p(r; \xi)}{\partial \xi_i} \right\} = 0, \text{ for all } \xi_i \in \xi.$$

The pdf of the vector $r$ is written as

$$p(r; \xi) = \frac{1}{\pi^{N L_r} \det(\sigma^2 I_{N L_r})} e^{-\frac{1}{\sigma^2} (r-M(\alpha, \Delta f)t)^*(r-M(\alpha, \Delta f)t)},$$

and the log-likelihood function (LLF) as

$$\ln p(r; \xi) = -N L_r \ln \pi \sigma^2 - \frac{1}{\sigma^2} \left( r^* r - 2 \text{Re}\{t^* M^*(\alpha, \Delta f) r\} + t^* M^*(\alpha, \Delta f) M(\alpha, \Delta f) t \right).$$

Before computing the gradient of the LLF, it is important to note that

$$t^* M(\alpha, \Delta f) M(\alpha, \Delta f) t = \sum_{n=0}^{N-1} s^C[n] \alpha^* \Omega^*_n \Omega_n \alpha s[n]$$

$$= \sum_{n=0}^{N-1} s^C[n] \text{trace}\{P\} s[n]$$

$$= N \text{trace}\{P\},$$

where $P = \Lambda \Lambda^*$.

Therefore, for $\alpha_i$. 
\[ \frac{\partial \ln p(\mathbf{r}; \xi)}{\partial \alpha_i} = -\frac{1}{\sigma^2} \left( -2 \text{Re} \left\{ \mathbf{t}^* \frac{\partial \mathbf{M}^*(\mathbf{\alpha}, \Delta f)}{\partial \alpha_i} \mathbf{r} \right\} + NL_i \frac{\partial \text{trace}\{\mathbf{P}\}}{\partial \alpha_i} \right) \]

\[ = -\frac{1}{\sigma^2} \left( -2 \sum_{n=0}^{N-1} \text{Re} \left\{ s^*[n] e^{-j\beta_i} \mathbf{e}_i^T \Omega_n^* \mathbf{r}[n] \right\} + 2\alpha_i N \right) \]

\[ = -\frac{1}{\sigma^2} \left( -2 \sum_{n=0}^{N-1} \text{Re} \left\{ s^*[n] e^{-j\beta_i} \mathbf{e}_i^T \Omega_n^* \mathbf{\alpha}_s[n] \right\} + 2\alpha_i N \right) \]

\[ = \mathbb{E} \left\{ \frac{\partial \ln p(\mathbf{r}; \xi)}{\partial \alpha_i} \right\} = -\frac{1}{\sigma^2} \left( -2 \sum_{n=0}^{N-1} \text{Re} \{ \alpha_i \} + 2\alpha_i N \right) = 0. \]  

(12)

Now, for \( \sigma^2 \),

\[ \frac{\partial \ln p(\mathbf{r}; \xi)}{\partial \sigma^2} = -\frac{NL_r}{\sigma^2} + \frac{1}{\sigma^4} \| \mathbf{r} - \mathbf{M}(\mathbf{\alpha}, \Delta f) \mathbf{t} \|^2 \]

\[ = -\frac{NL_r}{\sigma^2} + \frac{1}{\sigma^4} \sum_{n=0}^{N-1} \mathbb{E} \{ \| \mathbf{w}_n[n] \|^2 \} \]

\[ = -\frac{NL_r}{\sigma^2} + \frac{1}{\sigma^4} \sum_{n=0}^{N-1} L_r \underbrace{\mathbb{E} \{ \| w_{l,n}[n] \|^2 \}}_{\sigma^2} + \mathbb{E} \{ \| w_{q,n}[n] \|^2 \} = 0. \]  

(13)

For \( \Delta f \),

\[ \frac{\partial \ln p(\mathbf{r}; \xi)}{\partial \Delta f} = -\frac{1}{\sigma^2} \left( -2 \text{Re} \left\{ \mathbf{t}^* \frac{\partial \mathbf{M}^*(\mathbf{\alpha}, \Delta f)}{\partial \Delta f} \mathbf{r} \right\} \right) \]

\[ = -\frac{1}{\sigma^2} \left( -2 \sum_{n=0}^{N-1} \text{Re} \left\{ s^*[n] \mathbf{\alpha}^* \Omega_n^* \Omega_n \mathbf{\alpha}_s[n] \right\} \right) \]

\[ = -\frac{1}{\sigma^2} \left( -2 \sum_{n=0}^{N-1} \text{Re} \left\{ \text{trace}\{\mathbf{P}\}(\Omega_n^* \mathbf{\alpha}_s[n]) \right\} \right) = 0 \]  

(14)

Finally, for \( \beta_i \),

\[ \frac{\partial \ln p(\mathbf{r}; \xi)}{\partial \beta_i} = -\frac{1}{\sigma^2} \left( -2 \text{Re} \left\{ \mathbf{t}^* \frac{\partial \mathbf{M}^*(\mathbf{\alpha}, \Delta f)}{\partial \beta_i} \mathbf{r} \right\} \right) \]

\[ = -\frac{1}{\sigma^2} \left( -2 \sum_{n=0}^{N-1} \text{Re} \left\{ s^*[n] \mathbf{\alpha}_i(-j) e^{-j\beta_i} \mathbf{e}_i^T \Omega_n^* \Omega_n \mathbf{\alpha}_s[n] \right\} \right) \]

\[ = -\frac{1}{\sigma^2} \left( -2 \sum_{n=0}^{N-1} \text{Re}\{ \alpha_i^2(-j) \} \right) = 0. \]  

(15)
since the term inside brackets is also purely imaginary.
Thus, since the regularity condition holds, the CRLB is to be found in the next section.

III. CRAMÉR-RAO LOWER BOUND

Since the model for the received signal is Gaussian, the Slepian-Bangs formula can be used to find the elements in the Fisher Information Matrix. This formula is given by [2]

$$F_{i,j}(\xi) = 2 \text{Re} \left\{ \frac{\partial \mu^*(\xi) C^{-1}(\xi) \frac{\partial \mu(\xi)}{\partial \xi_j}}{\partial \xi_i} \right\} + \text{trace} \left\{ C^{-1}(\xi) \frac{\partial C(\xi)}{\partial \xi_i} C^{-1}(\xi) \frac{\partial C(\xi)}{\partial \xi_j} \right\}. \quad (16)$$

Thus, the elements in the FIM can be found as

$$F_{\alpha_i,\alpha_i}(\xi) = 2 \frac{\sigma^2}{\sigma^2} \text{Re} \left\{ t^* \frac{\partial M^*(\alpha, \Delta f)}{\partial \alpha_i} \frac{\partial M(\alpha, \Delta f)}{\partial \alpha_i} t \right\}. \quad (17)$$

The partial derivative of $M(\alpha, \Delta f)$ is given by

$$\frac{\partial M^*(\alpha, \Delta f)}{\partial \alpha_i} \frac{\partial M(\alpha, \Delta f)}{\partial \alpha_i} = \begin{bmatrix} e_i^T e_i \\ \vdots \\ e_i^T e_i \end{bmatrix} = \alpha_i I_N, \quad (18)$$

being $e_i \in \mathbb{R}^{L_r}$ a vector of zeros with a single one in the $i$-th position. Therefore,

$$F_{\alpha_i,\alpha_i}(\xi) = \frac{2}{\sigma^2} \sum_{n=0}^{N-1} \text{Re} \{ |s[n]|^2 \} = \frac{2N}{\sigma^2}. \quad (19)$$

For the phase offset parameter,

$$F_{\beta_i,\beta_i}(\xi) = \frac{2}{\sigma^2} \text{Re} \left\{ t^* \frac{\partial M^*(\alpha, \Delta f)}{\partial \beta_i} \frac{\partial M(\alpha, \Delta f)}{\partial \beta_i} t \right\}, \quad (20)$$

where

$$\frac{\partial M^*(\alpha, \Delta f)}{\partial \beta_i} \frac{\partial M(\alpha, \Delta f)}{\partial \beta_i} = \begin{bmatrix} \alpha_i^2 e_i^T e_i \\ \vdots \\ \alpha_i^2 e_i^T e_i \end{bmatrix} = \alpha_i I_N. \quad (21)$$
Then, the Fisher Information for $\beta_i$ yields

$$F_{\beta_i, \beta_i}(\xi) = \frac{2}{\sigma^2} \sum_{n=0}^{N-1} \text{Re}\{\alpha_i^2 |s[n]|^2\} = \frac{2N\alpha_i^2}{\sigma^2}. \quad (22)$$

For the carrier frequency offset parameter,

$$F_{\Delta f, \Delta f}(\xi) = \frac{2}{\sigma^2} \text{Re}\left\{t^* \frac{\partial M^*(\alpha, \Delta f)}{\partial \Delta f} \frac{\partial M(\alpha, \Delta f)}{\partial \Delta f} t\right\}, \quad (23)$$

where

$$\frac{\partial M^*(\alpha, \Delta f)}{\partial \Delta f} \frac{\partial M(\alpha, \Delta f)}{\partial \Delta f} = \text{trace}\{P\} \begin{bmatrix} (2\pi 0)^2 \\ \vdots \\ (2\pi (N-1))^2 \end{bmatrix}. \quad (24)$$

Therefore,

$$F_{\Delta f, \Delta f}(\xi) = \frac{2}{\sigma^2} \text{trace}\{P\} \sum_{n=0}^{N-1} (2\pi n)^2. \quad (25)$$

For the noise variance,

$$F_{\sigma^2, \sigma^2}(\xi) = \frac{1}{\sigma^4} \text{trace}\{I_{NL_r}\} = \frac{NL_r}{\sigma^4}. \quad (26)$$

For the non-diagonal elements in the FIM, it can be checked that all of them are zero-valued except for

$$F_{\Delta f, \beta_i}(\xi) = \frac{2}{\sigma^2} \text{Re}\left\{t^* \frac{\partial M^*(\alpha, \Delta f)}{\partial \Delta f} \frac{\partial M(\alpha, \Delta f)}{\partial \beta_i} t\right\}, \quad (27)$$

where

$$\frac{\partial M^*(\alpha, \Delta f)}{\partial \Delta f} \frac{\partial M(\alpha, \Delta f)}{\partial \beta_i} = \alpha_i e^{j\beta_i} \begin{bmatrix} (2\pi 0)\alpha^* e_i \\ \vdots \\ (2\pi (N-1))\alpha^* e_i \end{bmatrix}. \quad (28)$$
Therefore,

\[ F_{\Delta f, \beta_i}(\xi) = \frac{2}{\sigma^2} \sum_{n=0}^{N-1} \text{Re}\{(2\pi n)\alpha_i^2\} = \frac{2\alpha_i^2}{\sigma^2} \sum_{n=0}^{N-1} (2\pi n). \] (29)

Then, the FIM is found to be

\[ F(\xi) = \begin{bmatrix} F_1(\xi) & 0 \\ 0 & F_2(\xi) \end{bmatrix}, \] (30)

with

\[ F_1(\xi) = \text{diag}\left\{ \frac{2N}{\sigma^2} \mathbf{1}_L^T, \frac{L_r N}{\sigma^4} \right\} \] (31)

\[ F_2(\xi) = \begin{bmatrix} \frac{2 \text{trace}\{\mathbf{P}\}}{\sigma^2} \sum_{n=0}^{N-1} (2\pi n)^2 & \frac{2}{\sigma^2} \sum_{n=0}^{N-1} 2\pi n \mathbf{1}_L^T \mathbf{P} \\ \frac{2}{\sigma^2} \sum_{n=0}^{N-1} 2\pi n \mathbf{P} \mathbf{1}_L & \frac{2N}{\sigma^2} \mathbf{P} \end{bmatrix}. \] (32)

Finally, upon inverting the block diagonal matrix \( F(\xi) \), the CRLB for the parameters is given by

\[ \text{var}\{\hat{\alpha}_i\} \geq \frac{\sigma^2}{2N}, \] (33)

\[ \text{var}\{\hat{\sigma}^2\} \geq \frac{\sigma^4}{L_r N} \] (34)

\[ \text{var}\{\hat{\Delta f}\} \geq \frac{\sigma^2}{2} \cdot \frac{N_{L_r}}{(N(N-1)(2N-1)/6) \text{trace}\{\mathbf{P}\} N_{L_r} - \sum_{k=1}^{L_r} \alpha_k^2 (N(N+1)^2/4) N_{L_r}} \] (35)

\[ \text{var}\{\hat{\beta}_i\} \geq \frac{\sigma^2}{2} \cdot \frac{(N(N-1)(2N-1)/6) \text{trace}\{\mathbf{P}\} N_{L_r} \prod_{j \neq i} \alpha_j^2 - \sum_{k=1}^{L_r} \alpha_k^2 (N(N+1)^2/4) N_{L_r} \prod_{k \neq i} \alpha_k^2}{(N(N-1)(2N-1)/6) \text{trace}\{\mathbf{P}\} N_{L_r} - \sum_{k=1}^{L_r} \alpha_k^2 (N(N+1)^2/4) N_{L_r}} \] (36)

Finally, the SNR in the \( i \)-th RF chain can be distinguished from the average of the individual SNRs of the different signals involved in the problem. If the SNR of the \( i \)-th signal is denoted by \( \gamma_i \), then

\[ \gamma_i = \frac{\alpha_i^2}{\sigma^2}, \] (37)

\[ \text{SNR} = \frac{\text{trace}\{\mathbf{P}\}}{L_r \sigma^2}, \] (38)
for which, according to the Transformed Parameters Property, the CRLB can be found to be

\[
\text{var}\{\hat{\gamma}_i\} \geq \frac{4\gamma_i}{N} + \frac{\gamma_i^2}{N} \quad (39)
\]

\[
\text{var}\{\hat{\text{SNR}}\} \geq \frac{2\text{SNR}}{L_i N} + \frac{\text{SNR}^2}{L_i N} \quad (40)
\]

The last formulas provide a clear insight on how the estimation of \(\gamma_i\) actually works. For lower values of \(\gamma_i\), the CRLB is dominated by the term growing linearly with the parameter, whereas for higher values the CRLB is dominated by the second term, making the estimation of the parameter much harder. The same happens for \(\hat{\text{SNR}}\), although it actually depends on the different \(\gamma_i\) for the different streams.

Henceforth, our effort is focused on searching for suitable estimators of these parameters. Owing to the non-linear dependence between the data and the parameters, an efficient estimator cannot be found, in general. The practical approach to follow is to seek the ML estimators for these parameters, for which the next subsection is devoted.

**IV. Maximum Likelihood Estimation**

The problem of finding the ML Estimator for the parameters in \(\xi\) can be formalized as

\[
\left\{\hat{\alpha}_{i,\text{ML}}, \hat{\sigma}_{\text{ML}}^2, \hat{\Delta f}_{\text{ML}}, \hat{\beta}_{i,\text{ML}}\right\} = \arg \max_{\xi} \ln p(r; \xi), \quad (41)
\]

which involves a joint maximization over \(2(L_r + 1)\) scalar variables. This problem can be solved in four different steps by splitting the original problem into four interconnected optimization problems.

**A. MLE for the Phase Offset**

Recall that the LLF is given by

\[
\ln p(r; \xi) = -NL_r \ln \pi \sigma^2 - \frac{1}{\sigma^2} \left( r^* r - 2 \text{Re}\{t^* M^*(\alpha, \Delta f) r\} + N \text{trace}\{P\} \right), \quad (42)
\]

such that the term that depends on \(\beta_i\) is the second variable inside brackets. Therefore, the value \(\hat{\beta}_i\) that maximizes the function above is found from

\[
\hat{\beta}_i = \arg \max_{\beta_i} \text{Re}\{t^* M^*(\alpha, \Delta f) r\}. \quad (43)
\]
The first derivative of the function to maximize yields

\[
\frac{\partial \ln p(r; \xi)}{\partial \beta} = \sum_{n=0}^{N-1} \text{Re}\{s^C[n]\alpha_i(-j)e^{-j\beta_i}\Omega_n r[n]\}\]

(44)

\[
\frac{\partial \ln p(r; \xi)}{\partial \beta} = \sum_{n=0}^{N-1} \text{Im}\{s^C[n]\alpha_i e^{-j\beta_i} e^{-j2\pi \Delta f n} r_i[n]\}\]

(45)

Now the following result can be applied.

**Lemma.** Given two complex numbers, \(z_1, z_2\), the imaginary part of their product is defined as

\[
\text{Im}\{z_1 z_2\} = \text{Re}\{z_1\} \text{Im}\{z_2\} + \text{Im}\{z_1\} \text{Re}\{z_2\}.
\]

Therefore, setting the previous derivative to zero allows us to obtain

\[
\sum_{n=0}^{N-1} \text{Im}\{s^C[n]r_i[n]e^{-j2\pi \Delta f n}\} \text{Re}\{\alpha_i e^{-j\beta_i}\} = \sum_{n=0}^{N-1} \text{Re}\{s^C[n]r_i[n]e^{-j2\pi \Delta f n}\} \text{Im}\{\alpha_i e^{-j\beta_i}\}
\]

(46)

\[
\hat{\beta}_i = \tan^{-1}\left\{\frac{\sum_{n=0}^{N-1} \text{Im}\{s^C[n]r_i[n]e^{-j2\pi \Delta f n}\}}{\sum_{n=0}^{N-1} \text{Re}\{s^C[n]r_i[n]e^{-j2\pi \Delta f n}\}}\right\}
\]

(47)

which can be interpreted as a matched-filtering operation with the training sequence \(s[n]\) after the carrier frequency offset has been corrected. Notice that the ML estimator of \(\beta_i\) requires the knowledge of the true value \(\Delta f\). Since it is impossible to know the exact value of \(\Delta f\), the ML estimator of \(\Delta f\) is to be substituted in (47) instead, such that the final estimator can be applied in a practical scenario.

**B. MLE for the Amplitude**

In this subsection we provide a closed-form expression for the amplitude parameters, which have already been denoted as \(\alpha_i\). By inspection of the LLF

\[
\ln p(r; \xi) = -NLr \ln \pi \sigma^2 - \frac{1}{\sigma^2} \left(r^* r - 2 \text{Re}\{t^*M^*(\alpha, \Delta f) r\} + N \text{trace}\{P\}\right),
\]

(48)

it can be seen that the terms that depend on this parameter are the two last within brackets. Therefore, the problem of finding \(\alpha_i\) is formalized as

\[
\hat{\alpha}_i = \arg \max_{\alpha_i} \quad 2 \text{Re}\{t^*M^*(\alpha, \Delta f) r\} - N \text{trace}\{P\}.
\]

(49)

Taking the first derivative of the objective function yields

\[
\frac{\partial f(\alpha_i)}{\partial \alpha_i} = 2 \sum_{n=0}^{N-1} \text{Re}\{r_i[n]s^C[n]e^{-j\beta_i} e^{-j2\pi \Delta f n}\} - 2\alpha_i N \]

(50)
such that the value of $\alpha_i$ that maximizes the function above is

$$\hat{\alpha}_i = \frac{1}{N} \sum_{n=0}^{N-1} \text{Re}\{r_i[n]s[n]e^{-j\beta_i}e^{-j2\pi\Delta fn}\}. \quad (51)$$

From (51), it is clear that the statistic $\hat{\alpha}_i$ depends on $\Delta f$ and $\beta_i$, similarly as with the estimator of $\beta_i$. Therefore, the estimators of these parameters are to be substituted in (51) to find the ML estimator.

C. MLE for the Carrier Frequency Offset

In this subsection we will find the ML estimator of the carrier frequency offset. In (48), the only term that depends on $\Delta f$ is the second one. Therefore, the problem of finding the ML estimator for this parameter can be stated as

$$\hat{\Delta f} = \arg \max_{\Delta f} \text{Re}\{t^*M^*(\alpha, \Delta f)r\}. \quad (52)$$

We can express the objective function in (52) as

$$f(\Delta f) = \text{Re}\left\{\sum_{n=0}^{N-1} \sum_{i=1}^{L_r} \alpha_i r_i[n]s[n]e^{-j2\pi\Delta fn}\right\}, \quad (53)$$

in which the statistics $\hat{\alpha}_i$ and $\hat{\beta}_i$ are to be substituted. These are given by

$$\hat{\beta}_i = \angle\left\{\sum_{n=0}^{N-1} r_i[n]s[n]e^{-j2\pi\Delta fn}\right\} \quad (54)$$

$$\hat{\alpha}_i = \frac{1}{N} \left|\sum_{n=0}^{N-1} r_i[n]s[n]e^{-j2\pi\Delta fn}\right|, \quad (55)$$

where the statistic $\hat{\alpha}_i$ follows from substituting (54) into (51). Therefore, (54) and (55) can be substituted into (53) to yield the value of $\Delta f$ that maximizes the objective function. This can be written as

$$\hat{\Delta f} = \arg \max_{\Delta f} \left\{\frac{1}{N} \sum_{i=1}^{L_r} \left|\sum_{n=0}^{N-1} r_i[n]s[n]e^{-j2\pi\Delta fn}\right|^2\right\}. \quad (56)$$

Accordingly, the ML estimator for the carrier frequency offset can be understood as the bin corresponding to the maximum of the sum of the squared-periodograms of the received signals, after a temporal matched filter. Since the periodogram is defined as the Fourier Transform of the autocorrelation, it is necessary to find its maximum numerically, by using the Fast - Fourier Transform (FFT).
The FFT yields a sampled version of the DTFT, such that there is no guarantee that the carrier frequency offset falls within a specific integer bin of the FFT. Thus, the three largest points in the FFT can be found and parabolic interpolation can be performed thereafter. The motivation behind this approach is trying to find the real maximum of the function. The actual maximum is not generally found, but an accurate estimate can be found, instead.

D. Quadratic Interpolation of Spectral Peaks

This subsection is meant to explaining how parabolic interpolation can be applied to obtain a better estimate of the carrier frequency offset. In general, assume that the shape of the Energy Spectral Density (ESD) of the received stochastic process is the one pictured in Figure 2. Then, the equation of a parabola is given by

\[ y(x) \triangleq a(x-p)^2 + b, \quad \text{(57)} \]

where \( p \) is the interpolated peak location (in FFT bins). At the three nearest samples, the images of the parabola at these points are

\[ y(-1) = \alpha, \quad y(0) = \beta, \quad y(1) = \gamma. \quad \text{(58)} \]

Then, expressing the three samples in terms of the interpolating parabola, it is found that

\[ \alpha = ap^2 + 2ap + a + b \]
\[ \beta = ap^2 + b \quad \text{(59)} \]
\[ \gamma = ap^2 - 2ap + a + b. \]

Therefore, the interpolated peak location in bins is given by

\[ p = \frac{1}{2} \frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \in \left[ -\frac{1}{2}, \frac{1}{2} \right], \quad \text{(60)} \]

such that the amplitude at that bin is

\[ y(p) = \beta - \frac{1}{4}(\alpha - \gamma)p. \quad \text{(61)} \]

Finally, by using (60), an estimate of the interpolated peak location can be obtained. If \( k^* \) denotes the FFT bin that yields the maximum value, then \((k^* + p)\) yields the absolute position of the interpolated bin. Then, the analog frequency estimate is expressed as

\[ \tilde{\omega} = \frac{k^* + p}{N} f_s, \quad \text{(62)} \]

where \( f_s \) is the sampling frequency in Hertz and \( N \) is the number of points in the FFT.
Fig. 2: Illustration of parabolic peak interpolation using the three samples closest to the peak.

E. MLE for the Noise Variance

This subsection is devoted to find the MLE for the noise variance parameter, $\sigma^2$. Recalling the Log-Likelihood Function

$$
\ln p(r; \xi) = -NL_r \ln \pi \sigma^2 - \frac{1}{\sigma^2} \|r - M(\alpha, \Delta f)t\|^2_2,
$$

the MLE for the noise variance can be easily found from

$$
\frac{\partial \ln p(r; \xi)}{\partial \sigma^2} = -\frac{NL_r}{\sigma^2} + \frac{1}{\sigma^4} \|r - M(\alpha, \Delta f)t\|^2_2.
$$

The condition for the derivative to vanish is that

$$
\hat{\sigma}^2 = \frac{1}{NL_r} \|r - \hat{M}(\alpha, \Delta f)t\|^2_2,
$$

where $\hat{M}(\alpha, \Delta f)$ is the estimate of the matrix $M(\alpha, \Delta f)$ that models how the signal energy is transferred from the transmitter to the receiver. The estimate $\hat{M}(\alpha, \Delta f)$ is computed using the ML estimators of the previous parameters.
F. MLE for the Signal to Noise Ratio

This last subsection is devoted to provide the ML estimator for the SNR metric. To find this statistic, the following property of ML estimation can be applied [2].

**Lemma 1: Invariance Property of the MLE:** The MLE of the parameter $\alpha = g(\xi)$, where the pdf $p(x; \xi)$ is parameterized by $\xi$, is given by

$$\hat{\alpha} = g(\hat{\xi}),$$

where $\hat{\xi}$ is the MLE of $\xi$.

This nice property allows us to find the MLE for $\text{SNR} = \frac{\alpha_i^2}{\sigma^2}$, which is simply

$$\hat{\gamma}_i = \frac{\hat{\alpha}_i^2}{\hat{\sigma}^2},$$

where $\hat{\alpha}_i$ and $\hat{\sigma}^2$ are the MLEs for $\alpha_i$ and $\sigma^2$ provided in the previous subsections. As a summary, the ML estimators for the parameters are

$$\hat{\Delta f}_{\text{ML}} = \arg \max_{\Delta f} \left| \sum_{i=1}^{L_t} \sum_{n=0}^{N-1} s^C[n] r_i[n] e^{-j2\pi \Delta f n} \right|^2,$$

$$\hat{\beta}_{i,\text{ML}} = \angle \left\{ \sum_{n=0}^{N-1} s^C[n] r_i[n] e^{-j2\pi \Delta f n} \right\},$$

$$\hat{\alpha}_{i,\text{ML}} = \frac{1}{N} \left| \sum_{n=0}^{N-1} s^C[n] r_i[n] e^{-j2\pi \Delta f n} \right|,$$

$$\hat{\sigma}^2_{\text{ML}} = \frac{1}{L_t N} \left\| \mathbf{r} - \hat{\mathbf{M}}(\alpha, \Delta f) \mathbf{t} \right\|^2_2,$$

$$\hat{\gamma}_{i,\text{ML}} = \frac{\hat{\alpha}_i^2}{\hat{\sigma}^2},$$

$$\hat{\text{SNR}}_{\text{ML}} = \frac{\text{trace}\{\hat{\mathbf{P}}\}}{L_t \hat{\sigma}^2},$$

To assess the performance of the different estimators, some results are presented hereafter. In the simulations, a range from $-15$ dB to $10$ dB has been considered for SNR, which is a typical range in millimeter wave communications. The amplitude parameters are generated from a $\mathcal{U}[0, 1]$ distribution. The phase-offsets are generated from a $\mathcal{U}(0, 2\pi)$ and the carrier frequency offset is generated as $\mathcal{U}[-1/2, 1/2]$. The number of antennas is set to $N_t = 32$ and $N_r = 32$, and the number of RF chains employed at both transmitter and receiver are set to $L_t = 4$ and $L_r = 4$. The training sequence $t[n]$, $0 \leq n \leq N - 1$ contains normalized QPSK symbols, with
$N = 64$ time-domain samples. The results have been averaged over $N_{MC} = 1000$ Monte Carlo realizations.

The normalized sample bias of the different parameters is shown in Figure 4. As predicted by the Estimation Theory, at low SNR regime the sample bias of the parameters is non zero whereas it decreases to zero asymptotically with SNR. The estimators are, thereby, asymptotically unbiased.

![Normalized sample bias for the amplitude and SNR parameters.](image)

**Fig. 3**: Normalized sample bias for the amplitude and SNR parameters.

Now, the efficiency of the estimators is evaluated through the normalized sample variance. The normalized sample variances for the different estimators is shown in Figure 5, Figure 6, Figure 7 and Figure 8, along with their corresponding normalized CRLBs. It can be noticed that the normalized sample variance of the estimators does not lie within the NCRLB for low values of SNR, as it is expected. Nevertheless, for high values of SNR, the estimators present a normalized sample variance that lies within the NCRLB. Therefore, the estimators are asymptotically efficient.
Fig. 4: Normalized sample bias for the angular parameters $\beta_i$, $i = 1, \ldots, 4$ and $\Delta f$.

Fig. 5: Normalized sample variance and NCRLB for the amplitude parameters.
Fig. 6: Normalized sample variance and NCRLB for SNR.

Fig. 7: Normalized sampled variance and NCRLB for the phase offsets $\beta_i$, $i = 1, \ldots, 4$. 
Fig. 8: Normalized sample variance and NCRLB obtained for the carrier frequency offset parameter.

REFERENCES

[1] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath Jr., "Hybrid MIMO architectures for millimeter wave communications", in Proc. IEEE Globecom, 2014.

[2] S. M. Kay, Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory. Prentice Hall PTR, 1993.