Chapter 2

Spatially Distributed Source Localization

2.1 Sensor Array Processing

Sensor array processing is regarded as a common framework for taking into account several problems involved with the use of sensor array. The sensor elements are located at different points in space, to receive a lot of incoming signals. The problem of finding several arrival directions impinging on an array antenna or sensor array, called *direction finding*, is of interest long times ago since it is a useful parameter in several systems such as wireless commination, radar, navigation and *etc.* It therefore plays a prominent role in the phenomenons due to the propagation of plane waves transmitted through a media.

The fundamental of sensor array processing is to infer the information implicit in the incident signals by collecting multiple observations from a number of sensor elements. When modelling the array output as a parameterized quantity, it is plausible to retrieve the location information of the impinged waves. Literature review on the area can be found in [3], [4] and [5]. It is conceivable that most DOA (direction of arrival) estimators can be classified into four groups, such as, beamforming, subspace (also known as Eigenstructure), maximum likelihood,and combinatorial scheme. For more details on these, we refer to [5]. The intention presented in this section is to reveiew an alternative development to the classical (point source) model of sensor array processing as illustrated in Fig. 2.1.

In some radio propagation scenarios, especially in urban areas when there is no line of sight between the source and the receiving antenna array, the array receives many rays from the vicinity of source and therefore the assumption on number of rays becomes too restrictive. This constitutes the term *local scattering* which is often used to describe a wireless channel in the form of a physical phenomenon. In the presence of local scattering around the vicinity of source, the classical point source model will suffer from the lack of identifiability accounted for large number of directions. To deviate from the problem given, it is preferred to assume that a number of multipaths is large enough so that their path gains can be characterized under the central limit theorem. As governed by a Gaussian random variable and associated random directions, the channel response beholds stochastic process. It seemed in general possible to govern deviated angles into a parametric model as well, whenever a priori knowledge of angle deviation is available. Note that such a model development rather coincides with the following quotation.

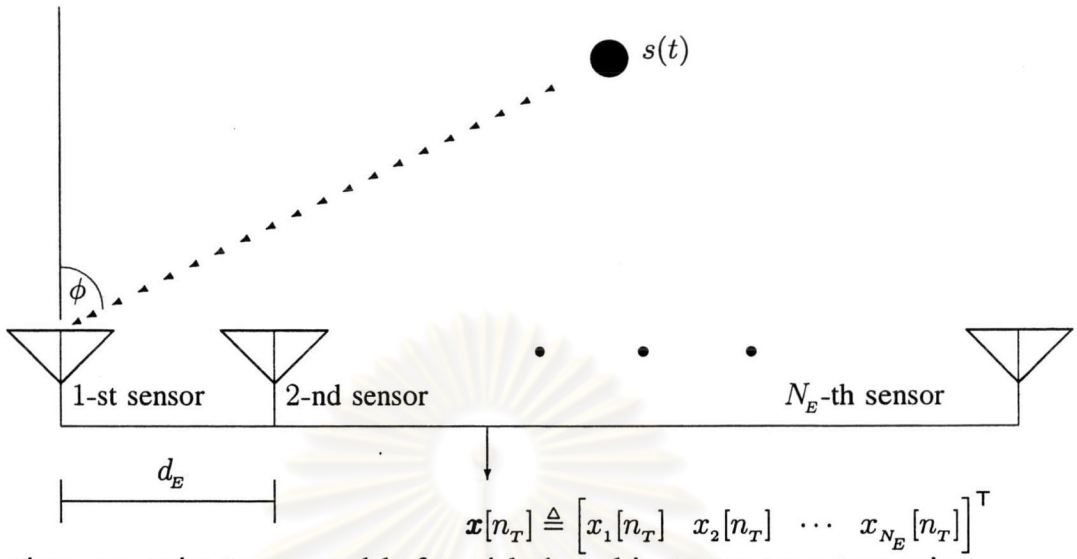

Figure 2.1 Point Source Model of Spatial Channel in Sensor Array Processing

> "*The purpose of a scientific model is to enable the analyst to determine how one or more changes in various aspects of the modelled system may affect other aspects of the system or the system as a whole.*
>
> .....
>
> *Often we think that the more details a model includes the better it resembles reality.*
>
> .....
>
> *All that is required is that there be a high correlation between the prediction by the model and what would actually happen with the real system. To ascertain whether this requirement is satisfied or not, it is important to test and establish control over the solution*"
>
> Reuven Y. Rubinstein In *Simulation and the Monte Carlo Method* (1981)

Later, a brief on spatially distributed localization model is provided. In an experiment of multibeam echo sounder system, Jäntti [6] studied the superposition of plane waves and then modelled it as a continuum of multiple directions. This work was further studied in [7] to find the nominal direction and its associating maximum deviation in two frameworks, such as coherently and incoherently distributed sources.

During the post doctoral degree at Royal Institute of Technology, Trump considered the same problem in communication applications when there is no *l*ine-*o*f-*s*ight (LOS) between the source and the receiving antenna array [8]. His attempt was to make use of the second-order statistic and to model all deviated directions as a reasonable spatial

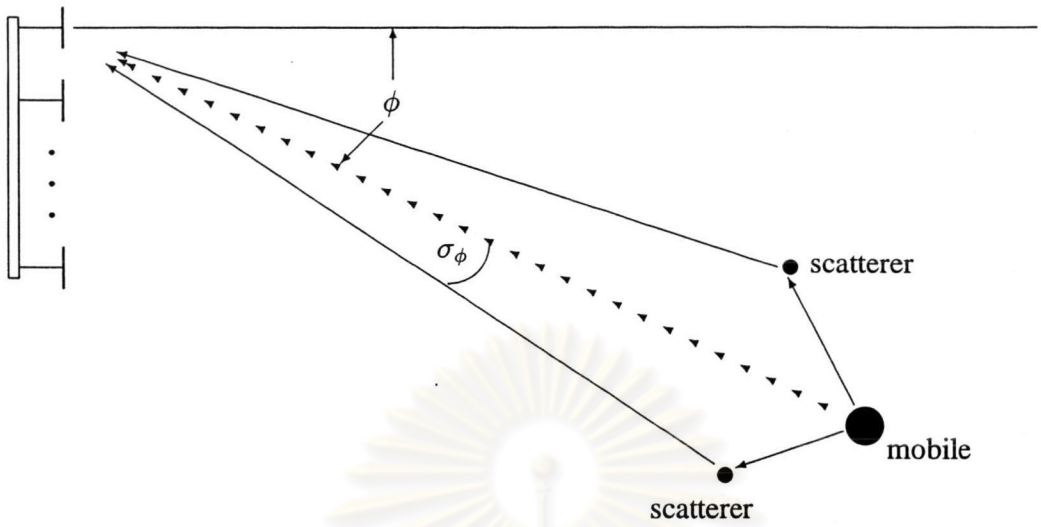

Figure 2.2 Distributed Source Model of Spatial Channel in Sensor Array Processing

distribution function, that is Gaussian PDF. Then, he tried to estimate the parameters of this distribution instead of marginal directions of arrival [9]. This work implicitly addressed the situations where most of classical point source models will suffer from non-identifiability in the presence of large number of directions. To deviate from this problem as well, Zetterberg's propagation model [10] assumed that the number of multipaths is large enough in order that their path gains can be characterized, under the central limit theorem, by a Gaussian random variable whose associated directions are also random. By means of local scattering around the vicinity of source, Asztély proposed a generalized array manifold model in term of deterministic quantity for estimating the parameterized spatial signature [11]. To relieve estimation complexity, Bengtsson's *spatial-frequency* approximation model [12] later accounted for partitioning the covariance matrix of spatial fading channel into a separable parameterization.

The spatial channel model described above has been validated against experimental data collected by Ericsson Radio Systems. In the field experiments, a transmitter has been placed in urban areas approximately 1km from the receiving array [13]. Furthermore, a lot of measurements have shown that local scattering in the vicinity of a mobile is a non-negligible phenomenon [14].

In the vicinity of the mobile sources, local scattering leads to fading statistic and angular spread of the signals incident on the array. Herein, a time-invariant scenario with small angular spread (see *e.g.*, [8] [10] [11] and [12]) is considered as illustrated in Fig. 2.2.

2.2 Mathematical Abstraction

Restrict our attention to a signal transmitting through a dispersive channel and then impinging on the ULA. With phase reference at the first element, or *f*irst *e*lement *r*eference (FER), the array response vector $\boldsymbol{a}(\phi) : [-\frac{\pi}{2}, \frac{\pi}{2}] \mapsto \mathbb{C}^{N_E \times 1}$ can be, in general, written ideally as (see *e.g.*, [15, ch. 2], [16], [17], [18])

$$\boldsymbol{a}(\phi) \triangleq \left[1 \quad e^{\imath k d_E \sin(\phi)} \quad \cdots \quad e^{\imath k d_E (N_E - 1)\sin(\phi)}\right]^{\mathsf{T}} \tag{2.1}$$

where $k = \frac{2\pi}{\lambda}$ designates the wave number with associating wavelength λ, d_E signifies the equi-distance between two adjacent elements, and N_E is the number of sensor elements. As previously developed, most local scattering models assume that the nominal angle ϕ is deterministic while angular deviation δ_ϕ and associating path gain γ are considered as stochastic quantities. In Fig. 2.2, the array output at time instant n_T can be characterized in a flat fading channel by the snapshot $\boldsymbol{x}[n_T] \in \mathbb{C}^{N_E \times 1}$. Mathematically speaking, it can be represented according to linear regression as [12, p. 25]

$$\boldsymbol{x}[n_T] = s[n_T] \sum_{n_P=1}^{N_P} \gamma_{n_P}[n_T]\boldsymbol{a}(\phi + \delta_{\phi_{n_P}}[n_T]) + \boldsymbol{n}[n_T] \tag{2.2}$$

where N_P denotes the number of scattering paths and $\boldsymbol{n}[n_T] \in \mathbb{C}^{N_E \times 1}$ designates the additive noise at sensor array. It can be collected in a data burst $\boldsymbol{X} \in \mathbb{C}^{N_E \times N_T}$ as

$$\boldsymbol{X} \triangleq \left[\boldsymbol{x}[1] \quad \boldsymbol{x}[2] \quad \cdots \quad \boldsymbol{x}[N_T]\right]. \tag{2.3}$$

For a large number of incoming rays, the channel vector [10]

$$\boldsymbol{h}[n_T] \triangleq \sum_{n_P=1}^{N_P} \gamma_{n_P}[n_T]\boldsymbol{a}(\phi + \delta_{\phi_{n_P}}[n_T]) \tag{2.4}$$

seemed, under the central limit theorem, plausible to hold a circularly-symmetric complex-valued Gaussian process, *i.e.*, $\boldsymbol{h}[n_T] \sim \mathcal{N}_c(\boldsymbol{0}; \boldsymbol{\Sigma}_{hh}, \boldsymbol{O})$. This N_E-dimensional variate implicitly provides the statistic[1] $\boldsymbol{\Sigma}_{hh} \triangleq \mathcal{E}\langle \tilde{\boldsymbol{h}}[n_T]\tilde{\boldsymbol{h}}^{\mathsf{H}}[n_T]\rangle \in \mathbb{C}_{\mathbb{H}}^{N_E \times N_E}$, where $\tilde{\boldsymbol{h}}[n_T] \triangleq \boldsymbol{h}[n_T] - \mathcal{E}\langle \boldsymbol{h}[n_T]\rangle = \boldsymbol{h}[n_T]$. For taking into account an incoherently distributed channel [7], the second-order statistic of a certain incoming ray yields [12]

$$\mathcal{E}\langle \gamma_{n_P}[n_T]\gamma^*_{\acute{n}_P}[\acute{n}_T]\rangle = \sigma^2_\gamma \delta_{n_P,\acute{n}_P}\delta_{n_T,\acute{n}_T} \tag{2.5}$$

[1] By means of spatially extended or distributed source, the indicator to assess the extension of such a source is given from the eigenvalue distribution of the channel covariance matrix $\boldsymbol{\Sigma}_{hh}$.

where $\delta_{\cdot,\cdot}$ signifies the Krönecker delta function, $\mathcal{E}\langle\cdot\rangle$ denotes the statistical expectation operator and σ_γ^2 designates the power in each path. Over the spatial continuum of interest, the channel covariance can be approximated as

$$\boldsymbol{\Sigma}_{hh}(\rho,\phi,\sigma_\phi) \approx \rho \int f(\delta_\phi|0;\sigma_\phi^2)\boldsymbol{a}(\phi+\delta_\phi)\boldsymbol{a}^{\mathsf{H}}(\phi+\delta_\phi)d\delta_\phi \tag{2.6}$$

where $\rho \triangleq N_P\sigma_\gamma^2$ signifies the cluster power due to all paths and $f(\delta_\phi|0;\sigma_\phi^2)$ denotes the conditional PDF for random deviation δ_ϕ given a priori knowledge of the angular spread σ_ϕ. As encountered, a family of symmetric distributions with zero mean and variance σ_ϕ^2 is in most modelled as follows

$$f(\delta_\phi|0;\sigma_\phi^2) = \begin{cases} \frac{1}{2\sqrt{3}\sigma_\phi} \sqcap [-\sqrt{3}\sigma_\phi, \sqrt{3}\sigma_\phi] & ;\text{Uniform} \\ \frac{1}{\sqrt{2\pi}\sigma_\phi} e^{-\frac{1}{2}\frac{\delta_\phi^2}{\sigma_\phi^2}} & ;\text{Gaussian} \\ \frac{1}{\sqrt{2}\sigma_\phi} e^{-\frac{1}{\sigma_\phi}\sqrt{2}|\delta_\phi|} & ;\text{Laplacian.} \end{cases} \tag{2.7}$$

Rather than estimating the directional parameter through physical angles ϕ and σ_ϕ, the spatial frequency response is preferable due to the better accuracy of truncating the Taylor series with first-order around the array broadside [12]. In general, the spatial frequency ω and its associating standard deviation σ_ω are introcuced by

$$\omega(\phi) = kd_{_E}\sin(\phi) \tag{2.8a}$$

$$\sigma_\omega(\phi,\sigma_\phi) = kd_{_E}\cos(\phi)\sigma_\phi. \tag{2.8b}$$

Accounting for small angular spread, the so-called *spatial frequency* approximation results in a separable form as

$$\boldsymbol{\Sigma}_{hh}(\rho,\omega,\sigma_\omega) \simeq \rho\boldsymbol{D}_a(\omega)\boldsymbol{B}(\sigma_\omega)\boldsymbol{D}_a^{\mathsf{H}}(\omega) \tag{2.9}$$

where $\boldsymbol{D}_a(\omega) : [-kd_{_E}, kd_{_E}] \mapsto \mathbb{C}_{\mathbb{D},\mathbb{U}}^{N_E\times N_E}$ is diagonal and unitary matrix parameterized by nominal angle and $\boldsymbol{B}(\sigma_\omega) : \mathbb{R}_+^{1\times 1} \mapsto \mathbb{R}_{\mathbb{S},\mathbb{T}}^{N_E\times N_E}$ is symmetric Toeplitz matrix parameterized by angular spread. Their $(n_{_E},\acute{n}_{_E})$-th elements can be expressed by [12, p. 22]

$$[\boldsymbol{D}_a(\omega)]_{[n_{_E},\acute{n}_{_E}]} = e^{\imath(n_{_E}-1)\omega}\delta_{n_{_E},\acute{n}_{_E}} \tag{2.10a}$$

$$[\boldsymbol{B}(\sigma_\omega)]_{[n_{_E},\acute{n}_{_E}]} = f_{\mathcal{F}}((n_{_E}-\acute{n}_{_E})\sigma_\omega|0,1) \tag{2.10b}$$

whence characteristic function $f_{\mathcal{F}}(t|,0,1) \triangleq \mathcal{F}(f(\delta_\omega|,0,1))$ is equivalent to the Fourier transform $\mathcal{F}(\cdot)$ of the associating PDF whose random variable holds zero-mean and

unit variance. In a certain situation, the $(n_E, \acute{n}_E)$-th element in $\boldsymbol{B}(\sigma_\omega)$ can be expressed as (see *e.g.*, [12] and [19])

$$[\boldsymbol{B}(\sigma_\omega)]_{[n_E,\acute{n}_E]} = \begin{cases} \frac{\sin((n_E-\acute{n}_E)\sqrt{3}\sigma_\omega)}{(n_E-\acute{n}_E)\sqrt{3}\sigma_\omega} & ;\text{uniform} \\ e^{-\frac{1}{2}(n_E-\acute{n}_E)^2\sigma_\omega^2} & ;\text{Gaussian} \\ \frac{1}{1+\frac{1}{2}(n_E-\acute{n}_E)^2\sigma_\omega^2} & ;\text{Laplacian.} \end{cases} \tag{2.11}$$

If additive noise assumed is spatially uncorrelated noise and absolutely uncorrelated from channels, it results in

$$\boldsymbol{\Sigma}_{xx}[n_T] = p[n_T]\boldsymbol{D}_a(\omega)\boldsymbol{B}(\sigma_\omega)\boldsymbol{D}_a^{\mathsf{H}}(\omega) + \sigma_n^2\boldsymbol{I} \tag{2.12}$$

where $p[n_T] \triangleq \rho|s[n_T]|^2$ stands for the total power observed at the sensor array. In what follows, we shall consider only the deterministic signal with constant modulus so that $\boldsymbol{\Sigma}_{xx}[n_T] = \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}_\circ)$; $\forall n_T$, where $\boldsymbol{\theta}_\circ$ is the true value of model parameter. For all N_S signals, the array covariance matrix becomes

$$\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}_\omega) = \sum_{n_S=1}^{N_S} p_{n_S}\boldsymbol{D}_a(\omega_{n_S})\boldsymbol{B}(\sigma_{\omega_{n_S}})\boldsymbol{D}_a^{\mathsf{H}}(\omega_{n_S}) + \sigma_n^2\boldsymbol{I}. \tag{2.13}$$

2.3 Problem Formulation

Now suppose that, based on the second-order statistic $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}_\circ)$, our problem is to find the nominal directions of arrival, ϕ, and/or the underlying angular spreads, $\boldsymbol{\sigma}_\phi$, given the collected data $\boldsymbol{x}[n_T]$; $\forall n_T$, where true-valued parameter vector $\boldsymbol{\theta}_\circ \in \mathbb{R}^{(3N_S+1)\times 1}$ due to multiple signals can be expressed as

$$\boldsymbol{\theta}_\phi \triangleq \begin{bmatrix} \boldsymbol{\phi}^{\mathsf{T}} & \boldsymbol{\sigma}_\phi^{\mathsf{T}} & \boldsymbol{p}^{\mathsf{T}} & \sigma_n^2 \end{bmatrix}^{\mathsf{T}} \tag{2.14a}$$

$$\boldsymbol{\theta}_\omega \triangleq \begin{bmatrix} \boldsymbol{\omega}^{\mathsf{T}} & \boldsymbol{\sigma}_\omega^{\mathsf{T}} & \boldsymbol{p}^{\mathsf{T}} & \sigma_n^2 \end{bmatrix}^{\mathsf{T}} \tag{2.14b}$$

for the physical and spatial frequency models, respectively.

2.4 Optimal Solution

The term "optimal" in most framework is quoted when comparing the statistical error variance with the inherent accuracy limitation in the model. Next section will be devoted to the Cramér-Rao bound which is well-known as an inherent accuracy limitation of sensor array processing. In this section, we shall concern the exact maximum likelihood estimator because its asymptotic performance is plausible to attain the Cramér-Rao bound.

Assume that the Nyquist-based sampling rate is conducted in such a way that several samples are all statistically independent *i.e.*,

$$f_{\boldsymbol{X}}(\boldsymbol{x}[1],\boldsymbol{x}[2],\ldots,\boldsymbol{x}[N_T]|\boldsymbol{\theta}) = \prod_{n_T=1}^{N_T} f_{\boldsymbol{x}}(\boldsymbol{x}[n_T]|\boldsymbol{\theta}). \tag{2.15}$$

As mentioned above, all available snapshots are based on the central limit theorem such that

$$\mathcal{E}\left\langle \boldsymbol{x}[n_T]\right\rangle = \boldsymbol{0} \tag{2.16a}$$

$$\mathcal{E}\left\langle (\boldsymbol{x}[n_T]-\mathcal{E}\left\langle \boldsymbol{x}[n_T]\right\rangle)(\boldsymbol{x}[\acute{n}_T]-\mathcal{E}\left\langle \boldsymbol{x}[\acute{n}_T]\right\rangle)^{\mathsf{T}}\right\rangle = \boldsymbol{O} \tag{2.16b}$$

$$\mathcal{E}\left\langle (\boldsymbol{x}[n_T]-\mathcal{E}\left\langle \boldsymbol{x}[n_T]\right\rangle)(\boldsymbol{x}[\acute{n}_T]-\mathcal{E}\left\langle \boldsymbol{x}[\acute{n}_T]\right\rangle)^{\mathsf{H}}\right\rangle = \delta_{n_T,\acute{n}_T}\boldsymbol{\Sigma}_{xx}. \tag{2.16c}$$

Therefore, the complex N_E-dimensional variate $\boldsymbol{x}[n_T] \sim \mathcal{N}_c\left(\boldsymbol{0};\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}),\boldsymbol{O}\right)$ is immediately governed by [20]

$$f_{\boldsymbol{x}}(\boldsymbol{x}[n_T]|\boldsymbol{\theta}) = \frac{1}{\pi^{N_E}|\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta})|}e^{-\boldsymbol{x}^{\mathsf{H}}[n_T]\boldsymbol{\Sigma}_{xx}^{-1}(\boldsymbol{\theta})\boldsymbol{x}[n_T]}. \tag{2.17}$$

The joint probability density function of all snapshots $\boldsymbol{x}[n_T]$ allows us to the likelihood function (see *e.g.*, [21], [22] and [23])

$$\begin{aligned} l_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X}) &\triangleq f_{\boldsymbol{X}}(\boldsymbol{x}[1],\boldsymbol{x}[2],\ldots,\boldsymbol{x}[N_T]|\boldsymbol{\theta}) \\ &= \prod_{n_T=1}^{N_T} f_{\boldsymbol{x}}(\boldsymbol{x}[n_T]|\boldsymbol{\theta}). \end{aligned} \tag{2.18}$$

Let the negative normalized log likelihood with neglecting the constant term be

$$\ell_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}) \triangleq -\frac{1}{N_T}\ln l_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X}) + N_E \ln \pi. \tag{2.19}$$

It was proposed that the *m*aximum *l*ikelihood (ML) estimator of $\boldsymbol{\theta}$ is equivalent to [9]

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\min_{\boldsymbol{\theta}} \ell_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}) \tag{2.20a}$$

$$\ell_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}) = \lceil \boldsymbol{\Sigma}_{xx}^{-1}(\boldsymbol{\theta})\hat{\boldsymbol{\Sigma}}_{xx}\rfloor + \ln|\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta})| \tag{2.20b}$$

where the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{xx} \in \mathbb{C}_{\mathbb{H}}^{N_E\times N_E}$ is computed from

$$\hat{\boldsymbol{\Sigma}}_{xx} = \frac{1}{N_T}\sum_{n_T=1}^{N_T}\boldsymbol{x}[n_T]\boldsymbol{x}^{\mathsf{H}}[n_T]. \tag{2.21}$$

It is important to note that the exact likelihood function $\ell_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}) : \mathbb{R}^{(3N_S+1)\times 1} \mapsto \mathbb{R}^{1\times 1}$ straightforwardly requires $(3N_S+1)$-dimensional search to arrive at the critical point for the ML solution $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$.

Most works involved the direction finding problem were based on the exact ML estimator due to its producible optimality [4], [20]. In order to arrive at extremal quantity, the optimization search of likelihood function seems, in general, inevitable for a complex model. However, the physical model with well-described characterization would, in nature, requires large number of model parameters. As a consequence, the larger the number of model parameters, the larger the dimension of optimization search over parameter space. Unfortunately, this might prohibit the exact ML estimator from being incorporated into real-word applications.

2.5 Sub-Optimal Solutions

For jointly estimating the nominal direction and its associating angular spread, there is a *w*eighted *l*east *s*quares (WLS) estimator proposed in [9]. Since this requires $2N_s$-dimensional search, it is reasonable to review, under the same dimension, a *l*east *s*quares (LS) approach [40] which is called therein as *r*edundancy *a*veraging *c*ovariance *m*atching (RACM) method. Note that the subspace-based methods such as [24], [25], [26], [27], [28], [29], [30] and [31] can be neglected herein because the local scattering model leads to full-rank channel covariance matrix. Hence, all subspace-based estimators are biased eventually (see *e.g.*, [12] and [32]). Moreover, it was shown in [33] and [34] that beamforming and linear prediction can carry out the distributed source localization only in small angular spread.

2.5.1 WLS Estimator

Trying to reduce the optimization tasks, the WLS approach is preferred instead of the criterion based on likelihood function [35]. Such an idea stems from the fact that both ML and WLS methods yield the same asymptotic performance when taking into acount any Gauss-Markov model (see *e.g.*, [22, pp. 127–128], [23, pp. 566–567] and [36]). For making the WLS more attractive in computational cost than the ML, there were various efforts to reduce the dimension of optimization during the computation of the parameter estimate. A reasonable way is to replace the optimal weight with other one which is consistent, or obviously, converges to the optimal weight. The most famous way usually makes use of a nonparametric estimation of such value, *e.g.*, the sample covariance estimate. The cause of this is that in a fairly general situation, such a statistic is not only easy to compute but also holds the performance of asymptotic consistency [20].

The WLS estimation is to find a parametric argument which provides the smallest residual in matching criteria. In spatially distributed source localization, most work is

relied upon the second-order statistics. Therefore, the array covariance matrix is one to be matched between theoretical and receivable quantities [9]. Being concentrated on the source signal power and noise variance, the WLS loss function is arrived at [35]

$$\hat{\boldsymbol{\vartheta}}_{\text{WLS}} = \arg\min_{\boldsymbol{\vartheta}_\omega} \hat{\boldsymbol{\xi}}_x^{\mathsf{H}} \boldsymbol{W}^{-\frac{1}{2}} \boldsymbol{\Pi}^{\perp}_{\mathcal{S}_{\mathcal{R}}(\boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{\Omega}(\boldsymbol{\vartheta}_\omega))} \boldsymbol{W}^{-\frac{1}{2}} \hat{\boldsymbol{\xi}}_x \tag{2.22}$$

where $\hat{\boldsymbol{\xi}}_x = \upsilon_{\mathsf{c}}(\hat{\boldsymbol{\Sigma}}_{xx}) \in \mathbb{C}^{N_E^2 \times 1}$ is the covariance vectorization with column-stacking vectorization operator $\upsilon_{\mathsf{c}}(\mathbf{A}) : \mathbb{C}^{N_{rA} \times N_{cA}} \mapsto \mathbb{C}^{(N_{rA}N_{cA})\times 1}$, and $\boldsymbol{\vartheta}_\omega \triangleq \left[\boldsymbol{\omega}^{\mathsf{T}} \quad \boldsymbol{\sigma}_\omega^{\mathsf{T}}\right]^{\mathsf{T}} \in \mathbb{R}^{(2N_S)\times 1}$ results in $\boldsymbol{\Omega}(\boldsymbol{\vartheta}_\omega)$ expressed by

$$\begin{aligned} \boldsymbol{\Omega}(\boldsymbol{\omega}, \boldsymbol{\sigma}_\omega) \triangleq \Big[& \upsilon_{\mathsf{c}}(\boldsymbol{D}_a(\omega_1)\boldsymbol{B}(\sigma_{\omega_1})\boldsymbol{D}_a^{\mathsf{H}}(\omega_1)) \quad \cdots \\ & \upsilon_{\mathsf{c}}(\boldsymbol{D}_a(\omega_{N_S})\boldsymbol{B}(\sigma_{\omega_{N_S}})\boldsymbol{D}_a^{\mathsf{H}}(\omega_{N_S})) \quad \upsilon_{\mathsf{c}}(\boldsymbol{I}) \Big]. \end{aligned} \tag{2.23}$$

To deal with single signal scenario, the parameter of interest, $\boldsymbol{\vartheta}_\omega$, will be reduced to ω and thus its corresponding $\boldsymbol{\Omega}(\omega)$ is given from [37]

$$\boldsymbol{\Omega}(\omega) \triangleq \left(\boldsymbol{D}_a^{\mathsf{H}}(\omega) \otimes \boldsymbol{D}_a(\omega)\right) \boldsymbol{\Xi} \tag{2.24}$$

where the binary selection matrix $\boldsymbol{\Xi} \in \mathbb{B}_{\mathbb{F}}^{N_E^2 \times N_E}$ being of full rank is

$$\boldsymbol{\Xi} \triangleq \left[\upsilon_{\mathsf{c}}(\boldsymbol{I}) \quad \upsilon_{\mathsf{c}}(\boldsymbol{L}_1 + \boldsymbol{L}_1^{\mathsf{T}}) \quad \cdots \quad \upsilon_{\mathsf{c}}(\boldsymbol{L}_{N_E-1} + \boldsymbol{L}_{N_E-1}^{\mathsf{T}})\right] \tag{2.25}$$

with the block-lower triangular matrix $\boldsymbol{L}_{n_L} \in \mathbb{B}^{N_E \times N_E}$ according to

$$\boldsymbol{L}_{n_L} \triangleq \begin{bmatrix} \boldsymbol{O}_{(n_L \times (N_E - n_L))} & \boldsymbol{O}_{(n_L)} \\ \boldsymbol{I}_{(N_E - n_L)} & \boldsymbol{O}_{((N_E - n_L)\times n_L)} \end{bmatrix}. \tag{2.26}$$

Now the problem of interest belongs to what the weight matrix $\boldsymbol{W}$ should be. To make the residual $\tilde{\boldsymbol{\vartheta}}_{\text{WLS}} \triangleq \hat{\boldsymbol{\vartheta}}_{\text{WLS}} - \boldsymbol{\vartheta}_\circ$ minimal, the optimal weight should be satisfied by [35]

$$\boldsymbol{W} = \lim_{N_T \to \infty} N_T \mathcal{E}\left\langle \tilde{\boldsymbol{\xi}}_x \tilde{\boldsymbol{\xi}}_x^{\mathsf{H}} \right\rangle = \boldsymbol{\Sigma}_{xx}^{\mathsf{T}} \otimes \boldsymbol{\Sigma}_{xx} \tag{2.27}$$

where $\tilde{\boldsymbol{\xi}}_x \triangleq \hat{\boldsymbol{\xi}}_x - \boldsymbol{\xi}_x$ is the the sample covariance residual. Since the exact weight depends itself on the model parameter, it is preferable to make use of nonparametric estimate $\hat{\boldsymbol{W}} = \hat{\boldsymbol{\Sigma}}_{xx}^{\mathsf{T}} \otimes \hat{\boldsymbol{\Sigma}}_{xx}$ rather than $\boldsymbol{W}$ in (2.22) without loss of asymptotic performance.

2.5.2 RACM Estimator

Unlike the exact ML and WLS presented in the previous, the redundancy-averaged covariance-matching method was not designed for handling the local scattering model. Instead of such, it was developed from the *imperfect spatial wavefront* model which is

an extension of the classical point source model in array signal processing. It borrowed the key idea from ordinary (unweighted) least squares. Interesting enough, it is rather sophisticate because the vector to be fitted in its matching criterion is given by sufficient statistic, *i.e.*, the first column of the Toeplitz-constrained covariance estimate. This is given by preprocessing the sample covariance in such a way that the new covariance estimate holds the Toeplitz structure. The technique employed in this step is borrowed from the redundancy-averaging approach which will be described in details later. Here we shall recall the objective function of RACM method.

Let $\boldsymbol{\mu}_{n_S}[n_T] \in \mathbb{C}^{N_E \times 1}$ be the time-varying perturbations of the n_S-th source. Its second-order statistic is governed by $\boldsymbol{\mu}_{n_S}[n_T] \sim \mathcal{N}_c\,(\boldsymbol{0}; \boldsymbol{\Sigma}_{uu}, \boldsymbol{O})$, where $\boldsymbol{\Sigma}_{uu} \in \mathbb{R}_{\mathbb{T},\mathbb{S}}^{N_E \times N_E}$ is the associating covariance matrix. In the imperfect spatial wavefront model, the output snapshot is modelled as (see *e.g.*, [38], [39] and [40])

$$\boldsymbol{x}[n_T] = \sum_{n_S=1}^{N_S} s_{n_S}[n_T]\big(\boldsymbol{\mu}_{n_S}[n_T] \odot \boldsymbol{a}(\phi_{n_S})\big) + \boldsymbol{n}[n_T]. \tag{2.28}$$

In the same fashion, this leads to the array covariance matrix as

$$\boldsymbol{\Sigma}_{xx}[n_T] = \sum_{n_S=1}^{N_S} |\check{p}_{n_S}[n_T]|^2 \big(\boldsymbol{a}(\phi_{n_S})\boldsymbol{a}^{\mathsf{H}}(\phi_{n_S}) \odot \boldsymbol{\Sigma}_{uu}(\nu_{n_S})\big) + \sigma_n^2 \boldsymbol{I} \tag{2.29}$$

where $\nu \in \mathbb{R}^{1\times 1}$ is a sufficient parameter in $\boldsymbol{\Sigma}_{uu}(\nu)$. We refer to [38], [39] and [40] for more details on the paramterization of ν in $\boldsymbol{\Sigma}_{uu}(\nu)$. Over the batch data $\boldsymbol{X}$, the model parameter $\boldsymbol{\theta} \in \mathbb{R}^{(3N_S+1)\times 1}$ can be expressed as

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\phi}^{\mathsf{T}} & \boldsymbol{\nu}^{\mathsf{T}} & \check{\boldsymbol{p}}^{\mathsf{T}} & \sigma_n^2 \end{bmatrix}^{\mathsf{T}} \tag{2.30}$$

where $\check{\boldsymbol{p}} \in \mathbb{R}^{N_S \times 1}$ signifies the signal power, defined as

$$\check{\boldsymbol{p}} \triangleq \frac{1}{N_T} \sum_{n_T=1}^{N_T} \boldsymbol{s}[n_T] \odot \boldsymbol{s}^*[n_T]. \tag{2.31}$$

It is important to note that this model can be regarded as the unstructured model of the local scattering model. However, the inherent accuracy in this model seems to be different from the local scattering model because it was assumed to possess the spatial perturbation in different idea.

Let us introduce the array response matrix $\boldsymbol{A}(\boldsymbol{\phi}) : [-\frac{\pi}{2}, \frac{\pi}{2}]^{N_S \times 1} \mapsto \mathbb{C}^{N_E \times N_S}$ and the imperfect spatial parameterization matrix $\boldsymbol{U}(\boldsymbol{\nu}) \in \mathbb{R}^{N_E \times N_S}$ according to

$$\boldsymbol{A}(\boldsymbol{\phi}) = \begin{bmatrix} \boldsymbol{a}(\phi_1) & \boldsymbol{a}(\phi_2) & \cdots & \boldsymbol{a}(\phi_{N_S}) \end{bmatrix} \tag{2.32a}$$

$$\boldsymbol{U}(\boldsymbol{\nu}) = \begin{bmatrix} \boldsymbol{u}_1(\nu_1) & \boldsymbol{u}_2(\nu_2) & \cdots & \boldsymbol{u}_{N_S}(\nu_{N_S}) \end{bmatrix} \tag{2.32b}$$

If $\hat{\varsigma}_{\mathrm{RA}} \in \mathbb{C}^{N_E \times 1}$ is the RA estimate of the first column vector in the Toeplitz-constrained covariance estimate of $\bar{\boldsymbol{\Sigma}}_{xx}$, the the RACM objective function is formulated as

$$\{\phi, \nu, \check{\boldsymbol{p}}\}_{\mathrm{RACM}} \triangleq \arg\min_{\phi,\nu,\check{\boldsymbol{p}}} \|(\boldsymbol{A}(\phi) \odot \boldsymbol{U}(\nu))\check{\boldsymbol{p}} - \hat{\varsigma}_{\mathrm{RA}}\|_{\mathbb{E}}^2. \tag{2.33}$$

For estimating both ϕ and ν, one can concentrate on the RACM function in order to derive [40]

$$\{\phi, \nu\}_{\mathrm{RACM}} = \arg\max_{\phi,\nu} \Re\big(\hat{\tau}_{\mathrm{RA}}^{\mathsf{H}} \boldsymbol{A}_U(\phi,\nu)\big)\Re^{-1}\big(\boldsymbol{A}_U^{\mathsf{H}}(\phi,\nu)\boldsymbol{A}_U(\phi,\nu)\big)\Re\big(\boldsymbol{A}_U^{\mathsf{H}}(\phi,\nu)\hat{\varsigma}_{\mathrm{RA}}\big) \tag{2.34}$$

where $\boldsymbol{A}_U(\phi,\nu) \triangleq \boldsymbol{A}(\phi) \odot \boldsymbol{U}(\nu) \in \mathbb{C}^{N_E \times N_S}$.

2.6 Cramér-Rao Bound

The *C*ramér-*R*ao *b*ound (CRB) is, in most, a lower bound on the error variance of estimated parameter from which the estimator will be concerned. It might be considered as a benchmark of achievable performances from efficient estimators. Herein, the CRB is accounted for vector parameter version due to the model taken several parameters into account. Firstly, we would recall the Cramér-Rao inequality theorem for inspiring the further specifications of our scenario at hand.

Definition 1. *Let* $\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta}) : \mathbb{R}^{(3N_S+1)\times 1} \mapsto \mathbb{R}_{\mathbb{S}}^{(3N_S+1)\times(3N_S+1)}$ *be the Fisher information matrix (FIM) which is expressed as [20]*

$$\begin{aligned} \boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta}) &= \mathcal{E}\left\langle \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln l_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln l_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X})\right)^{\mathsf{T}} \right\rangle \\ &= -\mathcal{E}\left\langle \frac{\partial^2}{\partial \boldsymbol{\theta} \otimes \partial \boldsymbol{\theta}^{\mathsf{T}}} \ln l_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X}) \right\rangle. \end{aligned} \tag{2.35}$$

Here we introduce $\mathbf{A} \succeq \mathbf{B}$ as the positive semi-definiteness[2] of $\mathbf{A} - \mathbf{B}$.

Theorem 1 (Cramér-Rao inequality). *If any unbiased estimator* $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ *of the true-valued parameter vector* $\boldsymbol{\theta}_\circ$ *is estimable from the observation data* $\boldsymbol{X}$*, we then have a condition of*

$$\mathcal{E}\big\langle \hat{\boldsymbol{\theta}}(\boldsymbol{X}) \big\rangle = \boldsymbol{\theta}_\circ. \tag{2.36}$$

Under the regular condition $\mathcal{E}\left\langle \frac{\partial}{\partial \boldsymbol{\theta}} \ln l^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X}) \right\rangle = \boldsymbol{0}$*, there exists Cramér-Rao inequality given by*

$$\mathcal{E}\big\langle (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\circ)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\circ)^{\mathsf{T}} \big\rangle \succeq \boldsymbol{B}_{\mathrm{CR}(\hat{\boldsymbol{\theta}})}(\boldsymbol{\theta}_\circ) \tag{2.37}$$

---

[2] **Definition 2.** *If* $\mathbf{Z} \in \mathbb{C}_{\mathbb{H}}^{N_Z \times N_Z}$ *is positive semi-definite, then we have* $\mathbf{y}^{\mathsf{H}}\mathbf{Z}\mathbf{y} \geq 0$ *for any* $\mathbf{y} \in \mathbb{C}^{N_Z \times 1}$.

where the Cramér-Rao bound matrix $\boldsymbol{B}_{\mathrm{CR}(\hat{\boldsymbol{\theta}})}(\boldsymbol{\theta}_\circ) : \mathbb{R}^{(3N_S+1)\times 1} \mapsto \mathbb{R}_{\mathrm{S}}^{(3N_S+1)\times(3N_S+1)}$ *is available from*

$$\boldsymbol{B}_{\mathrm{CR}(\hat{\boldsymbol{\theta}})}(\boldsymbol{\theta}_\circ) = \boldsymbol{I}_{\mathrm{F}}^{-1}(\boldsymbol{\theta}_\circ). \tag{2.38}$$

Proof. See *e.g.*, [41, pp. 66–67], [36, Appendix 3B] and [42, pp. 288–289]. □

It is systematically preferable to emphasize that the first step for calculating the error covariance bound is to derive the FIM.

2.6.1 Explicit Derivatives

Lemma 1 (Slepian-Bangs formula). *Straightforward calculating the preceding theorem, the second-order derivative of the likelihood function* $l_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X})$ *or, equivalently,* $\ell_{\mathrm{ML}}^{[N_T]}(\boldsymbol{\theta}|\boldsymbol{X})$ *are combatable. This leads to the* $(n,\acute{n})$*-th element of the FIM as*[3]

$$[\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta})]_{[n,\acute{n}]} = N_T \lceil \boldsymbol{\Sigma}_{xx}^{-1} \left(\frac{\partial}{\partial \theta_n} \boldsymbol{\Sigma}_{xx}(\theta_n) \right) \boldsymbol{\Sigma}_{xx}^{-1} \frac{\partial}{\partial \theta_{\acute{n}}} \boldsymbol{\Sigma}_{xx}(\theta_{\acute{n}}) \rfloor \tag{2.40}$$

where the scalars θ_n *and* $\theta_{\acute{n}}$ *are the* n*-th and* $\acute{n}$*-th elements of the parameter vector* $\boldsymbol{\theta}$ *for any indices* $n, \acute{n} \in \{1, 2, \ldots, 3N_S + 1\}$ *correspondingly.*

Proof. See [20, p. 21]. □

2.6.2 Separable Derivatives

Proceeding on the Slepian-Bangs formula, a convenient way to represent the FIM in compact form is to express the derivative into a Jacobian matrix.

Lemma 2 (Jansson-Göransson formula). *Let* $\nabla_\theta(\boldsymbol{\xi}_x) \in \mathbb{C}^{N_E^2 \times (3N_S+1)}$ *be the derivative matrix defined by*

$$\nabla_\theta(\boldsymbol{\xi}_x) \triangleq \frac{\partial}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \boldsymbol{\xi}_x(\boldsymbol{\theta}) \tag{2.41}$$

[3] Actually, there has been a generalization of the Slepian-Bangs formula as

$$\begin{aligned}[\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta})]_{[n,\acute{n}]} = {} & N_T \lceil \boldsymbol{\Sigma}_{xx}^{-1} \left(\frac{\partial}{\partial \theta_n} \boldsymbol{\Sigma}_{xx}(\theta_n) \right) \boldsymbol{\Sigma}_{xx}^{-1} \frac{\partial}{\partial \theta_{\acute{n}}} \boldsymbol{\Sigma}_{xx}(\theta_{\acute{n}}) \rfloor \\ & + 2N_T \Re \left(\left(\frac{\partial}{\partial \theta_n} \boldsymbol{\mu}_x(\theta_n) \right)^{\mathsf{H}} \boldsymbol{\Sigma}_{xx}^{-1} \frac{\partial}{\partial \theta_{\acute{n}}} \boldsymbol{\mu}_x(\theta_{\acute{n}}) \right).\end{aligned} \tag{2.39}$$

However, two special scenarios where $\boldsymbol{\mu}_x \triangleq \mathcal{E}\langle \boldsymbol{x}[n_T] \rangle$ is a zero vector or it does not depend on $\boldsymbol{\theta}$ will reduce the above result to the lemma version or the Slepian-Bangs formula. We refer to [43, pp. 141–147], [36, Appendix 15C], [42, p. 292] and [46, pp. 926–927] for single snapshot derivations of such a generalization. Indeed it can be easily modified to all N_T snapshots based on the mutually independent snapshots assumed.

where $\boldsymbol{\xi}_x(\boldsymbol{\theta}) = \upsilon_{\mathsf{c}}(\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}))$. *To fulfill the dimension of FIM, we will arrive at (see e.g., [44] and [45])*

$$\frac{1}{N_T}\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta}) = \nabla_{\theta}^{\mathsf{H}}(\boldsymbol{\xi}_x)\,\boldsymbol{\Psi}_{xx}^{-1}(\boldsymbol{\theta})\nabla_{\theta}(\boldsymbol{\xi}_x) \tag{2.42}$$

where $\boldsymbol{\Psi}_{xx}(\boldsymbol{\theta}) \triangleq \boldsymbol{\Sigma}_{xx}^{\mathsf{T}}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) \in \mathbb{C}_{\mathbb{H}}^{N_E^2 \times N_E^2}$.

Proof. In a separable formulation of $\upsilon_{\mathsf{c}}(\boldsymbol{\Sigma}_{xx})$, the relations $\lceil \mathbf{AB} \rfloor = \upsilon_{\mathsf{c}}^{\mathsf{H}}(\mathbf{A}^{\mathsf{H}})\,\upsilon_{\mathsf{c}}(\mathbf{B})$ and $\upsilon_{\mathsf{c}}(\mathbf{ABC}) = (\mathbf{C}^{\mathsf{T}} \otimes \mathbf{A})\upsilon_{\mathsf{c}}(\mathbf{B})$ allow the Slepian-Bangs formula to

$$\frac{1}{N_T}[\boldsymbol{I}_{\mathrm{F}}(\boldsymbol{\theta})]_{[n,\acute{n}]} = \upsilon_{\mathsf{c}}^{\mathsf{H}}\left(\frac{\partial}{\partial\theta_n}\boldsymbol{\Sigma}_{xx}(\theta_n)\right)(\boldsymbol{\Sigma}_{xx}^{-\mathsf{T}} \otimes \boldsymbol{\Sigma}_{xx}^{-1})\upsilon_{\mathsf{c}}\left(\frac{\partial}{\partial\theta_{\acute{n}}}\boldsymbol{\Sigma}_{xx}(\theta_{\acute{n}})\right). \tag{2.43}$$

By invoking the transformable geometry in $\upsilon_{\mathsf{c}}\left(\frac{\partial}{\partial \mathrm{x}_n}\mathbf{A}(\mathrm{x}_n)\right) = \frac{\partial}{\partial \mathrm{x}_n}\upsilon_{\mathsf{c}}(\mathbf{A}(\mathrm{x}_n))$, it is attainable to acquire such a variation of FIM formula. □

2.7 Toeplitz Covariance Estimations

As stated, sensor array processing deals, in general, with inferring the parameter of source signal based on the use of *u*niform *l*inear *a*rray (ULA) [4]. When there is no a priori knowledge of Toeplitz[4] structure, the sample covariance $\hat{\boldsymbol{\Sigma}}_{xx}$ computed from (2.21) is conceivable to be the unstructured maximum likelihood estimate of $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta})$ [52]. With phase reference at the first element, this enables array covariance matrix to Toeplitz structure. More precisely, the signal and the additive noise must also be stochastically uncorrelated in the classical (point source) model. As taken into account the multipath directions, the classical model is further argued to be unrealistic. However, the array covariance matrix still be possible to hold Toeplitz structure.

In this section, we shall explore the imposed Toeplitz-Hermitian structure in the array covariance matrix. The way to explore the Toeplitz structure in $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta})$ can be given by rewriting as $\bar{\boldsymbol{\Sigma}}_{xx}(\boldsymbol{\tau}) : \mathbb{R}^{(2N_E-1)\times 1} \mapsto \mathbb{C}_{\mathrm{T},\mathbb{H}}^{N_E \times N_E}$, where $\boldsymbol{\tau} \in \mathbb{R}^{(2N_E-1)\times 1}$ denotes the fundamental vector of Toeplitz-Hermitian matrix $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta})$. Notice that the actual covariance matrix satisfies $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}_{\circ}) = \bar{\boldsymbol{\Sigma}}_{xx}(\boldsymbol{\tau})$ and therefore can be expressed in Toeplitz-Hermitian form as

$$\bar{\boldsymbol{\Sigma}}_{xx}(\boldsymbol{\tau}) = \begin{bmatrix} \tau_0 & \tau_1^* & \cdots & \tau_{N_E-1}^* \\ \tau_1 & \tau_0 & \ddots & \tau_{N_E-2}^* \\ \vdots & \ddots & \ddots & \vdots \\ \tau_{N_E-1} & \tau_{N_E-2} & \cdots & \tau_0 \end{bmatrix} \tag{2.44}$$

[4]The name is after Otto Toeplitz (1881-1940), a German mathematician.

where the fundamental vector $\boldsymbol{\tau} \in \mathbb{R}^{(2N_E-1)\times 1}$ of all lower triangular entries of the Toeplitz-Hermitian matrix $\boldsymbol{\Sigma}_{xx}$ is

$$\boldsymbol{\tau} \triangleq \left[\tau_0 \quad \Re(\tau_1) \quad \Im(\tau_1) \quad \cdots \quad \Re(\tau_{N_E-1}) \quad \Im(\tau_{N_E-1})\right]^{\mathsf{T}}. \tag{2.45}$$

Let us introduce the binary selection of Toeplitz-Hermitian structure as the full column-rank matrix $\breve{\boldsymbol{\Xi}} \in \mathbb{B}_{\mathbb{F}}^{N_E^2\times(2N_E-1)}$:

$$\breve{\boldsymbol{\Xi}} \triangleq \left[\boldsymbol{\upsilon}_{\mathsf{c}}(\boldsymbol{I}) \quad \boldsymbol{\upsilon}_{\mathsf{c}}(\boldsymbol{L}_1) \quad \boldsymbol{\upsilon}_{\mathsf{c}}^{\mathsf{T}}(\boldsymbol{L}_1) \quad \cdots \quad \boldsymbol{\upsilon}_{\mathsf{c}}(\boldsymbol{L}_{N_E-1}) \quad \boldsymbol{\upsilon}_{\mathsf{c}}^{\mathsf{T}}(\boldsymbol{L}_{N_E-1})\right] \tag{2.46}$$

where $\boldsymbol{L}_{n_L}$ is the block-lower triangular matrix according to (2.26). The full-rank matrix $\boldsymbol{\Upsilon} \in \mathbb{C}_{\mathbb{F}}^{(2N_E-1)\times(2N_E-1)}$ is also introduced as a complex-valued selection matrix, defined by

$$\boldsymbol{\Upsilon} \triangleq \begin{bmatrix} 1 & \boldsymbol{0}^{\mathsf{T}} \\ \boldsymbol{0} & \boldsymbol{I}_{(N_E-1)} \otimes \begin{bmatrix} 1 & \imath \\ 1 & -\imath \end{bmatrix} \end{bmatrix} \tag{2.47}$$

where $\otimes$ signifies the Krönecker product operator. In the form of linearly affine structure, we can write

$$\bar{\boldsymbol{\xi}}_x(\boldsymbol{\tau}) = \breve{\boldsymbol{\Xi}}\,\boldsymbol{\Upsilon}\boldsymbol{\tau} \tag{2.48}$$

where $\bar{\boldsymbol{\xi}}_x(\boldsymbol{\tau}) \triangleq \boldsymbol{\upsilon}_{\mathsf{c}}(\bar{\boldsymbol{\Sigma}}_{xx}(\boldsymbol{\tau})) : \mathbb{R}^{(2N_E-1)\times 1} \mapsto \mathbb{C}^{N_E^2\times 1}$ designates the array covariance vectorization with $\boldsymbol{\upsilon}_{\mathsf{c}}(\cdot)$ designating the column-stacking vectorization operator.

2.7.1 Redundancy-Averaging Toeplitzification

One of the easiest ways to Toeplitzize the covariance matrix is to notice that all entries aligned in a subdiagonal of $\hat{\boldsymbol{\Sigma}}_{xx}$ must be equivalent. Averaging all $N_E - n_L$ redundant lags of $[\hat{\boldsymbol{\Sigma}}_{xx}]_{[n_L+n_l,n_l]}$, it attributes the covariance estimate according to *redundancy averaging*. The averaged result is then designated as the n_L-th Toeplitz lag $\hat{\tau}_{n_L} \in \mathbb{C}^{1\times 1}$ computed by [47]

$$\hat{\tau}_{n_L} = \frac{1}{N_E - n_L} \sum_{n_l=1}^{N_E-n_L} [\hat{\boldsymbol{\Sigma}}_{xx}]_{[n_L+n_l,n_l]}. \tag{2.49}$$

Keeping it into the fundamental vector $\hat{\boldsymbol{\tau}}_{\mathrm{RA}} \in \mathbb{R}^{(2N_E-1)\times 1}$ as

$$\hat{\boldsymbol{\tau}}_{\mathrm{RA}} = \left[\hat{\tau}_0 \quad \Re(\hat{\tau}_1) \quad \Im(\hat{\tau}_1) \quad \cdots \quad \Re(\hat{\tau}_{N_E-1}) \quad \Im(\hat{\tau}_{N_E-1})\right]^{\mathsf{T}} \tag{2.50}$$

then the Toeplitz-Hermitian covariance estimate is written as

$$\boldsymbol{\upsilon}_{\mathsf{c}}(\hat{\bar{\boldsymbol{\Sigma}}}_{xx}(\hat{\boldsymbol{\tau}}_{\mathrm{RA}})) = \breve{\boldsymbol{\Xi}}\,\boldsymbol{\Upsilon}\hat{\boldsymbol{\tau}}_{\mathrm{RA}}. \tag{2.51}$$

2.7.2 Weighted Covariance-Matching Toeplitzification

Let $\|\mathbf{A}\|_{\mathbf{W}}^2 \triangleq \boldsymbol{\upsilon}_{\mathsf{c}}^{\mathsf{H}}(\mathbf{A})\,\mathbf{W}^{-1}\boldsymbol{\upsilon}_{\mathsf{c}}(\mathbf{A})$ be a weighted version of the Euclidean norm, where $\mathbf{W}$ is a positive-definite Hermitian weight matrix. Based on the extended invariance principle, the WCM can be reformulated as [48]

$$\begin{aligned}\hat{\boldsymbol{\tau}}_{\text{WCM}} &= \arg\min_{\boldsymbol{\tau}} \|\bar{\boldsymbol{\Sigma}}_{xx}(\boldsymbol{\tau}) - \hat{\boldsymbol{\Sigma}}_{xx}\|_{\boldsymbol{W}}^2 \\ &= (\boldsymbol{\Upsilon}^{\mathsf{H}}\breve{\boldsymbol{\Xi}}^{\mathsf{T}}\boldsymbol{W}^{-1}\breve{\boldsymbol{\Xi}}\,\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}^{\mathsf{H}}\breve{\boldsymbol{\Xi}}^{\mathsf{T}}\boldsymbol{W}^{-1}\hat{\boldsymbol{\xi}}_x \\ &= \boldsymbol{\Upsilon}^{-1}(\breve{\boldsymbol{\Xi}}^{\mathsf{T}}\boldsymbol{W}^{-1}\breve{\boldsymbol{\Xi}})^{-1}\breve{\boldsymbol{\Xi}}^{\mathsf{T}}\boldsymbol{W}^{-1}\hat{\boldsymbol{\xi}}_x\end{aligned} \tag{2.52}$$

where $\hat{\boldsymbol{\xi}}_x = \boldsymbol{\upsilon}_{\mathsf{c}}(\hat{\boldsymbol{\Sigma}}_{xx}) \in \mathbb{C}^{N_E^2\times 1}$ and $\boldsymbol{W} \in \mathbb{C}_{\mathbb{H}}^{N_E^2\times N_E^2}$ are the covariance vectorization and Hermitian weight, respectively. One can see that the WCM loss function can be concentrated on $\boldsymbol{\Upsilon}$ by noticing that $\boldsymbol{\Upsilon}$ is of full rank. To make the residual $\tilde{\boldsymbol{\tau}}_{\text{WCM}} \triangleq \hat{\boldsymbol{\tau}}_{\text{WCM}} - \boldsymbol{\tau}_{\text{o}}$ minimal, the optimal weight should be satisfied by [35]

$$\boldsymbol{W} = \lim_{N_T\to\infty} N_T\mathcal{E}\langle\tilde{\boldsymbol{\xi}}_x\tilde{\boldsymbol{\xi}}_x^{\mathsf{H}}\rangle = \boldsymbol{\Sigma}_{xx}^{\mathsf{T}}\otimes\boldsymbol{\Sigma}_{xx} \tag{2.53}$$

where $\tilde{\boldsymbol{\xi}}_x \triangleq \hat{\boldsymbol{\xi}}_x - \boldsymbol{\xi}_x$ is the the sample covariance residual. Since the exact weight depends itself on the model parameter, it is preferable to make use of nonparametric estimate $\hat{\boldsymbol{W}} = \hat{\boldsymbol{\Sigma}}_{xx}^{\mathsf{T}}\otimes\hat{\boldsymbol{\Sigma}}_{xx}$ rather than $\boldsymbol{W}$ without loss of asymptotic performance. Therefore, the WCM covariance estimate $\hat{\bar{\boldsymbol{\Sigma}}}_{xx}(\hat{\boldsymbol{\tau}}_{\text{WCM}})$ can be devectorized from [49]

$$\boldsymbol{\upsilon}_{\mathsf{c}}(\hat{\bar{\boldsymbol{\Sigma}}}_{xx}(\hat{\boldsymbol{\tau}}_{\text{WCM}})) \cong \breve{\boldsymbol{\Xi}}(\breve{\boldsymbol{\Xi}}^{\mathsf{T}}\hat{\boldsymbol{W}}^{-1}\breve{\boldsymbol{\Xi}})^{-1}\breve{\boldsymbol{\Xi}}^{\mathsf{T}}\hat{\boldsymbol{W}}^{-1}\hat{\boldsymbol{\xi}}_x. \tag{2.54}$$