On ordered normally distributed vector parameter estimates

Eric Chaumette, François Vincent, Olivier Besson

On ordered normally distributed vector parameter estimates

Eric Chaumette, François Vincent, Olivier Besson

University of Toulouse-ISAE, Department of Electronics, Optronics and Signal, 10 Avenue Edouard Belin, 31055 Toulouse, France

Keywords:
Multivariate normal distribution
Order statistics
Mean square error
Maximum likelihood estimator

A B S T R A C T

The ordered values of a sample of observations are called the order statistics of the sample and are among the most important functions of a set of random variables in probability and statistics. However the study of ordered estimates seems to have been overlooked in maximum-likelihood estimation. Therefore it is the aim of this communication to give an insight into the relevance of order statistics in maximum-likelihood estimation by providing a second-order statistical prediction of ordered normally distributed estimates. Indeed, this second-order statistical prediction allows to refine the asymptotic performance analysis of the mean square error (MSE) of maximum likelihood estimators (MLEs) of a subset of the parameters. A closer look to the bivariate case highlights the possible impact of estimates ordering on MSE, impact which is not negligible in (very) high resolution scenarios.

1. Introduction

The ordered values of a sample of observations are called the order statistics of the sample: if $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_M)^T$ is a vector[1] of M real valued random variables, then $\boldsymbol{\theta}_{(M)} = (\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(M)})^T$ denotes the vector of order statistics induced by $\boldsymbol{\theta}$ where $\theta_{(1)} \le \theta_{(2)} \le \cdots \le \theta_{(M)}$ [1,2]. Order statistics and extremes (smallest and largest values) are among the most important functions of a set of random variables in probability and statistics. There is natural interest in studying the highs and lows of a sequence, and the other order statistics help in understanding the concentration of probability in a distribution, or equivalently, the diversity in the population represented by the distribution. Order statistics are also useful in statistical inference, where estimates of parameters are often based on some suitable functions of the order statistics vector (robust location estimates, detection of outliers, censored sampling, characterizations, goodness of fit, etc.). However the study of ordered estimates seems to have been overlooked in maximum-likelihood estimation [4], which is at first sight a little bit surprising. Indeed, if $\mathbf{x}$ denotes the random observation vector and $p(\mathbf{x}; \boldsymbol{\Theta})$ denotes the probability density function (p.d.f.) of $\mathbf{x}$ depending on a vector of P real unknown parameters $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_P)$ to be estimated, then many estimation problems in this setting lead to estimation algorithms yielding ordered estimates $\widehat{\boldsymbol{\theta}}_{(M)}$ induced by a vector $\widehat{\boldsymbol{\theta}}$ of M estimates formed from a subset of the whole set of P estimates $\widehat{\boldsymbol{\Theta}}$. Among all various possible instances of this setting, the most studied in signal processing is that of separating the components of data formed from a linear superposition of individual signals and noise (nuisance). For the sake of illustration, let us consider the following simplified

E-mail addresses: eric.chaumette@isae.fr (E. Chaumette), francois.vincent@isae.fr (F. Vincent), olivier.besson@isae.fr (O. Besson).

[1] The notational convention adopted is as follows: italic indicates a scalar quantity, as in a; lower case boldface indicates a column vector quantity, as in $\mathbf{a}$; upper case boldface indicates a matrix quantity, as in $\mathbf{A}$. The n-th row and m-th column element of the matrix $\mathbf{A}$ will be denoted by $A_{n,m}$ or $(\mathbf{A})_{n,m}$. The n-th coordinate of the column vector $\mathbf{a}$ will be denoted by a_n or $(\mathbf{a})_n$. The matrix/vector transpose is indicated by a superscript T as in $\mathbf{A}^T$. For two vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \ge \mathbf{b}$ means that $\mathbf{a}-\mathbf{b}$ is positive componentwise. $\mathcal{M}_{\mathbb{R}}(N,P)$ denotes the vector space of real matrices with N rows and P columns. $\mathbf{1}_M^{m:m+l}$ denotes the M-dimensional vector with all components set to 0 except components from m to $m+l$ set to 1. $\mathbf{1}_M$ denotes the M-dimensional vector with all components set to 1. $\mathbf{I}_M \in \mathbb{R}^{M \times M}$ denotes the identity matrix. $1_{\{A\}}$ denotes the indicator function of the event A. $E_{\boldsymbol{\Theta}}[\mathbf{g}(\mathbf{x})] = \int \mathbf{g}(\mathbf{x}) p(\mathbf{x}; \boldsymbol{\Theta})\, d\mathbf{x}$ denotes the statistical expectation of the vector of functions $\mathbf{g}(\,)$ with respect to $\mathbf{x}$ parameterized by $\boldsymbol{\theta}$.

http://dx.doi.org/10.1016/j.sigpro.2015.02.026

example:

$$\mathbf{x}_t(\boldsymbol{\Theta}) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}_t + \mathbf{n}_t, \quad \boldsymbol{\Theta}^T = \left(\boldsymbol{\theta}^T, \mathbf{s}_1^T, \ldots, \mathbf{s}_T^T\right)^T \tag{1}$$

where $1 \le t \le T$, T is the number of independent observations, $\mathbf{x}_t$ is the vector of samples of size N, M is the number of signal sources, $\mathbf{s}_t$ is the vector of complex amplitudes of the M sources for the tth observation, $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_M)]$ and $\mathbf{a}(\ )$ is a vector of N parametric functions depending on a single parameter θ, $\mathbf{n}_t$ are Gaussian complex circular noises independent of the M sources. Since (1) is invariant over permutation of signal sources, i.e. for any permutation matrix $\mathbf{P}_i \in \mathbb{R}^{M \times M}$:

$$\mathbf{x}_t(\boldsymbol{\Theta}) = (\mathbf{A}(\boldsymbol{\theta})\mathbf{P}_i)(\mathbf{P}_i\mathbf{s}_t) + \mathbf{n}_t,$$

it is well known that (1) is an ill-posed unidentifiable estimation problem which can be regularized, i.e. transformed into a well-posed and identifiable estimation problem, by imposing the ordering of the unknown parameters θ_m: $\boldsymbol{\theta} \triangleq (\theta_1, \ldots, \theta_M), \theta_1 < \cdots < \theta_M$, and of their estimates as well: $\widehat{\boldsymbol{\theta}} \triangleq \widehat{\boldsymbol{\theta}}_{(M)}$. Therefore in the MSE sense, the correct statistical prediction is given by the computation of $E_{\boldsymbol{\Theta}}[(\widehat{\theta}_{(m)} - \theta_m)^2]$, $1 \le m \le M$. Unfortunately, the correct statistical prediction cannot be obtained from scratch since most of the available results in the open literature on order statistics [1–3] have been derived the other way round, i.e. they request the knowledge of the distribution of $\widehat{\boldsymbol{\theta}}$. The distribution of $\widehat{\boldsymbol{\theta}}$ can be obtained from a priori information on the problem at hand or may have been derived in some regions of operation of the observation model. For instance, it has been known for a while that, under reasonably general conditions on the observation model [4,7], the ML estimates are asymptotically Gaussian distributed when the number of independent observation tends to infinity. Additionally, if the observation model is linear Gaussian as in (1), some additional asymptotic regions of operation yielding Gaussian MLEs have also been identified: at finite number of independent observations [5–9] or when the number of samples and the number of independent observations increase without bound at the same rate, i.e. $N, T \to \infty$, $N/T \to c$, $0 < c < 1$ [10]. Nevertheless a close look at the derivations of these results reveals an implicit hypothesis: the asymptotic condition of operation considered yields resolvable estimates [11,12], what prevents from estimates re-ordering. Therefore, under this implicit hypothesis $\widehat{\boldsymbol{\theta}}_{(M)} = \widehat{\boldsymbol{\theta}}$. However when the condition of operation degrades, distribution spread and/or location bias of each $\widehat{\theta}_m$ increase and the hypothesis of resolvable estimates does not hold any longer yielding observation samples for which $\widehat{\boldsymbol{\theta}}_{(M)} \ne \widehat{\boldsymbol{\theta}}$ [11,12].

Therefore it is the aim of this communication to give an insight into the relevance of order statistics in maximum-likelihood estimation by providing a second-order statistical prediction of ordered normally distributed estimates. This second-order statistical prediction allows to refine the asymptotic performance analysis of the MSE of MLEs of a subset $\boldsymbol{\theta}$ of the parameters set $\boldsymbol{\Theta}$.[2] Indeed, in the setting of a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}$ and covariance matrix $\mathbf{C}_{\widehat{\boldsymbol{\theta}}}$, $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}_M\left(\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}, \mathbf{C}_{\widehat{\boldsymbol{\theta}}}\right)$, with p.d.f. denoted $p_{\mathcal{N}_M}(\widehat{\boldsymbol{\theta}}; \boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}, \mathbf{C}_{\widehat{\boldsymbol{\theta}}})$, the most general statistical characterization, i.e. including distribution and moments, have been derived for an exchangeable multivariate normal random vector [1,13,14], that is a normal distribution with a common mean μ, a common variance σ^2 and a common correlation coefficient ρ: $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}_M\left(\mu \mathbf{1}_M, \sigma^2\left((1-\rho)\mathbf{I}_M + \rho \mathbf{1}_M \mathbf{1}_M^T\right)\right)$, with $\rho \in [0, 1[$. If the focus is on distribution, then the most general result has been released recently in [3] where the exact distribution of linear combinations of order statistics (L-statistics) [15] of arbitrary dependent random variables has been derived (see also [16] for the joint distribution of order statistics in a set of univariate or bivariate observations). In particular, [3] examines the case where the random variables have a joint elliptically contoured distribution and the case where the random variables are exchangeable. Arellano-Vallea and Genton [3] investigate also the particular L-statistics that simply yield a set of order statistics, and study their joint distribution. Unfortunately, general derivations of closed form expressions for moments and cumulants of L-statistics were beyond the scope of [3] and were left for future research. However in the particular case of a multivariate normal distribution, it is possible to obtain closed forms for first and second order moments of its order statistics directly, i.e. without explicitly computing the order statistics distribution (see Section 2). These closed forms not only generalize the earlier work from the exchangeable case to the general case providing a second-order statistical prediction of L-statistics from multivariate normal distribution but are also required to characterize the MSE of normally distributed vector parameter estimates. Indeed, since it is always assumed that $\boldsymbol{\theta}$ has distinct components, any sensible estimation technique of $\boldsymbol{\theta}$ must preserve this resolvability requirement and yield distinct mean values, leading to asymptotically non-exchangeable multivariate normal random vector.

2. Second-order statistical prediction of ordered normally distributed estimates

First, note that $\widehat{\boldsymbol{\theta}}_{(M)} \in \mathrm{Per}(\widehat{\boldsymbol{\theta}})$, where $\mathrm{Per}(\widehat{\boldsymbol{\theta}}) = \{\widehat{\boldsymbol{\theta}}_i = \mathbf{P}_i\widehat{\boldsymbol{\theta}};\ i = 1, \ldots, M!\}$ is the collection of random vectors $\widehat{\boldsymbol{\theta}}_i$ corresponding to the $M!$ different permutations of the components of $\widehat{\boldsymbol{\theta}}$. Here $\mathbf{P}_i \in \mathbb{R}^{M \times M}$ are permutation matrices with $\mathbf{P}_i \ne \mathbf{P}_j$ for all $i \ne j$. Let $\boldsymbol{\Delta} \in \mathbb{R}^{(M-1) \times M}$ be the difference matrix such that $\boldsymbol{\Delta}\boldsymbol{\theta} = (\theta_2 - \theta_1, \theta_3 - \theta_2, \ldots, \theta_M - \theta_{M-1})^T$, i.e., the mth row of $\boldsymbol{\Delta}$ is $\mathbf{d}_{m+1}^T - \mathbf{d}_m^T$, $m = 1, \ldots, M-1$, where $\mathbf{d}_1, \ldots, \mathbf{d}_M$ are the M-dimensional unit basis vectors. Let $\mathcal{S}_i = \{\widehat{\boldsymbol{\theta}}: \boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i \ge \mathbf{0}\}$ where $\widehat{\boldsymbol{\theta}}_i \sim \mathcal{N}_M(\boldsymbol{\mu}_i, \mathbf{C}_i)$, $\boldsymbol{\mu}_i = \mathbf{P}_i\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}$, $\mathbf{C}_i = \mathbf{P}_i\ \mathbf{C}_{\widehat{\boldsymbol{\theta}}}\mathbf{P}_i^T$. Let $\mathcal{P}(\mathcal{D})$ be the probability of an event $\mathcal{D}$. As the set of events $\{\mathcal{S}_i\}_{i=1}^{M!}$ is a partition of $\mathbb{R}^M$, whatever the real valued function $f(\ ,\ )$, by the theorem of total probability we have

$$E\left[f\left(\widehat{\theta}_{(m)}, \widehat{\theta}_{(l)}\right)\right] = \sum_{i=1}^{M!} E\left[f\left(\widehat{\theta}_{(m)}, \widehat{\theta}_{(l)}\right) | \mathcal{S}_i\right] \mathcal{P}(\mathcal{S}_i)$$

[2] Note that these results are also applicable to other estimators, such as M-estimators, Bayesian estimators (MAP, MMSE), as long as their distribution is normal.

that is

$$E\left[f\left(\widehat{\theta}_{(m)},\widehat{\theta}_{(l)}\right)\right]=\sum_{i=1}^{M!}E\left[f\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_m,\left(\widehat{\theta}_i\right)_l\right)|\mathcal{S}_i\right]\mathcal{P}(\mathcal{S}_i). \tag{2}$$

However, from a computational point of view, it is wiser to express (2) as

$$E\left[f\left(\widehat{\theta}_{(m)},\widehat{\theta}_{(l)}\right)\right]=\sum_{i=1}^{M!}E\left[f\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_m,\left(\widehat{\theta}_i\right)_l\right)|\mathcal{U}_i\right]\mathcal{P}(\mathcal{U}_i) \tag{3}$$

where $\mathcal{U}_i=\{\widehat{\mathbf{u}}_i:\widehat{\mathbf{u}}_i\ge-\boldsymbol{\Delta\mu}_i\}$ and $\mathbf{u}_i=\boldsymbol{\Delta}\left(\widehat{\boldsymbol{\theta}}_i-\boldsymbol{\mu}_i\right)\sim\mathcal{N}_{M-1}\left(\mathbf{0},\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T\right)$. Then a smart exploitation of (3) (see Appendix for details) yields

$$E\left[\widehat{\theta}_{(m)}\right]=\sum_{i=1}^{M!}\left(\alpha_{m|i}\mathcal{P}_i+\boldsymbol{\beta}_{m|i}^T\mathbf{e}_i\right) \tag{4}$$

$$E\left[\widehat{\theta}_{(m)}^2\right]=\sum_{i=1}^{M!}\left(\left(\sigma_{m|i}^2+\alpha_{m|i}^2\right)\mathcal{P}_i+2\alpha_{m|i}\boldsymbol{\beta}_{m|i}^T\mathbf{e}_i+\boldsymbol{\beta}_{m|i}^T\mathbf{R}_i\boldsymbol{\beta}_{m|i}\right) \tag{5}$$

$$E\left[\widehat{\theta}_{(m+l)}\widehat{\theta}_{(m)}\right]=\frac{1}{2}\left(E\left[\widehat{\theta}_{(m+l)}^2\right]+E\left[\widehat{\theta}_{(m)}^2\right]-E\left[\left(\widehat{\theta}_{(m+l)}-\widehat{\theta}_{(m)}\right)^2\right]\right) \tag{6}$$

$$E\left[\left(\widehat{\theta}_{(m+l)}-\widehat{\theta}_{(m)}\right)^2\right]=\left(\mathbf{1}_{M-1}^{m:m+l-1}\right)^T\left(\sum_{i=1}^{M!}\begin{pmatrix}\boldsymbol{\Delta\mu}_i(\boldsymbol{\Delta\mu}_i)^T\mathcal{P}_i+\\ 2\mathbf{e}_i(\boldsymbol{\Delta\mu}_i)^T+\mathbf{R}_i\end{pmatrix}\right)\mathbf{1}_{M-1}^{m:m+l-1} \tag{7}$$

$$\mathcal{P}_i=\mathcal{P}(\mathcal{U}_i),\quad \mathbf{e}_i=E\left[\mathbf{u}_i1_{\mathcal{U}_i}\right],\quad \mathbf{R}_i=E\left[\mathbf{u}_i(\mathbf{u}_i)^T1_{\mathcal{U}_i}\right] \tag{8}$$

where $l\ge1$, $\alpha_{m|i}=\mathbf{d}_m^T\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_i=\mathbf{P}_i\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}$, $\mathbf{C}_i=\mathbf{P}_i\mathbf{C}_{\widehat{\boldsymbol{\theta}}}P_i^T$, $\boldsymbol{\beta}_{m|i}=(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T)^{-1}\boldsymbol{\Delta}\mathbf{C}_i\mathbf{d}_m$, $\sigma_{m|i}^2=\mathbf{d}_m^T(\mathbf{C}_{\widehat{\boldsymbol{\theta}}}-\mathbf{C}_i\boldsymbol{\Delta}^T\left(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T\right)^{-1}\ \boldsymbol{\Delta}\mathbf{C}_i)\ \mathbf{d}_m$.

Finally, the second order statistical prediction of any L-statistics, that is linear combinations of the vector of order statistics, $\widehat{\mathbf{z}}=\mathbf{L}\widehat{\boldsymbol{\theta}}_{(M)}$, $\mathbf{L}\in\mathcal{M}_{\mathbb{R}}(N,M)$, is given by

$$\boldsymbol{\mu}_{\widehat{\mathbf{z}}}=\mathbf{L}\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}_{(M)}},\quad \mathbf{C}_{\widehat{\mathbf{z}}}=\mathbf{L}\mathbf{C}_{\widehat{\boldsymbol{\theta}}_{(M)}}\mathbf{L}^T,\quad \left|\begin{array}{l}\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}_{(M)}}=E\left[\widehat{\boldsymbol{\theta}}_{(M)}\right]\\ \mathbf{C}_{\widehat{\boldsymbol{\theta}}_{(M)}}=E\left[\widehat{\boldsymbol{\theta}}_{(M)}\widehat{\boldsymbol{\theta}}_{(M)}^T\right]-E\left[\widehat{\boldsymbol{\theta}}_{(M)}\right]E\left[\widehat{\boldsymbol{\theta}}_{(M)}\right]^T\end{array}\right.$$

and can be computed from expressions (4)–(7) of $E\left[\widehat{\theta}_{(m)}\right]$, $E\left[\widehat{\theta}_{(m+l)}\widehat{\theta}_{(m)}\right]$, $E\left[\widehat{\theta}_{(m)}^2\right]$ when the L-statistics derive from multivariate normal distribution.

2.1. *Cramér–Rao bound for ordered normally distributed estimates*

Let $\mathbf{CRB}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ denote the Cramér–Rao bound (CRB) for unbiased estimates of $\boldsymbol{\theta}$ [4,7]. Then (4) allows for the computation of the ordered estimates bias $\left(\overline{\mathbf{b}}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\right)_m=E_{\Theta}\left[\widehat{\theta}_{(m)}\right]-\theta_m$ which may be different from the (a priori) theoretical bias $(\mathbf{b}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}))_m=E_{\boldsymbol{\Theta}}\left[\widehat{\theta}_m\right]-\theta_m$, leading to the CRB for ordered estimates given by [4,7]

$$\mathbf{CRB}_{\boldsymbol{\Theta}}^{\overline{\mathbf{b}}}(\boldsymbol{\theta})=\overline{\mathbf{b}}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\overline{\mathbf{b}}_{\boldsymbol{\Theta}}^T(\boldsymbol{\theta})+\left(\mathbf{I}_M+\frac{\partial\overline{\mathbf{b}}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right)$$

$$\times\,\mathbf{CRB}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\left(\mathbf{I}_M+\frac{\partial\overline{\mathbf{b}}_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right)^T, \tag{9}$$

which may be different from the (a priori) theoretical $\mathbf{CRB}_{\boldsymbol{\Theta}}^{\mathbf{b}}(\boldsymbol{\theta})$.

2.2. *Implementation*

Let $\mathbf{D}_i$ be the diagonal matrix such that $(\mathbf{D}_i)_{m,m}=\sqrt{(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T)_{m,m}}$, $m=1,\dots,M-1$. The computational cost of (4)–(7) can be reduced by reformulating $\mathcal{P}_i$, $\mathbf{e}_i$, $\mathbf{R}_i$ (8) as

$$\mathcal{P}_i=\mathcal{P}(\mathcal{V}_i),\quad \mathbf{e}_i=\mathbf{D}_iE\left[\widehat{\mathbf{v}}_i|\mathcal{V}_i\right],\quad \mathbf{R}_i=\mathbf{D}_iE\left[\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^T|\mathcal{V}_i\right]\mathbf{D}_i, \tag{10}$$

where $\mathcal{V}_i=\left\{\widehat{\mathbf{v}}_i:\widehat{\mathbf{v}}_i\ge-\mathbf{D}_i^{-1}\boldsymbol{\Delta\mu}_i\right\}$ and $\widehat{\mathbf{v}}_i\sim\mathcal{N}_{M-1}(\mathbf{0},\mathbf{D}_i^{-1}(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T)\mathbf{D}_i^{-1})$ is a vector of correlated standard normal random variables satisfying $\mathcal{P}(|(\widehat{\mathbf{v}}_i)_m|\ge10)<10^{-23}$. Thus, from a numerical point of view, the distribution support of each $(\widehat{\mathbf{v}}_i)_m$ can be restricted to the interval $[-10,10]$ leading to $|E[(\widehat{\mathbf{v}}_i)_m|\mathcal{V}_i]|\le10\mathcal{P}(\mathcal{V}_i)$ and $|E[(\widehat{\mathbf{v}}_i)_m(\widehat{\mathbf{v}}_i)_l|\mathcal{V}_i]|\le10^2\mathcal{P}(\mathcal{V}_i)$. Therefore, since

$$\mathcal{P}(\mathcal{V}_i)\le\min_{1\le m\le M-1}\left\{\mathcal{P}\left((\widehat{\mathbf{v}}_i)_m\ge-\left(\mathbf{D}_i^{-1}\boldsymbol{\Delta\mu}_i\right)_m\right)\right\},$$

from a numerical point of view:

- if $\exists m\in[1,M-1]|\left(\mathbf{D}_i^{-1}\boldsymbol{\Delta\mu}_i\right)_m\le-10$, then:
  $\mathcal{P}(\mathcal{V}_i)=0,\quad E\left[\widehat{\mathbf{v}}_i|\mathcal{V}_i\right]=\mathbf{0},\quad E\left[\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^T|\mathcal{V}_i\right]=\mathbf{0},$
- if $\forall m\in[1,M-1]$, $\left(\mathbf{D}_i^{-1}\boldsymbol{\Delta\mu}_i\right)_m\ge10$, then:
  $\mathcal{P}(\mathcal{V}_i)=1,\quad E\left[\widehat{\mathbf{v}}_i|\mathcal{V}_i\right]=E\left[\widehat{\mathbf{v}}_i\right]=\mathbf{0},$
  $ptE\left[\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^T|\mathcal{V}_i\right]=E\left[\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^T\right]=\mathbf{D}_i^{-1}\left(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T\right)\mathbf{D}_i^{-1}.$

In any other case, $\mathcal{P}(\mathcal{V}_i)$, $E\left[\widehat{\mathbf{v}}_i|\mathcal{V}_i\right]$ and $E\left[\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i^T|\mathcal{V}_i\right]$ can be computed by resorting to algorithms proposed by Genz [17] for numerical evaluation of multivariate normal distributions and moments over domains included in $[-10,10]^M$.

3. The bivariate (two sources) case

Let $\mathbf{C}\triangleq\mathbf{C}_{\widehat{\boldsymbol{\theta}}}$, $\boldsymbol{\mu}\triangleq\boldsymbol{\mu}_{\widehat{\boldsymbol{\theta}}}$, $d\mu=\mu_2-\mu_1$, and $d\widehat{\theta}=u_1=\widehat{\theta}_2-\widehat{\theta}_1$. As M=2 then $u_2=-u_1$, $\boldsymbol{\Delta\mu}_1=d\mu=-\boldsymbol{\Delta\mu}_2$, $\mathcal{P}_1+\mathcal{P}_2=1$, $\mathbf{1}_1^{1:1}=1$, $e_1-e_2=E[u_1]=0$, $R_1+R_2=E[u_1^2]=\sigma_{d\widehat{\theta}}^2$, $\boldsymbol{\Delta}\mathbf{C}_1\boldsymbol{\Delta}^T=\boldsymbol{\Delta}\mathbf{C}_2\boldsymbol{\Delta}^T=\sigma_{d\widehat{\theta}}^2=C_{1,1}+C_{2,2}-2C_{1,2}$. Moreover, if $\mathbf{C}=\sigma^2((1-\rho)\mathbf{I}_2+\rho\mathbf{1}_2\mathbf{1}_2^T)$, then $\sigma_{d\widehat{\theta}}=\sqrt{2\sigma^2(1-\rho)}$ and a few additional calculations yield

$$E\left[\widehat{\theta}_{(m)}\right]=E\left[\widehat{\theta}_m\right]+(-1)^mh(\tau)\sigma_{d\widehat{\theta}},$$

$$E\left[\widehat{\theta}_{(m)}^2\right]=E\left[\widehat{\theta}_m^2\right]+(-1)^mh(\tau)(\mu_2+\mu_1)\sigma_{d\widehat{\theta}},$$

$$\mathrm{Var}\left[\widehat{\theta}_{(m)}\right]=\mathrm{Var}\left[\widehat{\theta}_m\right]-\sigma_{d\widehat{\theta}}^2h(\tau)(\tau+h(\tau))$$

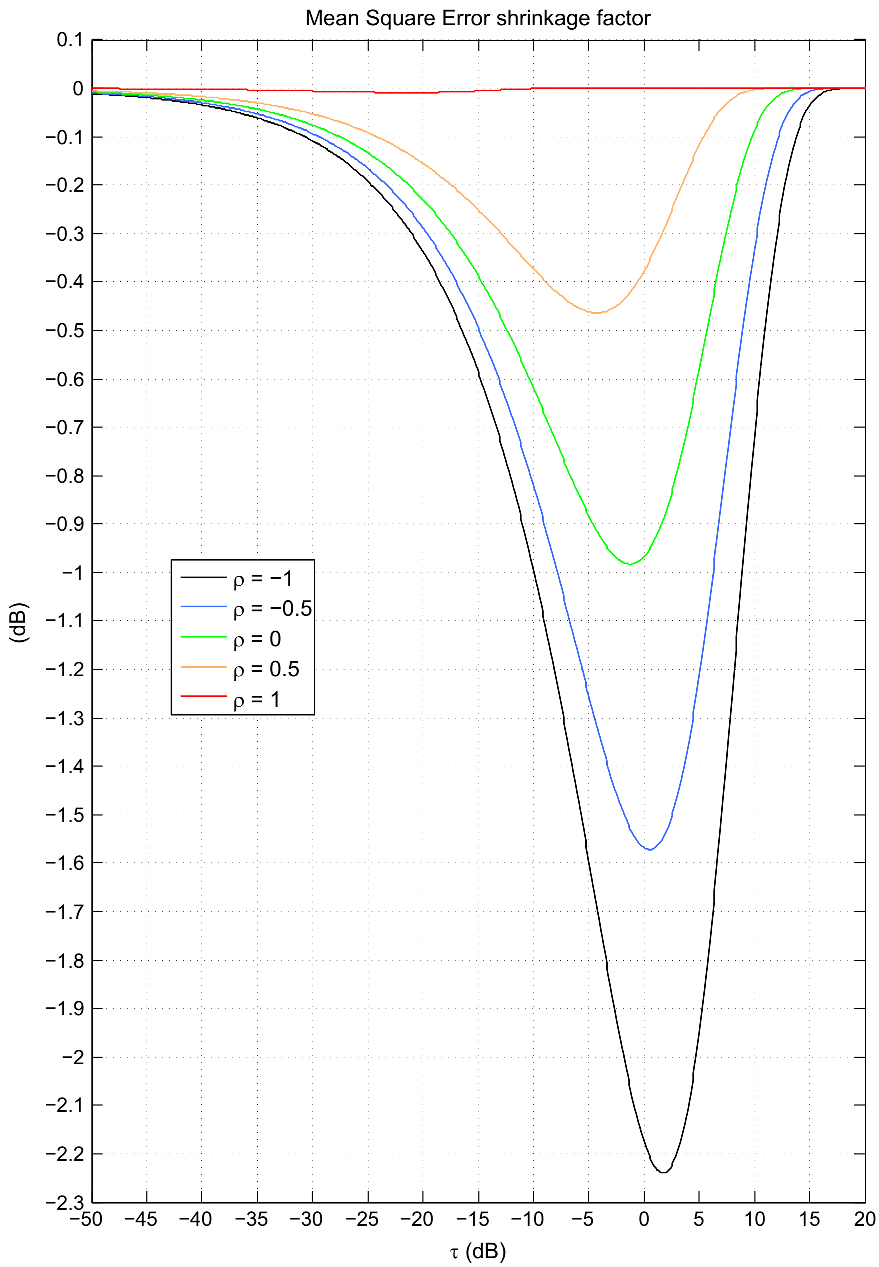

Fig. 1. MSE shrinkage factor $\xi(\tau,\rho)$ versus $\tau = d\mu/\sigma_{\widehat{d\theta}}$ for $\rho = -1, -0.5, 0, 0.5, 1$.

$$\mathrm{MSE}\left[\widehat{\theta}_{(m)}\right] = \mathrm{Var}\left[\widehat{\theta}_m\right] - \sigma^2_{\widehat{d\theta}}\tau h(\tau), \tag{11}$$

where $\tau = d\mu/\sigma_{\widehat{d\theta}}$, $h(y) = E\left[v1_{\{v \geq y\}}\right] - y\mathcal{P}(v \geq y)$ and $\mathrm{MSE}\left[\widehat{\theta}_{(m)}\right] \triangleq E\left[\left(\widehat{\theta}_{(m)} - \mu_m\right)^2\right]$. An interesting feature is the MSE shrinkage factor:

$$\xi(\tau,\rho) = \frac{\mathrm{MSE}\left[\widehat{\theta}_{(m)}\right]}{\mathrm{Var}\left[\widehat{\theta}_m\right]} = 1 - 2(1-\rho)\tau h(\tau). \tag{12}$$

The computation of the derivatives vector $\partial\xi(\tau,\rho)/\partial(\tau,\rho)$ shows that, for a given ρ, $\xi(\tau,\rho)$ admits a global minimum which is the solution of $E\left[v1_{\{v \geq y\}}\right] - 2y\mathcal{P}(v \geq y) = 0$, $y = \tau/\sqrt{2(1-\rho)}$, that is, after a numerical resolution, $y \simeq 0.61$ and leading to $\tau \simeq 0.863\sqrt{1-\rho}$, $\xi(\tau,\rho) \geq 1 - 0.2(1-\rho) \geq 0.6$ (see Fig. 1). Therefore the minimum of $\xi(\tau,\rho)$ is 0.6 (-2.2 dB) and is reached for couples (τ,ρ) belonging to the set $\{(\tau, \rho \to -1^{+}) | \tau \simeq 1.22\}$. This result is applicable to any instantiation of (1) for which $\widehat{\boldsymbol{\theta}}$ is normal with $\mathbf{C}_{\widehat{\boldsymbol{\theta}}} = \sigma^2\left((1-\rho)\mathbf{I}_2 + \rho\mathbf{1}_2\mathbf{1}_2^T\right)$, as for example, the asymptotic

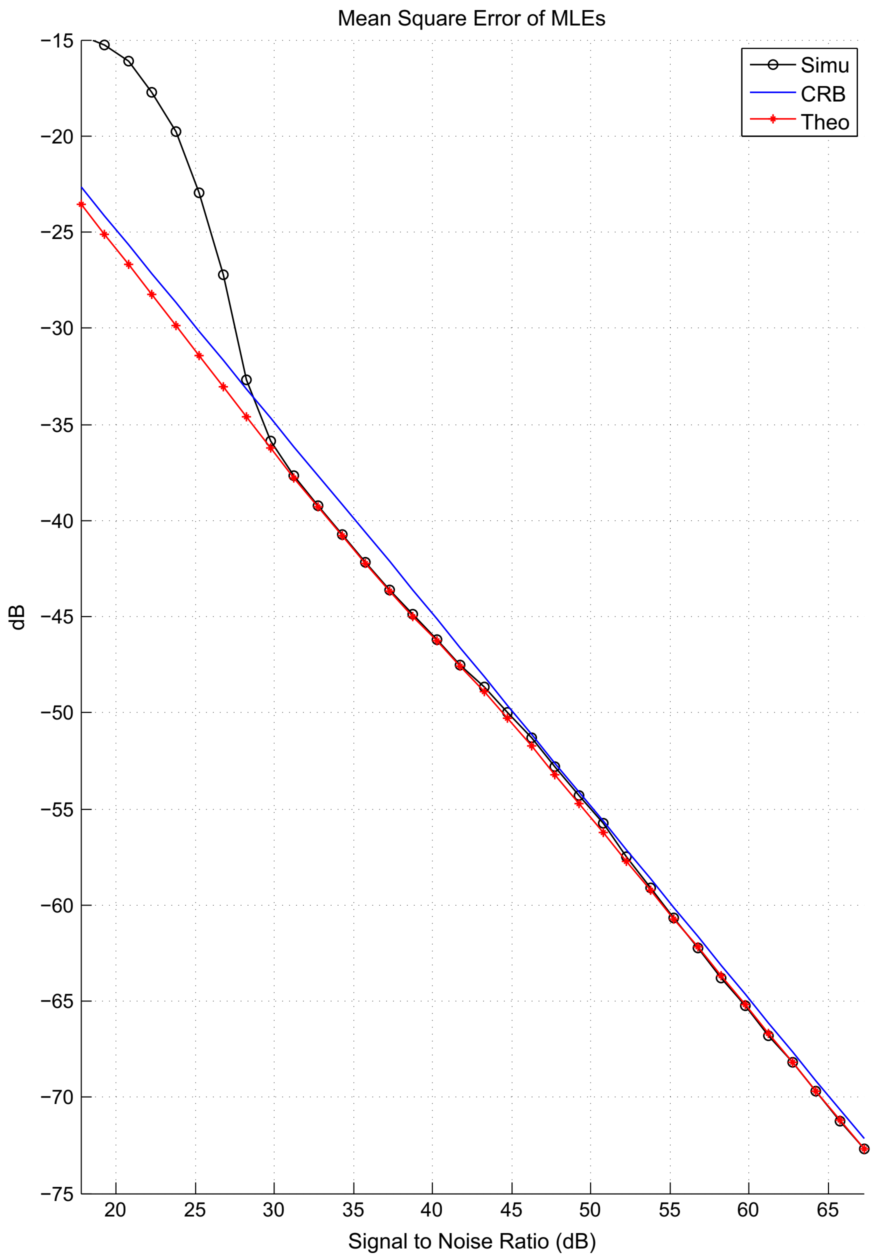

Fig. 2. Average empirical and theoretical MSE (11) of order MLEs of two closely spaced frequencies versus SNR.

behaviour of the MLEs of the direction of arrival (DoA) of two sources of equal power impinging on a uniform linear array (ULA) with symmetric DoAs relative to boresight, or two cisoids (tones) of equal power with opposite frequencies, whatever the observation model is conditional or unconditional [5]. Indeed in these two cases $\mathbf{CRB}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \sigma^2\left((1-\rho)\mathbf{I}_2+\rho\mathbf{1}_2\mathbf{1}_2^T\right)$ [4,5,7] and a priori, a violation of the CRB not attributable to the variance of the empirical MSE is expected in some scenarios, unless the ad hoc CRB (9) is computed. However, a closer look to Fig. 1 reveals that $\tau = d\mu/\sigma_{\widehat{d\theta}}$ must be inferior to $10^{15/20} \simeq 6$ in order to exhibit a measurable MSE shrinkage effect. As the normality of estimates and convergence to the CRB is obtained only in asymptotic conditions where $\sigma_{\widehat{d\theta}} \ll 1$, this means that $d\theta = d\mu \ll 1$ as well, that is the scenario considered is a (very) high resolution scenario. To highlight the possible impact of estimates ordering on MSE in (very) high resolution scenarios, we consider the estimation of two tones ($M=2$) of equal power with opposite frequencies where the observation model (1) is deterministic: $\mathbf{a}(\theta)^T = (1, e^{j2\pi\theta}, \ldots, e^{j2\pi(N-1)\theta})$, $N=8$, $T=2$, $d\theta = 1/12N$, $\mathbf{C}_{\mathbf{n}_t} = \mathbf{I}_2$, $\mathbf{C}_{\mathbf{s}_t} = (SNR/N)\ \left((1+\frac{1}{8})\mathbf{I}_2 - \frac{1}{8}\mathbf{1}_2\mathbf{1}_2^T\right)$ where SNR is the signal to noise ratio measured at the

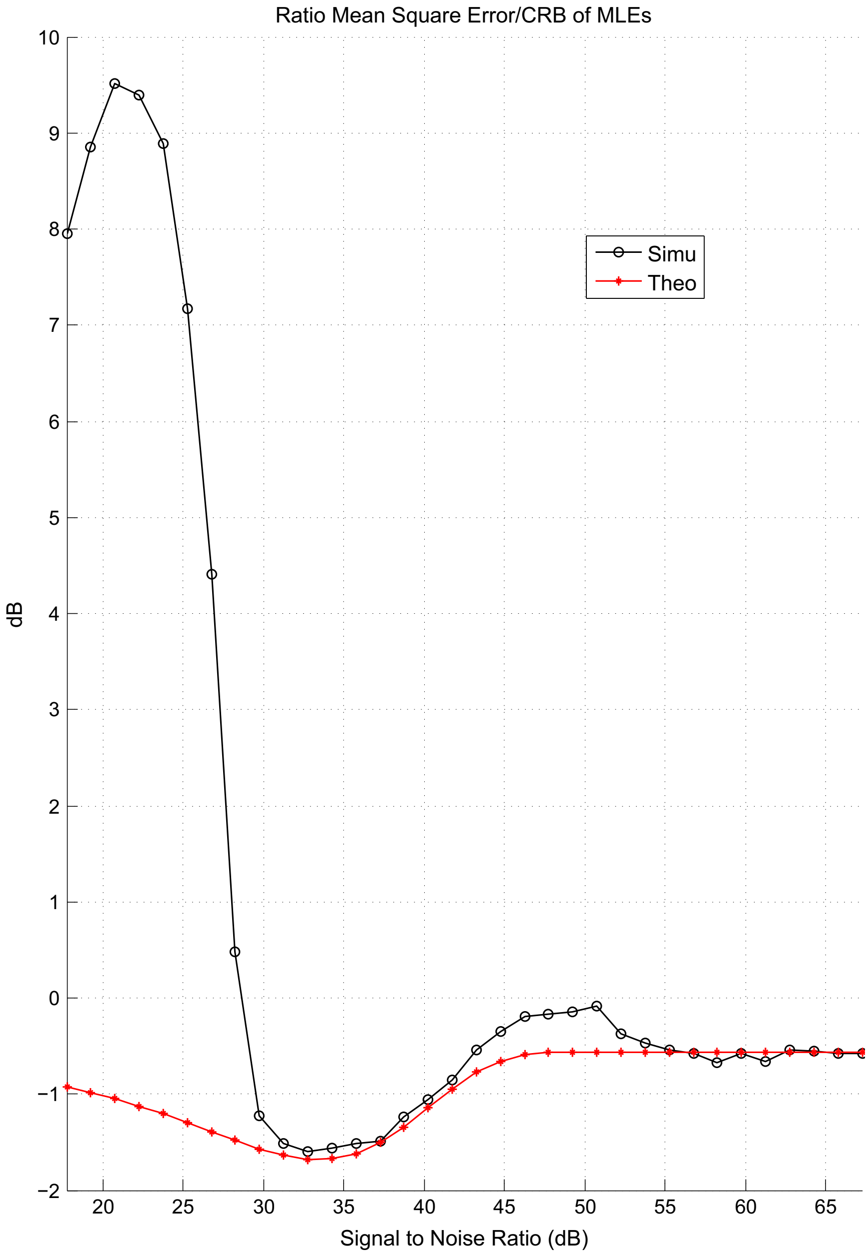

Fig. 3. Average empirical and theoretical MSE shrinkage factor (12) of order MLEs of two closely spaced frequencies versus SNR.

output of the frequency matched filter. Fig. 2 displays the empirical and theoretical MSE of ordered estimates (11) averaged over the two frequencies (since equal by symmetry) and the associated unbiased CRB. Fig. 3 displays the empirical and theoretical MSE shrinkage factor (12) averaged over the two frequencies as well. The empirical MSE are assessed with 10^4 Monte-Carlo trials and a frequency step $\delta\theta = (1/12N)\frac{1}{64}$. As expected there is a perfect match between the theoretical results and the empirical results in the asymptotic region (SNR $\geq$ 55 dB) where Gaussianity is reached by MLEs, highlighting the non-negligible impact (-0.6 dB) of estimates ordering on high resolution scenarios. As the SNR values enter the threshold area, the MLEs start to explore the side lobe peaks of the matched filter producing outliers and loosing gradually their Gaussian distribution [4]. Because of this phenomenon, as the SNR values enter the threshold area, the results derived in this section must be regarded as an approximation which accuracy is somewhat difficult to quantify in general. Indeed, it would require to recompute the second-order statistical prediction of ordered estimates taking into account the probability of outlier [4], which is far beyond the scope of this communication. We can simply notice, but without generality, that in our example, the approximation is quite good for SNR ≥ 30 dB.

4. Conclusion

In this communication we have provided a second-order statistical prediction of ordered normally distributed estimates which allows to refine the asymptotic performance analysis of the MSE of MLEs of a subset $\boldsymbol{\theta}$ of the parameters set $\boldsymbol{\Theta}$. Analytical expressions of the MSEs of the whole parameters set $\boldsymbol{\Theta}$ after re-ordering of the subset $\boldsymbol{\theta}$ should be the next topic to be addressed.

Appendix

As the vector $\widehat{\boldsymbol{\xi}}_{m,i} = ((\widehat{\boldsymbol{\theta}}_i)_m, \widehat{\mathbf{u}}_i^T)^T$, $\widehat{\mathbf{u}}_i = \boldsymbol{\Delta}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\mu}_i) \sim \mathcal{N}_{M-1}(\mathbf{0}, \boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T)$, is a M-dimensional normal vector resulting from a bijective affine transformation of $\widehat{\boldsymbol{\theta}}_i$, then we have [1]

$$p_{\mathcal{N}_M}\left(\widehat{\boldsymbol{\xi}}_{m,i}; \boldsymbol{\mu}_{m,i}, \mathbf{C}_{m,i}\right) = p_{\mathcal{N}_1}\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_m | \widehat{\mathbf{u}}_i; \mu_{m|i}, \sigma^2_{m|i}\right) p_{\mathcal{N}_{M-1}}\left(\widehat{\mathbf{u}}_i; \mathbf{0}, \boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T\right)$$

where $\mu_{m|i} = \mathbf{d}_m^T\left(\boldsymbol{\mu}_i + \mathbf{C}_i\boldsymbol{\Delta}^T\left(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T\right)^{-1}\widehat{\mathbf{u}}_i\right)$ and $\sigma^2_{m|i} = \mathbf{d}_m^T \left(\mathbf{C}_{\widehat{\boldsymbol{\theta}}} - \mathbf{C}_i\boldsymbol{\Delta}^T\left(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T\right)^{-1}\boldsymbol{\Delta}C_i\right)\mathbf{d}_m$. Let $\mathcal{U}_i = \{\widehat{\mathbf{u}}_i : \widehat{\mathbf{u}}_i \geq -\boldsymbol{\Delta}\boldsymbol{\mu}_i\}$, then whatever the real valued function $f(\,)$:

$$E\left[f\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_m\right)|\mathcal{S}_i\right] = E\left[f\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_m\right)|\mathcal{U}_i\right] = E\left[E\left[f\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_m\right)|\widehat{\mathbf{u}}_i\right]|\mathcal{U}_i\right]$$

In the particular cases where $f\left(\widehat{\theta}_{(m)}\right) = \widehat{\theta}^k_{(m)}$, $k \in \{1,2\}$:

$$E\left[\left(\widehat{\boldsymbol{\theta}}_i\right)_m|\widehat{\mathbf{u}}_i\right] = \mu_{m|i} = \alpha_{m|i} + \boldsymbol{\beta}_{m|i}^T\widehat{\mathbf{u}}_i$$

$$E\left[\left(\widehat{\boldsymbol{\theta}}_i\right)^2_m|\widehat{\mathbf{u}}_i\right] = \sigma^2_{m|i} + \mu^2_{m|i} = \sigma^2_{m|i} + \alpha^2_{m|i} + 2\alpha_{m|i}\boldsymbol{\beta}_{m|i}^T\widehat{\mathbf{u}}_i + \boldsymbol{\beta}_{m|i}^T\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^T\boldsymbol{\beta}_{m|i}$$

where $\alpha_{m|i} = \mathbf{d}_m^T\boldsymbol{\mu}_i$, $\boldsymbol{\beta}_{m|i} = \left(\boldsymbol{\Delta}\mathbf{C}_i\boldsymbol{\Delta}^T\right)^{-1}\boldsymbol{\Delta}C_i\mathbf{d}_m$, and (2) can finally be expressed as

$$E\left[\widehat{\theta}_{(m)}\right] = \sum_{i=1}^{M!}\left(\alpha_{m|i}\mathcal{P}(\mathcal{U}_i) + \boldsymbol{\beta}_{m|i}^T E\left[\widehat{\mathbf{u}}_i 1_{\mathcal{U}_i}\right]\right)$$

$$E\left[\widehat{\theta}^2_{(m)}\right] = \sum_{i=1}^{M!}\left(\left(\sigma^2_{m|i} + \alpha^2_{m|i}\right)\mathcal{P}(\mathcal{U}_i) + 2\alpha_{m|i}\boldsymbol{\beta}_{m|i}^T E\left[\widehat{\mathbf{u}}_i 1_{\mathcal{U}_i}\right] + \boldsymbol{\beta}_{m|i}^T E\left[\widehat{\mathbf{u}}_i(\widehat{\mathbf{u}}_i)^T 1_{\mathcal{U}_i}\right]\boldsymbol{\beta}_{m|i}\right) \tag{13}$$

From (2), $\forall m \in [1, M]$, $\forall l | m + l \in [1, M]$:

$$E\left[\left(\widehat{\theta}_{(m+l)} - \widehat{\theta}_{(m)}\right)^2\right] = \sum_{i=1}^{M!} E\left[\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_{m+l} - \left(\widehat{\boldsymbol{\theta}}_i\right)_m\right)^2 |\mathcal{S}_i\right]\mathcal{P}(\mathcal{S}_i)$$

Additionally, $\forall l \geq 1 | m + l \in [1, M]$:

$$\left(\widehat{\boldsymbol{\theta}}_i\right)_{m+l} - \left(\widehat{\theta}_i\right)_m = \left(\mathbf{1}_{M-1}^{m:m+l-1}\right)^T\boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i$$

therefore

$$\left(\left(\widehat{\boldsymbol{\theta}}_i\right)_{m+l} - \left(\widehat{\boldsymbol{\theta}}_i\right)_m\right)^2 = \left(\mathbf{1}_{M-1}^{m:m+l-1}\right)^T\boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i\boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i^T\mathbf{1}_{M-1}^{m:m+l-1}$$

Finally

$$E\left[\left(\widehat{\theta}_{(m+l)} - \widehat{\theta}_{(m)}\right)^2\right] = \sum_{i=1}^{M!}\left(\left(\mathbf{1}_{M-1}^{m:m+l-1}\right)^T E\left[\boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i\boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i^T|\mathcal{S}_i\right]\mathbf{1}_{M-1}^{m:m+l-1}\right)\mathcal{P}(\mathcal{S}_i) \tag{14}$$

Last, as

$$E\left[\boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i\boldsymbol{\Delta}\widehat{\boldsymbol{\theta}}_i^T|\mathcal{S}_i\right] = E\left[\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^T|\mathcal{U}_i\right] + E\left[\widehat{\mathbf{u}}_i|\mathcal{U}_i\right](\boldsymbol{\Delta}\boldsymbol{\mu}_i)^T + \boldsymbol{\Delta}\boldsymbol{\mu}_i E\left[\widehat{\mathbf{u}}_i|\mathcal{U}_i\right]^T + \boldsymbol{\Delta}\boldsymbol{\mu}_i(\boldsymbol{\Delta}\boldsymbol{\mu}_i)^T$$

an equivalent expression of (14) is (7).

References

[1] Y.L. Tong, The Multivariate Normal Distribution, Springer, New York, 1990.

[2] H.A. David, H.N. Nagaraja (Eds.), Order Statistics, 3rd ed. Wiley, New York, 2003.

[3] R.B. Arellano-Vallea, M.G. Genton, On the exact distribution of linear combinations of order statistics from dependent random variables, J. Multivar. Anal. 98 (2007) 1876–1894.

[4] H.L. van Trees, Optimum Array Processing, Wiley-Interscience, New-York, 2002.

[5] P. Stoica, A. Nehorai, Performances study of conditional and unconditional direction of arrival estimation, IEEE Trans. Signal Process. 38 (10) (1990) 1783–1795.

[6] M. Viberg, B. Ottersten, Sensor array processing based on subspace fitting, IEEE Trans. Signal Process. 39 (5) (1991) 1110–1121.

[7] B. Ottersten, M. Viberg, P. Stoica, A. Nehorai, Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing, in: S. Haykin, J. Litva, T.J. Shepherd (Eds.), Radar Array Processing, Springer-Verlag, Berlin, Germany, 1993, pp. 99–151. (Chapter 4).

[8] J. Li, R.T. Compton, Maximum likelihood angle estimation for signals with known waveforms, IEEE Trans. Signal Process. 41 (9) (1993) 2850–2862.

[9] A. Renaux, P. Forster, E. Chaumette, P. Larzabal, On the high snr cml estimator full statistical characterization, IEEE Trans. Signal Process. 54 (12) (2006) 4840–4843.

[10] X. Mestre, P. Vallet, P. Loubaton, On the resolution probability of conditional and unconditional maximum likelihood doa estimation, in: Proceedings of the Eurasip EUSIPCO, 2013.

[11] M.P. Clark, On the resolvability of normally distributed vector parameter estimates, IEEE Trans. Signal Process. 43 (12) (1995) 2975–2981.

[12] C. Ren, M.N. El Korso, J. Galy, E. Chaumette, P. Larzabal, A. Renaux, On the accuracy and resolvability of vector parameter estimates, IEEE Trans. Signal Process. 62 (14) (2014) 3682–3694.

[13] I. Olkin, M.A.G. Viana, Correlation analysis of extreme observations from a multivariate normal distribution, J. Am. Stat. Assoc. 90 (1995) 1373–1379.

[14] N. Loperfido, A note on skew-elliptical distributions and linear functions of order statistics, Stat. Probab. Lett. 78 (2009) 3184–3186.

[15] J.R.M. Hosking, L-moments: analysis and estimation of distributions using linear combinations of order statistics, J. R. Stat. Soc. 52 (1) (1990) 105–124.

[16] H.M. Barakat, Multivariate order statistics based on dependent and nonidentically distributed random variables, J. Multivar. Analysis 100 (2009) 81–90.

[17] A. Genz, Numerical computation of multivariate normal probabilities, J. Comput. Graph. Stat. 1 (1992) 141–149.