CONCOMITANT OF ORDERED MULTIVARIATE NORMAL DISTRIBUTION WITH APPLICATION TO PARAMETRIC INFERENCE

Eric Chaumette and François Vincent

University of Toulouse/Isae-Supaero, 10 Avenue Edouard Belin, 31400 Toulouse, France.
(eric.chaumette@isae.fr, francois.vincent@isae.fr)

ABSTRACT

In statistics, the concept of a concomitant, also called the induced order statistic, arises when one sorts the members of a random sample according to corresponding values of another random sample. Indeed, multivariate order statistics induced by the ordering of linear combinations of the components arises naturally in many instances. As a contribution, we provide a general second-order statistical prediction of concomitant of order statistics for multivariate normal distribution, generalizing earlier works. We exemplify its usefulness in parametric inference via two examples related to deterministic and Bayesian estimation.

Index Terms— Multivariate normal distribution, Order statistics, Concomitants, Parametric Inference, Mean Square Error

1. INTRODUCTION

The ordered values of a sample of observations are called the order statistics of the sample: if $\boldsymbol{\theta}^T = (\theta_1, \ldots, \theta_M)^1$ is a vector of M real valued random variables, then $\boldsymbol{\theta}_{(M)} = \left(\theta_{(1)}, \ldots, \theta_{(M)}\right)^T$ denotes the vector of order statistics induced by $\boldsymbol{\theta}$ where $\theta_{(1)} \leq \theta_{(2)} \leq \ldots \leq \theta_{(M)}$ [1]. Order statistics and extreme 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 concentration of probability in a distribution. Order statistics are also useful in statistical inference, where estimates of parameters can be based on some suitable functions of the order statistics vector (robust location estimates, detection of outliers, censored sampling, characterizations and goodness of fit....) [1] or be implicitly ordered as in maximum likelihood estimation for parametric inference (see Section 3 and [2]). Since there is no direct extension of order concept to multivariate random variables, the extension of procedure based on order statistics to such situations is inapplicable. However, if we consider a random sample arising from a bivariate distribution $\{(s_1, \theta_1), \ldots, (s_M, \theta_M)\}$, ordering of the values recorded on the first variable s generates a set of random variables associated with the corresponding θ variate [3]. These random variables obtained due to the ordering of the θ's are known as the concomitants of order statistics $\mathbf{s}_{(M)}$ and are denoted $\boldsymbol{\theta}_{[M]}^T = \left(\theta_{[1]}, \ldots, \theta_{[M]}\right)$. Hence the general concept of a concomitant in statistics, also called the induced order statistic, arising when one sorts the members of a random sample according to corresponding values of another random sample [1]. In that perspective, a generalization of the bivariate case, where the sample $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_M\}$ consists of M multivariate random variables, is obtained by resorting to a linear combination of the form $\mathbf{s}^T = \left(\boldsymbol{\theta}_1^T \mathbf{a}, \ldots, \boldsymbol{\theta}_M^T \mathbf{a}\right)$. Then the ordering of the sample $\mathbf{s}$, i.e. $\mathbf{s}_{(M)}$, induces the associate ordering of random vectors $\boldsymbol{\theta}_{[m]}$, i.e. $\boldsymbol{\Theta}_{[M]} = \left[\boldsymbol{\theta}_{[1]} \ \ldots \ \boldsymbol{\theta}_{[M]}\right]$ [1]. Multivariate order statistics induced by the ordering of linear combinations of the components arises naturally in many instances. For example, in the evaluation of the performance of students in a course, the final grade may be a weighted average of the scores in a mid-term test and the final examination. Other interesting examples arise in hydrology while analyzing extreme lake levels [4], in biological selection problem [5], ocean engineering [6], development of structural designs [7]. Therefore, the need to characterize the order statistics and their concomitants has led to a large body of work summarized in [1][8][9]. A fairly general second-order statistical prediction of concomitants of ordered multivariate normal distribution has been given in [4] and [10] for the situation in which the random vectors $\boldsymbol{\theta}_m$ are independent. Unfortunately the situation where vectors $\boldsymbol{\theta}_m$ are independent is not the common situation in many instances of the setting under consideration (see section 3).

Therefore, as a contribution, we provide the most general second-order statistical prediction of order statistics and their concomitants for multivariate normal distribution, whatever they are dependent or independent. These closed forms generalize the earlier work from [4] and [10].

We exemplify their usefulness in parametric inference. Indeed, the asymptotic performance analysis of the mean square error (MSE) of maximum likelihood estimators (MLEs) can be refined by the study of concomitants of ordered estimates (generalizing the single unknown parameter case addressed in [2]). In Kalman filtering for linear discrete state-space models, concomitants of ordered estimates can be used to monitor the range of the states vector.

2. STATISTICAL PREDICTION OF CONCOMITANTS OF ORDERED MULTIVARIATE NORMAL DISTRIBUTION

Let us consider the observation of M random Gaussian vectors with P components: $\{\boldsymbol{\theta}_m\}_{m=1}^M$. The vector gathering the PM Gaussian random variables is denoted by $\mathbf{v}_{\boldsymbol{\Theta}}$ where $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1 \ \ldots \ \boldsymbol{\theta}_M] \in \mathcal{M}_{\mathbb{R}}(P, M)$, and:

This work has been partially supported by the DGA/MRIS (2015.60.0090.00.470.75.01).

[1]The n-th coordinate of the column vector $\mathbf{a}$ is denoted by a_n or $(\mathbf{a})_n$. The n-th row and m-th column element of the matrix $\mathbf{A}$ is denoted by $A_{n,m}$ or $(\mathbf{A})_{n,m}$. If $\mathbf{A} = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_M]$, then $\mathbf{v_A} \triangleq \mathbf{vec}(\mathbf{A}) = \left(\mathbf{a}_1^T, \ldots, \mathbf{a}_M^T\right)^T$. $\mathbf{1}_M$ denotes the M-dimensional column vector with all components set to 1. $\mathbf{I}_M \in \mathbb{R}^{M \times M}$ denotes the identity matrix. $\mathcal{P}(\mathcal{A})$ and $1_{\{\mathcal{A}\}}$ denote the probability and the indicator function of an event $\mathcal{A}$. $E[\mathbf{g}(\mathbf{y})] = \int \mathbf{g}(\mathbf{y})\, p(\mathbf{y})\, d\mathbf{y}$ denotes the statistical expectation of the vector of functions $\mathbf{g}(\,)$ with respect to the random vector $\mathbf{y}$.

$$\mathbf{v}_{\boldsymbol{\Theta}}^T = \left(\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_M^T\right),\ \mathbf{v}_{\boldsymbol{\Theta}} \sim \mathcal{N}_{PM}\left(\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}}, \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}}\right), \tag{1}$$

$$\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}} = \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{\theta}_1} \\ \vdots \\ \boldsymbol{\mu}_{\boldsymbol{\theta}_M} \end{pmatrix}, \quad \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}} = \begin{bmatrix} \mathbf{C}_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_1} & \dots & \mathbf{C}_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_M} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{\boldsymbol{\theta}_M,\boldsymbol{\theta}_1} & \dots & \mathbf{C}_{\boldsymbol{\theta}_M,\boldsymbol{\theta}_M} \end{bmatrix}$$

As in [4], let us consider the following M-dimensional vector:

$$\mathbf{s} = \left(\boldsymbol{\theta}_1^T\mathbf{a}, \dots, \boldsymbol{\theta}_M^T\mathbf{a}\right)^T = \boldsymbol{\Theta}^T\mathbf{a},\ \mathbf{a} \in \mathcal{M}_{\mathbb{R}}(P,1), \tag{2a}$$

then, the concomitants of $\mathbf{s}_{(M)} = \left(s_{(1)}, \dots, s_{(M)}\right)$ are defined as:

$$\boldsymbol{\Theta}_{[M]} = \left[\boldsymbol{\theta}_{[1]}\ \dots\ \boldsymbol{\theta}_{[M]}\right] \mid \boldsymbol{\theta}_{[m]} = \boldsymbol{\theta}_{m'} \Leftrightarrow s_{(m)} = s_{m'}. \tag{2b}$$

Let $\{\mathbf{d}_1, ..., \mathbf{d}_M\}$ be the M-dimensional unit basis vectors; then:

$$\boldsymbol{\theta}_{[m]} = \mathbf{vec}\left(\boldsymbol{\Theta}_{[M]}\mathbf{d}_m\right) = \mathbf{S}_m\mathbf{v}_{\boldsymbol{\Theta}_{[M]}},\ \mathbf{S}_m = \mathbf{d}_m^T \otimes \mathbf{I}_P, \tag{3a}$$

$$E\left[\boldsymbol{\theta}_{[m]}\right] = \mathbf{S}_m E\left[\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right],\ \mathbf{C}_{\boldsymbol{\theta}_{[m]},\boldsymbol{\theta}_{[m']}} = \mathbf{S}_m\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}\mathbf{S}_{m'}^T. \tag{3b}$$

In other words, the first and second order statistical prediction of $\left\{\boldsymbol{\theta}_{[1]}, \dots, \boldsymbol{\theta}_{[M]}\right\}$ derive from the first and second order statistical prediction of $\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}$, which is introduced in this section. First, note that $\mathbf{s}_{(M)} \in Per(\mathbf{s})$, where $Per(\mathbf{s}) = \{\mathbf{s}_i = \mathbf{P}_i\mathbf{s}; i = 1, \dots, M!\}$ is the collection of random vectors $\mathbf{s}_i$ corresponding to the $M!$ different permutations of the components of $\mathbf{s}$. Here $\mathbf{P}_i \in \mathbb{R}^{M\times M}$ are permutation matrices with $\mathbf{P}_i \neq \mathbf{P}_j$ for all $i \neq j$. Let $\boldsymbol{\Delta} \in \mathbb{R}^{(M-1)\times M}$ be the difference matrix such that $\boldsymbol{\Delta}\mathbf{s} = (s_2 - s_1, s_3 - s_2, ..., s_M - s_{M-1})^T$, i.e., the mth row of $\boldsymbol{\Delta}$ is $\mathbf{d}_{m+1}^T - \mathbf{d}_m^T$, $m = 1, ..., M-1$. Let $\mathcal{S}_i = \{\mathbf{s} : \boldsymbol{\Delta}\mathbf{s}_i \geq \mathbf{0}\}$ where $\mathbf{s}_i \sim \mathcal{N}_M\left(\boldsymbol{\mu}_{\mathbf{s}_i}, \mathbf{C}_{\mathbf{s}_i}\right)$, $\boldsymbol{\mu}_{\mathbf{s}_i} = \mathbf{P}_i\boldsymbol{\mu}_{\mathbf{s}}$, $\mathbf{C}_{\mathbf{s}_i} = \mathbf{P}_i\mathbf{C}_{\mathbf{s}}\mathbf{P}_i^T$. As the set of events $\{\mathcal{S}_i\}_{i=1}^{M!}$ is a partition of $\mathbb{R}^M$, whatever the vector of real valued functions $\mathbf{f}\ (\ )$, by the theorem of total probability we have:

$$E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right)\right] = \sum_{i=1}^{M!} E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right)|\mathcal{S}_i\right]\mathcal{P}(\mathcal{S}_i)$$

$$E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right)\right] = \sum_{i=1}^{M!} E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_i}\right)|\mathcal{S}_i\right]\mathcal{P}(\mathcal{S}_i) \tag{4a}$$

where $\boldsymbol{\Theta}_i = \boldsymbol{\Theta}\mathbf{P}_i^T$. However, from a computational point of view, it is wiser to express (4a) as:

$$E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right)\right] = \sum_{i=1}^{M!} E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_i}\right)|\mathcal{U}_i\right]\mathcal{P}(\mathcal{U}_i) \tag{4b}$$

where $\mathcal{U}_i = \{\mathbf{u}_i : \mathbf{u}_i \geq -\boldsymbol{\Delta}_i\boldsymbol{\mu}_{\mathbf{s}}\}$ and $\mathbf{u}_i = \boldsymbol{\Delta}_i(\mathbf{s} - \boldsymbol{\mu}_{\mathbf{s}}) \sim \mathcal{N}_{M-1}\left(\mathbf{0}, \boldsymbol{\Delta}_i\mathbf{C}_{\mathbf{s}}\boldsymbol{\Delta}_i^T\right)$, $\boldsymbol{\Delta}_i = \boldsymbol{\Delta}\mathbf{P}_i$. As [1]:

$$\boldsymbol{\xi}_i = \begin{pmatrix} \mathbf{u}_i \\ \mathbf{v}_{\boldsymbol{\Theta}_i} \end{pmatrix} \sim \mathcal{N}_{(M-1)+MP}\left(\boldsymbol{\mu}_{\boldsymbol{\xi}_i}, \mathbf{C}_{\boldsymbol{\xi}_i}\right), \tag{5a}$$

$$\boldsymbol{\mu}_{\boldsymbol{\xi}_i} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} \end{pmatrix},\ \mathbf{C}_{\boldsymbol{\xi}_i} = \begin{bmatrix} \mathbf{C}_{\mathbf{u_i}} & \mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}} \\ \mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}^T & \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} \end{bmatrix}, \tag{5b}$$

where $\mathbf{v}_{\boldsymbol{\Theta}_i} = (\mathbf{P}_i \otimes \mathbf{I}_P)\,\mathbf{v}_{\boldsymbol{\Theta}}$, therefore:

$$E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_i}\right)|\mathcal{U}_i\right] = E\left[E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_i}\right)|\mathbf{u}_i\right]|\mathcal{U}_i\right]$$

and (4b) can be finally rewritten as:

$$E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right)\right] = \sum_{i=1}^{M!} E\left[E\left[\mathbf{f}\left(\mathbf{v}_{\boldsymbol{\Theta}_i}\right)|\mathbf{u}_i\right]|\mathcal{U}_i\right]\mathcal{P}(\mathcal{U}_i) \tag{6}$$

In particular:

$$\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} = E\left[\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right] = \sum_{i=1}^{M!} E\left[E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i\right]|\mathcal{U}_i\right]\mathcal{P}(\mathcal{U}_i) \tag{7a}$$

$$\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} = \sum_{i=1}^{M!} E\left[E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}\mathbf{v}_{\boldsymbol{\Theta}_i}^T|\mathbf{u}_i\right]|\mathcal{U}_i\right] - \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}^T \tag{7b}$$

where:

$$E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i\right] = \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} + \mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{u}_i \tag{7c}$$

$$\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i} = \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} - \mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}} \tag{7d}$$

$$E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}\mathbf{v}_{\boldsymbol{\Theta}_i}^T|\mathbf{u}_i\right] = \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i} + E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i\right]E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i\right]^T \tag{7e}$$

Then a smart exploitation of (7a-7e) yields [12]:

$$\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} = \sum_{i=1}^{M!}\left(\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}}\mathcal{P}_i + \mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{e}_i\right) \tag{8a}$$

$$\begin{aligned} \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} &= \sum_{i=1}^{M!}\left(\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i} + \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}}\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\right)\mathcal{P}_i \\ &+ \sum_{i=1}^{M!}\left(\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}}\mathbf{e}_i^T\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}} + \mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{e}_i\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\right) \\ &+ \sum_{i=1}^{M!}\left(\mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{R}_i\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}\right) - \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}^T \end{aligned} \tag{8b}$$

where:

$$\mathcal{P}_i = \mathcal{P}(\mathcal{U}_i), \mathbf{e}_i = E\left[\mathbf{u}_i 1_{\{\mathcal{U}_i\}}\right],\ \mathbf{R}_i = E\left[\mathbf{u}_i(\mathbf{u}_i)^T 1_{\{\mathcal{U}_i\}}\right],$$

$$\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i} = \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} - \mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}}^T\mathbf{C}_{\mathbf{u}_i}^{-1}\mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}},$$

$$\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} = (\mathbf{P}_i \otimes \mathbf{I}_P)\,\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}},\ \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} = (\mathbf{P}_i \otimes \mathbf{I}_P)\,\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}}\left(\mathbf{P}_i^T \otimes \mathbf{I}_P\right),$$

$$\mathbf{C}_{\mathbf{u}_i} = \boldsymbol{\Delta}_i\mathbf{C}_{\mathbf{s}}\boldsymbol{\Delta}_i^T,\ \mathbf{C}_{\mathbf{s}} = \left(\mathbf{I}_M \otimes \mathbf{a}^T\right)\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}}(\mathbf{I}_M \otimes \mathbf{a}),$$

$$\mathbf{C}_{\mathbf{u}_i,\mathbf{v}_{\boldsymbol{\Theta}_i}} = \boldsymbol{\Delta}_i\mathbf{C}_{\mathbf{s},\mathbf{v}_{\boldsymbol{\Theta}}}\left(\mathbf{P}_i^T \otimes \mathbf{I}_P\right),\ \mathbf{C}_{\mathbf{s},\mathbf{v}_{\boldsymbol{\Theta}}} = \left(\mathbf{I}_M \otimes \mathbf{a}^T\right)\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}}.$$

As shown in [2], $\{\mathcal{P}_i, \mathbf{e}_i, \mathbf{R}_i\}_{i=1}^{M!}$ can be computed by resorting to algorithms proposed by Genz [1] for numerical evaluation of multivariate normal distributions and moments over domains included in $[-10, 10]^M$. Note that the use of (3b) in conjunction with (8a-8b) yields a generalization of [2] obtained for $P = 1$ and $\mathbf{a} \triangleq a = 1$.
The correctness of expressions (8a) and (8b) can be checked (see Appendix) by inspection of the case where the column vectors of matrix $\boldsymbol{\Theta}$ are i.i.d., which has been addressed in [4].
In the two sources case, $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1\ \boldsymbol{\theta}_2] \in \mathcal{M}_{\mathbb{R}}(P,2)$ and:

$$\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}} = \begin{pmatrix} \boldsymbol{\mu}_1 \triangleq \boldsymbol{\mu}_{\boldsymbol{\theta}_1} \\ \boldsymbol{\mu}_2 \triangleq \boldsymbol{\mu}_{\boldsymbol{\theta}_2} \end{pmatrix}, \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}} = \begin{bmatrix} \mathbf{C}_1 \triangleq \mathbf{C}_{\boldsymbol{\theta}_1} & \mathbf{C}_{1,2} \triangleq \mathbf{C}_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2} \\ \mathbf{C}_{1,2}^T & \mathbf{C}_2 \triangleq \mathbf{C}_{\boldsymbol{\theta}_2} \end{bmatrix}.$$

Moreover $\widehat{u}_1 = -\widehat{u}_2$, $\mathcal{P}_1 + \mathcal{P}_2 = 1$, $e_1 - e_2 = E[\widehat{u}_1] = 0$, $R_1 + R_2 = E\left[\widehat{u}_1^2\right] = \sigma_{\widehat{u}}^2 = \mathbf{a}^T(\mathbf{C}_1 + \mathbf{C}_2 - 2\mathbf{C}_{1,2})\,\mathbf{a}$, leading to:

$$\mathbf{C}_{\boldsymbol{\theta}_{[m]}} = E\left[\boldsymbol{\theta}_{[m]}\boldsymbol{\theta}_{[m]}^T\right] - E\left[\boldsymbol{\theta}_{[m]}\right]E\left[\boldsymbol{\theta}_{[m]}\right]^T, \tag{9}$$

$$E\left[\boldsymbol{\theta}_{[m]}\right] = \boldsymbol{\mu}_m + (-1)^{m-1}\left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\,\mathcal{P}_2 + \tfrac{\mathbf{a}_1 + \mathbf{a}_2}{\sigma_{\widehat{u}}^2}e_2\right),$$

$$\begin{aligned} E\left[\boldsymbol{\theta}_{[m]}\boldsymbol{\theta}_{[m]}^T\right] &= \mathbf{C}_m + \boldsymbol{\mu}_m\boldsymbol{\mu}_m^T \\ &+ (-1)^{m-1}\left(\mathbf{C}_2 - \mathbf{C}_1 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2^T - \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T - \tfrac{\mathbf{a}_2\mathbf{a}_2^T - \mathbf{a}_1\mathbf{a}_1^T}{\sigma_{\widehat{u}}^2}\right)\mathcal{P}_2 \\ &+ (-1)^{m-1}\left(\tfrac{\mathbf{a}_1\boldsymbol{\mu}_1^T + \boldsymbol{\mu}_1\mathbf{a}_1^T + \mathbf{a}_2\boldsymbol{\mu}_2^T + \boldsymbol{\mu}_2^T\mathbf{a}_2^T}{\sigma_{\widehat{u}}^2}\,e_2 + \tfrac{\mathbf{a}_2\mathbf{a}_2^T - \mathbf{a}_1\mathbf{a}_1^T}{\sigma_{\widehat{u}}^4}R_2\right), \end{aligned}$$

where $\mathbf{a}_1 = (\mathbf{C}_{1,2} - \mathbf{C}_1)\,\mathbf{a}$ and $\mathbf{a}_2 = \left(\mathbf{C}_{1,2}^T - \mathbf{C}_2\right)\mathbf{a}$.

3. APPLICATION TO PARAMETRIC INFERENCE

3.1. Maximum likelihood estimation

The ongoing success of ML estimators (MLEs) originates from the fact that, under reasonably general conditions on the probabilistic observation model [13][14], the MLEs are, in the limit of large sample support, Gaussian distributed and efficient. Additionally, if the observation model is Gaussian, some additional asymptotic regions of operation yielding, for a subset of MLEs, Gaussian distributed and efficient estimates, have also been identified at finite sample support [15][16][17][18][19]. However, many estimation problems are actually unidentifiable unless they are regularized by imposing the ordering of some unknown parameters. For illustration purposes, let us consider L independent observations of the linear model [11]:

$$\mathbf{y}(l) = \mathbf{H}(\boldsymbol{\Theta})\mathbf{x}(l) + \mathbf{v}(l), \ 1 \leq l \leq L, \tag{10}$$

where $\mathbf{y}(l)$ is the vector of samples of size N, M is the number of signal sources, $\mathbf{x}(l)$ is the vector of complex amplitudes of the M sources for the l^{th} observation, $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1 \ \ldots \ \boldsymbol{\theta}_M]$, $\mathbf{H}(\boldsymbol{\Theta}) = [\mathbf{h}(\boldsymbol{\theta}_1) \ \ldots \ \mathbf{h}(\boldsymbol{\theta}_M)]$ and $\mathbf{h}(\,)$ is a vector of N parametric functions depending on a vector of P unknown parameters $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^P$, $\mathbf{v}(l)$ are complex noises independent of the M sources. Since (10) is invariant over permutation of signal sources amplitude $\mathbf{x}(l)$, (10) is an unidentifiable estimation problem which is regularized by imposing the ordering of the unknown parameters $\{\boldsymbol{\theta}_m\}_{m=1}^M$. A straightforward ordering of $\{\boldsymbol{\theta}_m\}_{m=1}^M$ arises in the computation of MLEs which requires a multidimensional non linear optimization for which analytical solutions are in general not available. For instance, if (10) is a Gaussian conditional model [11][15]:

$$\widehat{\boldsymbol{\Theta}} = \left[\widehat{\boldsymbol{\theta}}_1 \ \ldots \ \widehat{\boldsymbol{\theta}}_M\right] = \arg\max_{\boldsymbol{\Theta}} \left\{ \sum_{l=1}^{L} \mathbf{y}(l)^H \, \boldsymbol{\Pi}_{\mathbf{H}(\boldsymbol{\Theta})} \mathbf{y}(l) \right\}, \tag{11}$$

where $\boldsymbol{\Pi}_{\mathbf{A}} = \mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$. Therefore one has to resort to numerical search techniques, generally compatible with computer programming, such as the conversion of a PM-dimensionnal search grid over Ω^M into a 1-dimensionnal search grid. For instance, in Matlab, this is done by the sub2ind.m function which returns the linear index equivalents to the specified subscripts for each dimension of an N-dimensional array. For example, if $P = 2$ and $\boldsymbol{\theta} \in \Omega = [a_1, b_1] \times [a_2, b_2]$, one can generate a rectangular search grid over Ω [11]:

$$\mathcal{G} = \left\{ \begin{pmatrix} a_1 + i_1\delta_1 \\ a_2 + i_2\delta_2 \end{pmatrix}, \middle| \begin{array}{l} \delta_1 = \frac{b_1 - a_1}{I_1}, \ 0 \leq i_1 \leq I_1 \\ \delta_2 = \frac{b_2 - a_2}{I_2}, \ 0 \leq i_2 \leq I_2 \end{array} \right\} \tag{12}$$

and convert each $\boldsymbol{\theta} \in \mathcal{G}$ into an equivalent linear search index $s = i_1 + (I_1 + 1)\, i_2$. Therefore, in practice (11) becomes $\widehat{\mathbf{s}} = \arg\max_{\mathbf{s}} \left\{ \sum_{l=1}^{L} \mathbf{y}(l)^H \, \boldsymbol{\Pi}_{\mathbf{H}(\boldsymbol{\Theta}(\mathbf{s}))} \mathbf{y}(l) \right\}$ and the issue of model identifiability is solved by ordering $\mathbf{s}$ yielding $\widehat{\mathbf{s}}_{(M)}$. If δ_1 and δ_2 are small enough, then $s \simeq \boldsymbol{\theta}^T\mathbf{a} - s_0$ where $\mathbf{a}^T = (1/\delta_1, (I_1+1)/\delta_2)$ and $s_0 = a_1/\delta_1 + (I_1 + 1)\, a_2/\delta_2$. Then, since the ordering does not depend on s_0, $\widehat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}\left(\widehat{\mathbf{s}}_{(M)}\right)$ are induced order statistic of $\widehat{\mathbf{s}}_{(M)}$ (2b), that is concomitants of $\widehat{\mathbf{s}}_{(M)}$.
Therefore, the asymptotic performance analysis of the MSE of MLEs[2] can be refined by the study of concomitant of ordered multivariate normal distribution.

For illustration purposes, let us consider a radar system consisting of a 1-element antenna array receiving scaled, timedelayed, and Doppler-shifted echoes of a known complex bandpass signal $e(t)\, e^{-j2\pi f_c t}$, where f_c is the carrier frequency. A standard observation model of a radar antenna receiving a pulse train of I pulses of duration δt_0 and bandwidth B, with a pulse repetition interval δt is given by (10) where [20] $L = 1$, $N = \lfloor \delta t / B \rfloor$, $\boldsymbol{\theta}^T = (\tau, \omega)$, $\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{\psi}(\omega) \otimes \boldsymbol{\phi}(\tau)$, $\boldsymbol{\psi}(\omega)^T = \left(1, \ldots, e^{j2\pi\omega(I-1)\delta t}\right)$, $\boldsymbol{\phi}(\tau)^T = \left(e(-\tau), \ldots, e\left(\frac{N-1}{B} - \tau\right)\right)$, τ and ω denoting the delay and the Doppler-shift associated to a target. The MLEs of $\boldsymbol{\Theta}$ are asymptotically efficient and Gaussian, and for 2 targets [21]:

$$\mathbf{C}_{\mathbf{v}_{\widehat{\boldsymbol{\Theta}}}} = \mathbf{CRB}_{\mathbf{v}_{\boldsymbol{\Theta}}} = 2\,\mathrm{Re}\left\{ \mathbf{J}(\boldsymbol{\Theta}) \odot \left(\left(\mathbf{x}_1^T \mathbf{x}_1^* \right) \otimes \mathbf{1}_{2\times 2} \right) \right\}^{-1},$$

where $\mathbf{J}(\boldsymbol{\Theta})$ is given in [22]. We consider a high resolution scenario in terms of $\boldsymbol{\theta}$, that is a small Doppler-Shift $d\omega = 1/(12I)$ $(I = 8)$ and a small delays difference $d\tau = 1/(8B)$ $(\delta t_0 = 32/B)$. $e(t)$ is a linear chirp. Figure (1) displays the empirical and theoretical MSE to CRB ratio (shrinkage factor) averaged over the two targets for both the delay and Doppler shift. The empirical MSE are assessed with 10^5 Monte-Carlo trials from the normally distributed vector associated with the asymptotic behavior of $\mathbf{v}_{\boldsymbol{\Theta}} \sim \mathcal{N}(\mathbf{v}_{\boldsymbol{\Theta}}, \mathbf{CRB}_{\mathbf{v}_{\boldsymbol{\Theta}}})$. The theoretical MSE is computed from (9). The match between theoretical and empirical results provides an empirical proof of the exactness of $\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}$ and $E\left[\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}\right]$ given in (8a-8b).

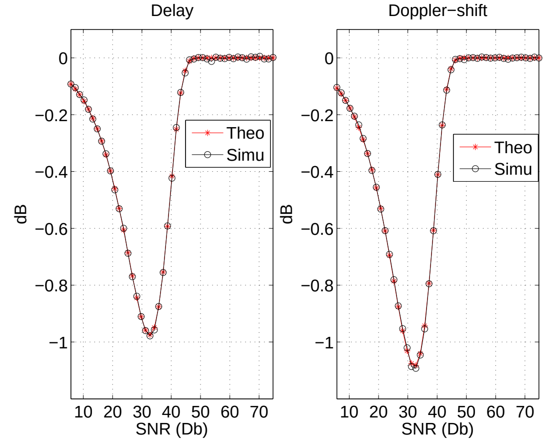

Fig. 1. Empirical and theoretical MSE to CRB ratio versus SNR

3.2. Monitoring of the states range of Kalman filters

We consider the class of real linear discrete state-space (LDSS) models represented with the state and measurement equations:

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad \mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, \tag{13}$$

where the time index $k \geq 1$, $\mathbf{x}_k$ is the M-dimensional state vector, $\mathbf{y}_k$ is the N-dimensional measurement vector and the model matrices $\mathbf{F}_k$ and $\mathbf{H}_k$ are known. The process noise sequence $\{\mathbf{w}_k\}$ and the measurement noise sequence $\{\mathbf{v}_k\}$, as well as the initial state $\mathbf{x}_0$ are Gaussian random vectors. $\widehat{\mathbf{x}}_{k|k} \triangleq \widehat{\mathbf{x}}_{k|k}(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ denotes an estimate of $\mathbf{x}_k$ based on measurements up to and including time k. If $\{\mathbf{w}_k\}$, $\{\mathbf{v}_k\}$ and $\mathbf{x}_0$ are uncorrelated, then the minimum MSE

[2] This refinement is applicable to any other estimators, such as M-estimators, Bayesian estimators (MAP, MMSE), as long as their distribution is normal multivariate

estimator $\widehat{\mathbf{x}}_{k|k}$ for LDSS models has a recursive predictor/corrector format, aka the Kalman filter (KF) [23][24, §9.1][25, §7.1]:

$$\widehat{\mathbf{x}}_{k|k} = \mathbf{F}_{k-1}\widehat{\mathbf{x}}_{k-1|k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{F}_{k-1}\widehat{\mathbf{x}}_{k-1|k-1}\right) \tag{14a}$$

$$E\left[\widehat{\mathbf{x}}_{k|k}\right] = E\left[\mathbf{x}_k\right], \quad \mathbf{C}_{\widehat{\mathbf{x}}_{k|k}} = \mathbf{C}_{\mathbf{x}_k} - \mathbf{P}_{k|k}, \tag{14b}$$

where $E\left[\mathbf{x}_k\right] = \mathbf{F}_{k-1}E\left[\mathbf{x}_{k-1}\right]$, $\mathbf{C}_{\mathbf{x}_k} = \mathbf{F}_{k-1}\mathbf{C}_{\mathbf{x}_{k-1}}\mathbf{F}_{k-1}^T + \mathbf{C}_{\mathbf{w}_{k-1}}$, $\mathbf{P}_{k|k} = E\left[\left(\widehat{\mathbf{x}}_{k|k} - \mathbf{x}_k\right)\left(\widehat{\mathbf{x}}_{k|k} - \mathbf{x}_k\right)^T\right]$ and $\mathbf{K}_k$ verify:

$$\begin{aligned}
\mathbf{P}_{k|k} &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_{k|k-1}, \\
\mathbf{K}_k &= \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^H + \mathbf{C}_{\mathbf{v}_k}\right)^{-1}, \\
\mathbf{P}_{k|k-1} &= \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^H + \mathbf{C}_{\mathbf{w}_{k-1}}.
\end{aligned} \tag{14c}$$

At each time index k, the range of states vector $\mathbf{x}_k$ is:

$$\gamma_k = \max\left\{\mathbf{x}_k\right\} - \min\left\{\mathbf{x}_k\right\} = \left(\mathbf{x}_k\right)_{(M)} - \left(\mathbf{x}_k\right)_{(1)}. \tag{15}$$

The ability to monitor the states range of a Kalman filter is of first importance if the actual system modeled by the linear state equations (13) can break down or enter a non-linear mode when the states range exceeds a given range limit. In that perspective, if L i.i.d. LDSS models are available, that is $\mathbf{x}_k(l) = \mathbf{F}_{k-1}\mathbf{x}_{k-1}(l) + \mathbf{w}_{k-1}(l)$ and $\mathbf{y}_k(l) = \mathbf{H}_k\mathbf{x}_k(l) + \mathbf{v}_k(l)$, $1 \le l \le L$, then one can derive from the concomitants of:

$$\mathbf{s} = \boldsymbol{\Theta}^T\mathbf{a},\ \boldsymbol{\Theta} \triangleq \left[\widehat{\mathbf{x}}_{k|k}(1)\ldots\widehat{\mathbf{x}}_{k|k}(L)\right]^T,\ \mathbf{a} = L^{-1}\mathbf{1}_L, \tag{16a}$$

the following estimator of γ_k [12]:

$$\widehat{\gamma}_k = \mathbf{a}^T\left(\boldsymbol{\theta}_{[M]} - \boldsymbol{\theta}_{[1]}\right). \tag{16b}$$

Then $\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}}$ and $\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}}$ in (1) are obtained from:

$$\boldsymbol{\mu}_{\boldsymbol{\theta}_m} = E\left[x_m\right]\mathbf{1}_L,\ \mathbf{C}_{\boldsymbol{\theta}_m,\boldsymbol{\theta}_{m'}} = \left(\mathbf{C}_{\widehat{\mathbf{x}}_{k|k}}\right)_{m,m'}\mathbf{I}_L, \tag{17a}$$

yielding generally correlated column vectors of $\boldsymbol{\Theta}$ (16a) since $\mathbf{C}_{\widehat{\mathbf{x}}_{k|k}}$ are not diagonal matrices in general (14b); hence the need of (8a-8b) to assess the second order statistical prediction of $\widehat{\gamma}_k$ (16b).

4. APPENDIX

If the column vectors of matrix $\boldsymbol{\Theta} = \left[\boldsymbol{\theta}_1\ \ldots\ \boldsymbol{\theta}_M\right]$ are i.i.d., then $\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} = \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}} = \mathbf{1}_M \otimes \boldsymbol{\mu}_{\boldsymbol{\theta}}$, $\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}} = \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}} = \mathbf{I}_M \otimes \mathbf{C}_{\boldsymbol{\theta}}$, where $\boldsymbol{\mu}_{\boldsymbol{\theta}} \in \mathcal{M}_{\mathbb{R}}(P,1)$ and $\mathbf{C}_{\boldsymbol{\theta}} \in \mathcal{M}_{\mathbb{R}}(P,P)$. As a consequence: $\boldsymbol{\mu}_{\mathbf{s}_i} = \boldsymbol{\mu}_{\mathbf{s}} = \mathbf{a}^T\boldsymbol{\mu}_{\boldsymbol{\theta}}\mathbf{1}_M$, $\boldsymbol{\Delta}_i\boldsymbol{\mu}_{\mathbf{s}} = \mathbf{0}$, $\mathcal{U}_i = \left\{\mathbf{u}_i | \mathbf{u}_i \ge \mathbf{0}\right\}$, $\mathbf{C}_{\mathbf{s}} = \sigma_{\boldsymbol{\theta}}^2\mathbf{I}_M$, $\sigma_{\boldsymbol{\theta}}^2 = \mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}$, $\mathbf{C}_{\mathbf{s},\mathbf{v}_{\boldsymbol{\Theta}}} = \mathbf{I}_M \otimes \left(\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\right)$, leading to $\boldsymbol{\xi}_i \sim \mathcal{N}_{(M-1)+MP}\left(\boldsymbol{\mu}_{\boldsymbol{\xi}}, \mathbf{C}_{\boldsymbol{\xi}}\right)$ where $\boldsymbol{\mu}_{\boldsymbol{\xi}} = \binom{\mathbf{0}}{\mathbf{1}_M \otimes \boldsymbol{\mu}_{\boldsymbol{\theta}}}$ and:

$$\mathbf{C}_{\boldsymbol{\xi}} = \left[\begin{array}{cc} \sigma_{\boldsymbol{\theta}}^2\boldsymbol{\Delta}\boldsymbol{\Delta}^T & \boldsymbol{\Delta}\left(\mathbf{I}_M \otimes \left(\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\right)\right) \\ \left(\mathbf{I}_M \otimes \mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}\right)\boldsymbol{\Delta}^T & \mathbf{I}_M \otimes \mathbf{C}_{\boldsymbol{\theta}} \end{array}\right].$$

Thus, $\mathbf{u}_i = \boldsymbol{\Delta}_i\mathbf{s} \sim \mathbf{u} = \boldsymbol{\Delta}\mathbf{s} \sim \mathcal{N}_{M-1}\left(\mathbf{0}, \sigma_{\boldsymbol{\theta}}^2\boldsymbol{\Delta}\boldsymbol{\Delta}^T\right)$ and:

$$E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i\right] = \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}} + \mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}^T\mathbf{C}_{\mathbf{u}}^{-1}\mathbf{u}_i, \tag{18a}$$

$$\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i} = \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}} - \mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}^T\mathbf{C}_{\mathbf{u}}^{-1}\mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}, \tag{18b}$$

$$E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}\mathbf{v}_{\boldsymbol{\Theta}_i}^T|\mathbf{u}_i\right] = \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}|\mathbf{u}} + E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i\right]E\left[\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathbf{u}_i\right]^T, \tag{18c}$$

yielding the following simplified form of (8a):

$$\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} = \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}}\textstyle\sum_{i=1}^{M!}\mathcal{P}\left(\mathcal{U}_i\right) + \mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}^T\mathbf{C}_{\mathbf{u}}^{-1}\sum_{i=1}^{M!}E\left[\mathbf{u}_i 1_{\{\mathcal{U}_i\}}\right]$$

Let $\widehat{\mathbf{v}}_i = \frac{\mathbf{u}_i}{\sigma_{\boldsymbol{\theta}}} \sim \mathcal{N}_{M-1}\left(\mathbf{0}, \boldsymbol{\Delta}\boldsymbol{\Delta}^T\right)$ and $\mathcal{V}_i = \left\{\widehat{\mathbf{v}}_i : \widehat{\mathbf{v}}_i \ge \mathbf{0}\right\}$. Since:

$$E\left[\mathbf{u}_i 1_{\{\mathcal{U}_i\}}\right] = E\left[\mathbf{u}_i|\mathcal{U}_i\right]\mathcal{P}\left(\mathcal{U}_i\right) = E\left[\widehat{\mathbf{v}}_i|\mathcal{V}_i\right]\mathcal{P}\left(\mathcal{V}_i\right) = E\left[\widehat{\mathbf{v}}_i 1_{\{\mathcal{V}_i\}}\right]$$

then, denoting $\boldsymbol{\Delta}^{\#} = \boldsymbol{\Delta}^T\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^T\right)^{-1}$:

$$\mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}^T\mathbf{C}_{\mathbf{u}}^{-1}\textstyle\sum_{i=1}^{M!}E\left[\mathbf{u}_i 1_{\{\mathcal{U}_i\}}\right] = \sigma_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\mathbf{s},\mathbf{v}_{\boldsymbol{\Theta}}}^T\boldsymbol{\Delta}^{\#}\displaystyle\sum_{i=1}^{M!}E\left[\widehat{\mathbf{v}}_i 1_{\{\mathcal{V}_i\}}\right].$$

In point of fact, it has be shown in [2] that:

$$\boldsymbol{\Delta}^{\#}\textstyle\sum_{i=1}^{M!}E\left[\widehat{\mathbf{v}}_i 1_{\{\mathcal{V}_i\}}\right] = E\left[\mathbf{z}_{(M)}\right],\ \mathbf{z} \sim \mathcal{N}_M\left(\mathbf{0}, \mathbf{I}_M\right),$$

therefore (8a) in the i.i.d. case is simply:

$$\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} = \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}} + \sigma_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\mathbf{s},\mathbf{v}_{\boldsymbol{\Theta}}}^T E\left[\mathbf{z}_{(M)}\right] \tag{19}$$

The connection between $\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}$ (7b)(8b) and [4] is a little bit more tricky to establish and for sake of space, we will only provide a sketch of the rationale detailed in [12]. Starting from the following alternative expression of (7b):

$$\begin{aligned}
\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} &= \textstyle\sum_{i=1}^{M!}\left(\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathcal{U}_i} + \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathcal{U}_i}\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathcal{U}_i}^T\right)\mathcal{P}\left(\mathcal{U}_i\right) \\
&\qquad - \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}}^T \\
\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathcal{U}_i} &= \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}|\mathbf{u}} + \mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}^T\mathbf{C}_{\mathbf{u}}^{-1}\mathbf{C}_{\mathbf{u}_i|\mathcal{U}_i}\mathbf{C}_{\mathbf{u}}^{-1}\mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}} \\
\boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}_i}|\mathcal{U}_i} &= \boldsymbol{\mu}_{\mathbf{v}_{\boldsymbol{\Theta}}} + \mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}^T\mathbf{C}_{\mathbf{u}}^{-1}E\left[\mathbf{u}_i|\mathcal{U}_i\right]
\end{aligned}$$

allows to prove that:

$$\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} = \mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}}} + \mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}}^T\mathbf{C}_{\mathbf{u}}^{-1}\left(\mathbf{C}_{\mathbf{u}_{[M-1]}} - \mathbf{C}_{\mathbf{u}}\right)\mathbf{C}_{\mathbf{u}}^{-1}\mathbf{C}_{\mathbf{u},\mathbf{v}_{\boldsymbol{\Theta}}},$$

that is:

$$\begin{aligned}
\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} &= \mathbf{I}_M \otimes \mathbf{C}_{\boldsymbol{\theta}} + \tfrac{\mathbf{I}_M \otimes \left(\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\right)}{\sigma_{\mathbf{s}}}\mathbf{Q}\tfrac{\mathbf{I}_M \otimes \left(\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\right)}{\sigma_{\mathbf{s}}}, \\
\mathbf{Q} &= \boldsymbol{\Delta}^{\#}\left(\sigma_{\mathbf{s}}^{-2}\mathbf{C}_{\mathbf{u}_{[M-1]}} - \boldsymbol{\Delta}\boldsymbol{\Delta}^T\right)\left(\boldsymbol{\Delta}^{\#}\right)^T,
\end{aligned}$$

where in fact [12] $\mathbf{Q} = \mathbf{C}_{\mathbf{z}_{(M)}} - \mathbf{I}_M$, leading to:

$$\mathbf{C}_{\mathbf{v}_{\boldsymbol{\Theta}_{[M]}}} = \mathbf{I}_M \otimes \mathbf{C}_{\boldsymbol{\theta}} + \frac{\mathbf{I}_M \otimes \left(\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}\right)}{\sigma_{\mathbf{s}}}\left(\mathbf{C}_{\mathbf{z}_{(M)}} - \mathbf{I}_M\right)\frac{\mathbf{I}_M \otimes \left(\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\right)}{\sigma_{\mathbf{s}}} \tag{20}$$

Finally, remembering that $\mathbf{vec}\left(\mathbf{AXB}\right) = \left(\mathbf{B}^T \otimes \mathbf{A}\right)\mathbf{vec}\left(\mathbf{X}\right)$ and $\left(\mathbf{A} \otimes \mathbf{B}\right)\left(\mathbf{C} \otimes \mathbf{D}\right) = \mathbf{AC} \otimes \mathbf{BD}$, one obtains from (3a-3b):

$$\begin{aligned}
E\left[\boldsymbol{\theta}_{[m]}\right] &= \left(\mathbf{d}_m^T \otimes \mathbf{I}_P\right)\left(\mathbf{1}_M \otimes \boldsymbol{\mu}_{\boldsymbol{\theta}}\right) + \\
&\qquad \sigma_{\boldsymbol{\theta}}^{-1}\left(\mathbf{d}_m^T \otimes \mathbf{I}_P\right)\left(\mathbf{I}_M \otimes \left(\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}\right)\right)E\left[\mathbf{z}_{(m)}\right] \\
E\left[\boldsymbol{\theta}_{[m]}\right] &= \left(1 \otimes \boldsymbol{\mu}_{\boldsymbol{\theta}}\right) + \sigma_{\boldsymbol{\theta}}^{-1}\left(\mathbf{d}_m^T \otimes \mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}\right)E\left[\mathbf{z}_{(m)}\right] \\
E\left[\boldsymbol{\theta}_{[m]}\right] &= \boldsymbol{\mu}_{\boldsymbol{\theta}} + \sigma_{\boldsymbol{\theta}}^{-1}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}\mathbf{d}_m^T E\left[\mathbf{z}_{(m)}\right]
\end{aligned}$$

and:

$$\begin{aligned}
\mathbf{C}_{\boldsymbol{\theta}_{[m]},\boldsymbol{\theta}_{[m']}} &= \left(\mathbf{d}_m^T \otimes \mathbf{I}_P\right)\left(\mathbf{I}_M \otimes \mathbf{C}_{\boldsymbol{\theta}}\right)\left(\mathbf{d}_{m'} \otimes \mathbf{I}_P\right) + \\
&\quad \tfrac{\left(\mathbf{d}_m^T \otimes \mathbf{I}_P\right)\left(\mathbf{I}_M \otimes \left(\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}\right)\right)}{\sigma_{\boldsymbol{\theta}}}\left(\mathbf{C}_{\mathbf{z}_{(M)}} - \mathbf{I}_M\right)\tfrac{\left(\mathbf{I}_M \otimes \left(\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\right)\right)\left(\mathbf{d}_{m'} \otimes \mathbf{I}_P\right)}{\sigma_{\boldsymbol{\theta}}} \\
\mathbf{C}_{\boldsymbol{\theta}_{[m]},\boldsymbol{\theta}_{[m']}} &= \mathbf{d}_m^T\mathbf{d}_{m'}\mathbf{C}_{\boldsymbol{\theta}} + \tfrac{\mathbf{d}_m^T \otimes \mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}}{\sigma_{\boldsymbol{\theta}}}\left(\mathbf{C}_{\mathbf{z}_{(M)}} - \mathbf{I}_M\right)\tfrac{\mathbf{d}_{m'} \otimes \mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}}{\sigma_{\boldsymbol{\theta}}} \\
\mathbf{C}_{\boldsymbol{\theta}_{[m]},\boldsymbol{\theta}_{[m']}} &= \mathbf{d}_m^T\mathbf{d}_{m'}\mathbf{C}_{\boldsymbol{\theta}} + \tfrac{\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}}{\sigma_{\boldsymbol{\theta}}}\left(\mathbf{d}_m^T\left(\mathbf{C}_{\mathbf{z}_{(M)}} - \mathbf{I}_M\right)\mathbf{d}_{m'}\right)\tfrac{\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}}{\sigma_{\boldsymbol{\theta}}}
\end{aligned}$$

Finally:

$$E\left[\boldsymbol{\theta}_{[m]}\right] = \boldsymbol{\mu}_{\boldsymbol{\theta}} + \frac{\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}}{\sqrt{\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}}}E\left[\mathbf{z}_{(m)}\right] \tag{21}$$

$$\mathbf{C}_{\boldsymbol{\theta}_{[m]},\boldsymbol{\theta}_{[m']}} = \delta_m^{m'}\mathbf{C}_{\boldsymbol{\theta}} + \frac{\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}}{\mathbf{a}^T\mathbf{C}_{\boldsymbol{\theta}}\mathbf{a}}\left(\mathbf{d}_m^T\mathbf{C}_{\mathbf{z}_{(M)}}\mathbf{d}_{m'} - \delta_m^{m'}\right) \tag{22}$$

which include [4, (1)(2)(17)].

5. REFERENCES

[1] H. A. David and H. N. Nagaraja, editors. *Order Statistics (3rd Ed.)*, Wiley, New York, 2003

[2] E. Chaumette, F. Vincent, O. Besson, ”On ordered normally distributed vector parameter estimates”, Signal Processing (Elsevier) 115: 20-26, 2015

[3] David, H. A., Concomitants of order statistics. Bull. Inst. Internat. Statist. 45(1):295-300, 1973

[4] R. Song and J. A. Deddens, ”A note on moments of variables summing to normal order statistics”, Statistics & Probability Letters, 17(5):337-341, 1993

[5] W.B. Yeo and H.A. David, ”Selection through an associated characteristic with application to the random effects model”, Journal of American Statistical Association, 79: 399-405, 1984

[6] E. Castillo, *Extreme Value Theory in Engineering*, Academic Press, New York, 1988

[7] S.G. Coles and A. J. Tawn, ”Statistical methods for multivariate extremes: an application to structural design”, Applied Statistics, 43: 1-48, 1994

[8] K. Wang, ’”On concomitants of order statistics”, Phd Thesis, The Ohio State University, 2008.

[9] A. Sheikhi, M. Tata, ”The exact joint distribution of concomitants of order statistics and their order statistics under normality”, Revstat - Statistical Journal, 11(2): 121-134, 2013

[10] S. Eryilmaz, “Concomitants in a Sequence of Independent Nonidentically Distributed Random Vectors”, Communications in Statistics: Theory and Methods, 34:1925-1933, 2005

[11] H. L. Van Trees, *Optimum Array Processing*, Wiley-Interscience, New-York, 2002.

[12] E. Chaumette and F. Vincent , ”On ordered normally distributed vector parameter estimates: the general case”, in preparation for IEEE Trans. on SP.

[13] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946

[14] E. L. Lehmann and G. Casella, *Theory of Point Estimation (2nd ed.)*. Springer, 1998

[15] P. Stoica, A. Nehorai, ”Performances study of conditional and unconditional direction of arrival estimation”, IEEE Trans. on SP, 38(10): 1783-1795, 1990

[16] M. Viberg and B. Ottersten, ”Sensor array processing based on subspace fitting”, IEEE Trans. on SP, 39(5): 1110-1121, 1991

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

[18] J. Li and R. T. Compton, ”Maximum likelihood angle estimation for signals with known waveforms”, IEEE Trans. on SP, 41(9): 2850-2862, 1993

[19] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal, ”On the high snr cml estimator full statistical characterization”, IEEE Trans. on SP, 54(12): 4840-4843, 2006

[20] N Levanon, E. Mozeson, *Radar Signals*, Wiley-Interscience 2004

[21] S. F. Yau and Y. Bresler, ”A Compact Cramér-Rao Bound Expression for Parametric Estimation of Superimposed Signals”, IEEE Trans. on SP, vol. 40, no. 5, pp 1226-1230, May 1992.

[22] T. Menni, E. Chaumette, P. Larzabal, ”Reparameterization and constraints for CRB: duality and a major inequality for system analysis and design in the asymptotic region”, in Proc. IEEE ICASSP, 2012

[23] R. Kalman, ”A new approach to linear filtering and prediction problems”, ASME Journal of Basic Engineering, 82: 35-45, 1960

[24] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, Upper Saddle River, New Jersey, 2000

[25] D. Simon, *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*, Wiley InterScience, 2006