High-Order Analysis of the Efficiency Gap for Maximum Likelihood Estimation in Nonlinear Gaussian Models

Arie Yeredor, *Senior Member, IEEE*, Amir Weiss, *Student Member, IEEE*, and Anthony J. Weiss, *Fellow, IEEE*

***Abstract*—In Gaussian measurement models, the measurements are given by a known function of the unknown parameter vector, contaminated by additive zero-mean Gaussian noise. When the function is linear, the resulting maximum likelihood estimate (MLE) is well-known to be efficient [unbiased, with a mean square estimation error (MSE) matrix attaining the Cramér–Rao lower bound (CRLB)]. However, when the function is nonlinear, the MLE is only asymptotically efficient. The classical derivation of its asymptotic efficiency uses a first-order perturbation analysis, relying on a "small-errors" assumption, which under subasymptotic conditions turns inaccurate, rendering the MLE generally biased and inefficient. Although a more accurate (higher-order) performance analysis for such cases is of considerable interest, the associated derivations are rather involved, requiring cumbersome notations and indexing. Building on the recent assimilation of tensor computations into signal processing literature, we exploit the tensor formulation of higher-order derivatives to derive a tractable formulation of a higher (up to third-) order perturbation analysis, predicting the bias and MSE matrix of the MLE of parameter vectors in general nonlinear models under subasymptotic conditions. We provide explicit expressions depending on the first three derivatives of the nonlinear measurement function, and demonstrate the resulting ability to predict the "efficiency gap" (relative excess MSE beyond the CRLB) in simulation experiments. We also provide MATLAB code for easy computation of our resulting expressions.**

***Index Terms*—Maximum likelihood estimation (MLE), efficiency, high-order performance analysis, tensor calculus.**

I. Introduction

MEASUREMENT models with additive Gaussian noise are fundamental and ubiquitous in signal processing, as well as in other engineering fields and in other sciences. A few prominent examples are classical estimation problems such as frequency estimation [1]–[5], Direction of Arrival (DOA) estimation [6]–[8] and other array-processing tasks [9], [10]. In such models the measurement vector $\boldsymbol{z} \in \mathbb{R}^M$ is related to an unknown (but deterministic) parameters vector $\boldsymbol{\theta} \in \mathbb{R}^K$ via

$$\boldsymbol{z} = \boldsymbol{g}(\boldsymbol{\theta}) + \boldsymbol{v}, \quad \boldsymbol{v} \sim N(\boldsymbol{0}, \boldsymbol{\Lambda}) \tag{1}$$

where $\boldsymbol{g}(\cdot) : \mathbb{R}^K \to \mathbb{R}^M$ is some given (known) function, and $\boldsymbol{v} \in \mathbb{R}^M$ is a multivariate, zero-mean Gaussian vector ("noise") with a known (positive definite) covariance matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{M\times M}$.

The Maximum Likelihood (ML) Estimate (MLE) $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ of $\boldsymbol{\theta}$, which coincides in this case with the Optimally Weighted Least Squares (OWLS) estimate, is well-known to be *asymptotically* efficient (unbiased, attaining the Cramér-Rao Lower Bound (CRLB)), under some mild regularity conditions [11], [12]. We define *asymptotic conditions* in this context as conditions under which all elements of $\boldsymbol{\Lambda}$ can become arbitrarily small. For example, when T independent, identically distributed (i.i.d.) measurements are observed under the same model (with the same parameters), namely $\boldsymbol{z}_t = \boldsymbol{g}(\boldsymbol{\theta}) + \boldsymbol{v}_t, t = 1, \ldots T$, where $\boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{\Lambda}_1)$ are independent noise realizations, the sufficient statistic $\boldsymbol{z} \triangleq T^{-1} \sum_t \boldsymbol{z}_t$ is similarly modeled as $\boldsymbol{z} = \boldsymbol{g}(\boldsymbol{\theta}) + \boldsymbol{v}$, where $\boldsymbol{v} \triangleq T^{-1} \sum_t \boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{\Lambda}_T)$ is the equivalent noise, with $\boldsymbol{\Lambda}_T \triangleq T^{-1}\boldsymbol{\Lambda}_1$. In this case arbitrarily large values of T imply arbitrarily small elements of $\boldsymbol{\Lambda}_T$, so asymptotic conditions are established for an arbitrarily large number of independent measurements. In general, however, asymptotic conditions are not necessarily associated with repeated i.i.d. measurements, but may relate to any situation where elements of $\boldsymbol{\Lambda}$ can become arbitrarily small due to possible arbitrary reduction in noise power.

The asymptotic efficiency property of the MLE implies that asymptotically, the estimation error $\boldsymbol{\epsilon}_{\mathrm{ML}} \triangleq \hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta} \in \mathbb{R}^K$ has zero mean and its covariance coincides with the CRLB matrix. However, under non-asymptotic conditions the MLE is generally biased, and its covariance matrix may differ from the CRLB. In fact, despite recent claims in [13], non-asymptotic efficiency is attained by the MLE *if and only if* the model $\boldsymbol{g}(\boldsymbol{\theta})$ is a *linear* function of $\boldsymbol{\theta}$, a property that we prove in Appendix A. Thus, in all non-linear cases the deviation of the

Manuscript received January 22, 2018; revised June 3, 2018; accepted July 13, 2018. Date of publication July 31, 2018; date of current version August 16, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Youngchul Sung. This work was supported in part by the Israel Science Foundation under Grant 503/15 and Grant 965/15, and in part by the Pazi Foundation. *(Corresponding author: Arie Yeredor.)*

The authors are with the Tel-Aviv University, School of Electrical Engineering, Tel-Aviv 69978, Israel (e-mail: arie@eng.tau.ac.il; amirwei2@mail.tau.ac.il; ajw@eng.tau.ac.il).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author. This includes Matlab code for calculating the derived expressions as well as for repeating most of the simulation results and performance predictions presented in this paper, and a readme file to serve as a user's guide. The material is 465 KB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2860570

1053-587X © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

MLE's Mean Square (matrix) estimation Error (MSE) from the CRLB may be small, but cannot vanish (except possibly pointwise, for enumerable specific values of $\boldsymbol{\theta}$). The difference, under non-asymptotic conditions, between the MSE of the MLE and the CRLB is sometimes termed the *efficiency gap*.

Our goal in this paper is to derive analytic expressions, leading to an explicit computation scheme (supported by Matlab code) for both the bias and MSE of the MLE under non-asymptotic conditions for general nonlinear measurements models with additive noise, thereby providing an important tool for predicting the efficiency gap and/or accuracy threshold conditions in such models. Our derivations will be based on a high-order analysis, which considers moments of the noise up to fourth order and derivatives of the nonlinear functions involved up to third order. To the best of our knowledge, previous attempts to provide similar analytic expressions were limited to the scalar case of a single parameter ($K = 1$) with a scalar MSE [14]. However, for a vector of $K > 1$ parameters (with a $K \times K$ MSE matrix) the derivations become considerably more involved. A "brute-force" attempt to simply extend the scalar case into the vector case by direct use of multiple indices would result in rather cumbersome and error-prone expressions, involving heavy summations and indexing schemes. Therefore, in order to keep our derivations tractable and our notations as compact as possible, we take advantage of the recently emerging popularity of tensors computations in signal processing (see [15] and references therein, such as [16]–[18]). We use tensor formulations for the higher-order derivatives, which, together with some custom-defined tensor operations and their basic properties, enable to obtain relatively tractable and easily manageable expressions.

The expressions we provide can readily accommodate any nonlinear (three times differentiable) function $\boldsymbol{g}(\boldsymbol{\theta})$, as long as the user can provide the function and its derivatives tensors up to third order (see our explicit definitions in the sequel). Simple Matlab code for implementing the required elementary tensor operations is provided in Appendix B. The "supplemental material" for this paper also includes the complete code for computing the predicted bias and MSE.

The paper is structured as follows. In the next section we outline our tensors notations and operations, together with some basic properties which would come in handy in our derivations. Section III sets the premises for expressing derivatives of implicitly-defined vector functions in general. The score function (gradient of the log-likelihood) for our measurement model is identified as a particular case of such an implicit function in Section IV, where we obtain the required high-order derivatives for our error analysis. At this point we are fully equipped to express the approximated estimation error vector using a Taylor series expansion in terms of the noise vector, up to third order. Then in Section V we use our results in order to express the statistics (first and second moments) of the estimation error in terms of the statistics (up to fourth-order moments) of the noise - thereby obtaining the bias and MSE expressions. We demonstrate the use of our derived expressions in three simulation examples in Section VI, where we compare our analytically-predicted bias and MSE matrix to their empirically obtained counterparts. Some concluding remarks are summarized in Section VII.

II. Preliminaries: Notations and Some Basic Properties

An N-mode (sometimes termed "order-N", "N-way", or "N-dimensional") tensor is an array of numbers indexed by N indices, generalizing vectors ($N = 1$) and matrices ($N = 2$). We denote vectors, matrices and tensors with boldface lower-case, upper-case and script letters, viz. $\boldsymbol{a} \in \mathbb{R}^{I_1}$, $\boldsymbol{A} \in \mathbb{R}^{I_1 \times I_2}$ and $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_N}$, resp. Their scalar elements are denoted with plain letters, e.g., $a(i_1)$, $A(i_1, i_2)$ and $\mathcal{A}(i_1, i_2, ...i_N)$, resp. We use $\mathbf{0}$, $\boldsymbol{O}$ and $\boldsymbol{\mathcal{O}}$ to denote all-zeros vectors, matrices and tensors, resp.

We define the following elementary operations on / between tensors, matrices and vectors (short Matlab implementations of these operations can be found in Appendix B):

- Outer Products: The outer product of two tensors (with vectors and matrices regarded as particular cases of tensors) is defined as follows: Let $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{I_1 \times \cdots I_N}$ and $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{K_1 \times \cdots K_L}$ denote an N-mode tensor and an L-mode tensor, resp. Their outer product is an $(N + L)$-mode tensor, denoted
$$\boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{P}} \circ \boldsymbol{\mathcal{Q}} \in \mathbb{R}^{I_1 \times \cdots I_N \times K_1 \times \cdots K_L}$$
$$\Rightarrow \mathcal{R}(i_1, ...i_N, k_1, ...k_L) = \mathcal{P}(i_1, ...i_N)\mathcal{Q}(k_1, ...k_L).$$
- "n-th mode" Products: We define the following products between: tensors and vectors; tensors and matrices; and tensors and tensors. Let $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{I_1 \times \cdots I_N}$ denote an N-mode tensor.
  - Multiplication along its n-th mode by a vector $\boldsymbol{a} \in \mathbb{R}^{I_n}$, denoted by $\bar{\times}_n$, yields an $(N - 1)$-mode tensor:
$$\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}} \bar{\times}_n \boldsymbol{a} \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times I_{n+1} \times \cdots I_N}$$
$$\Rightarrow \mathcal{Q}(i_1, ...i_{n-1}, i_{n+1}, ...i_N)$$
$$= \sum_{i_n=1}^{I_n} a(i_n)\mathcal{P}(i_1, ...i_n, ...i_N).$$
  - Multiplication along its n-th mode by a matrix $\boldsymbol{A} \in \mathbb{R}^{J \times I_n}$, denoted by $\times_n$, yields an N-mode tensor:
$$\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}} \times_n \boldsymbol{A} \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times J \times I_{n+1} \times \cdots I_N}$$
$$\Rightarrow \mathcal{Q}(i_1, ...i_{n-1}, j, i_{n+1}, ...i_N)$$
$$= \sum_{i_n=1}^{I_n} A(j, i_n)\mathcal{P}(i_1, ...i_n, ...i_N)$$
  - Multiplication along its n-th mode by a $(2 + L)$-mode tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{J \times I_n \times K_1 \times \cdots K_L}$, denoted by $\times_n^+$, yields an $(N + L)$-mode tensor:
$$\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}} \times_n^+ \boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times J \times I_{n+1} \times \cdots I_N \times K_1 \times \cdots K_L}$$
$$\Rightarrow \mathcal{Q}(i_1, ...i_{n-1}, j, i_{n+1}, ...i_N, k_1, ...k_L)$$
$$= \sum_{i_n=1}^{I_n} \mathcal{A}(j, i_n, k_1, ...k_L)\mathcal{P}(i_1, ...i_n, ...i_N)$$

- Inner Products:
  - The "simple" inner product between two tensors (or vectors or matrices) is defined for tensors of equal dimensions: Let $\boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{Q}} \in \mathbb{R}^{I_1 \times \cdots I_N}$ denote two N-mode tensors with equal dimensions. Their inner product is a scalar, denoted
$$r = < \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{Q}} > = \sum_{i_1, \ldots i_N} \mathcal{P}(i_1, \ldots i_N)\mathcal{Q}(i_1, \ldots i_N).$$
  - The "partial" inner product between an N-mode tensor and an L-mode tensor ($L < N$) with compatible dimensions is a generalization of the inner product, defined as follows. Let $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{I_1 \times \cdots I_N}$ and $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{I_{N-L+1} \times \cdots I_N}$ denote an N-mode tensor and an L-mode tensor, resp. Their inner product is an $(N-L)$-mode tensor, denoted
$$\boldsymbol{\mathcal{R}} = \ll \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{Q}} \gg \in \mathbb{R}^{I_1 \times \cdots I_{N-L}}$$
$$\Rightarrow \mathcal{R}(i_1, \ldots i_{N-L})$$
$$= \sum_{k_1, \ldots k_L} \mathcal{P}(i_1, \ldots i_{N-L}, k_1, \ldots k_L)\mathcal{Q}(k_1, \ldots k_L).$$
- Transposing: We denote a transposition of a tensor $\boldsymbol{\mathcal{P}}$ in the sense of swapping its m-th and n-th modes as $\boldsymbol{\mathcal{P}}^{\mathrm{T}(m,n)}$. The (m,n)-transpose of an N-mode tensor $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{I_1 \times \cdots I_m \times \cdots I_n \times \cdots I_N}$ is also an N-mode tensor,
$$\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}}^{\mathrm{T}(m,n)} = \boldsymbol{\mathcal{P}}^{\mathrm{T}(n,m)} \in \mathbb{R}^{I_1 \times \cdots I_n \times \cdots I_m \times \cdots I_N}$$
$$\Rightarrow \mathcal{Q}(i_1, \ldots i_n, \ldots i_m, \ldots i_N) = \mathcal{P}(i_1, \ldots i_m, \ldots i_n, \ldots i_N).$$
  For abbreviated notation, we denote the sums of a tensor $\boldsymbol{\mathcal{P}}$ with differently transposed version thereof as (for example):
$$(\boldsymbol{\mathcal{P}})^{+\begin{bmatrix}(2,3)\\(4,5)\end{bmatrix}} \triangleq \boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{P}}^{\mathrm{T}(2,3)} + \boldsymbol{\mathcal{P}}^{\mathrm{T}(4,5)}.$$
- Permutation: Generalizing simple transposition, we denote an arbitrary permutation of the modes of a tensor as $\boldsymbol{\mathcal{P}}^{[n_1, \ldots n_N]}$, which is also an N-mode tensor,
$$\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}}^{[n_1, \ldots n_N]} \in \mathbb{R}^{I_{n_1} \times \cdots I_{n_N}}$$
$$\Rightarrow \mathcal{Q}(i_{n_1}, i_{n_2}, \ldots i_{n_N}) = \mathcal{P}(i_1, i_2, \ldots i_N).$$
  For example, let $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{3 \times 2 \times 5 \times 7}$ and let $\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{P}}^{[3,1,2,4]}$. Then $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{5 \times 3 \times 2 \times 7}$, with $\mathcal{Q}(i,j,k,\ell) = \mathcal{P}(j,k,i,\ell)$. For abbreviated notation, we shall denote the "rotate-left" permutation $\boldsymbol{\mathcal{P}}^{[2,3,\ldots N,1]}$ of an N-mode tensor as $\boldsymbol{\mathcal{P}}\uparrow$, whereas the "rotate-right" permutation $\boldsymbol{\mathcal{P}}^{[N,1,2,\ldots N-1]}$ will be denoted $\boldsymbol{\mathcal{P}}\downarrow$.
- Differentiation: For an N-mode tensor $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{I_1 \times \cdots I_N}$ whose elements depend on a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^K$, the derivative w.r.t. $\boldsymbol{\theta}$ at some $\boldsymbol{\theta}_0$ ($\boldsymbol{\theta}_0$ is implicit in the notation) is denoted as the $(N+1)$-mode tensor
$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\theta}} \triangleq \left.\frac{\partial \boldsymbol{\mathcal{P}}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \in \mathbb{R}^{I_1 \times \cdots I_N \times K}$$
$$\Rightarrow \mathcal{P}_{\boldsymbol{\theta}}(i_1, \ldots i_N, k) = \left.\frac{\partial}{\partial \theta(k)}\mathcal{P}(i_1, \ldots i_N)\right|_{\boldsymbol{\theta}_0}.$$
  Note that the notation propagates for higher-order differentiation, so that, for example, given some $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}) \in \mathbb{R}^L$, $\boldsymbol{z} \in \mathbb{R}^M$, $\boldsymbol{\theta} \in \mathbb{R}^K$, we have, e.g.,
$$\boldsymbol{F}_{\boldsymbol{\theta}} = \left.\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta})\right|_{\boldsymbol{z}_0, \boldsymbol{\theta}_0} \in \mathbb{R}^{L \times K}$$
$$\boldsymbol{F}_{\boldsymbol{z}} = \left.\frac{\partial}{\partial \boldsymbol{z}}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta})\right|_{\boldsymbol{z}_0, \boldsymbol{\theta}_0} \in \mathbb{R}^{L \times M}$$
$$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} = \left.\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta})\right|_{\boldsymbol{z}_0, \boldsymbol{\theta}_0} \in \mathbb{R}^{L \times K \times K}$$
$$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} = \left.\frac{\partial^2}{\partial \boldsymbol{z} \partial \boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta})\right|_{\boldsymbol{z}_0, \boldsymbol{\theta}_0} \in \mathbb{R}^{L \times K \times M}$$
$$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}} = \left.\frac{\partial^3}{\partial \boldsymbol{z} \partial \boldsymbol{z} \partial \boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta})\right|_{\boldsymbol{z}_0, \boldsymbol{\theta}_0} \in \mathbb{R}^{L \times K \times M \times M}$$
  and so on. Obviously, we also have, e.g., $\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta z}} = \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}}^{\mathrm{T}(2,3)}$.

The following properties, which will be used in the sequel, can be easily verified (assume compatible dimensions):

P1:
$$\boldsymbol{A} \times_1 \boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$$
P2:
$$\boldsymbol{A} \times_2 \boldsymbol{B} = \boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}$$
P3:
$$\boldsymbol{A} \times_2^+ \boldsymbol{\mathcal{B}} = (\boldsymbol{\mathcal{B}} \times_2 \boldsymbol{A})^{\mathrm{T}(1,2)}$$
P4:
$$\boldsymbol{A} \times_2^+ \boldsymbol{\mathcal{B}}^{\mathrm{T}(2,n)} = (\boldsymbol{\mathcal{B}} \times_n \boldsymbol{A})^{\mathrm{T}(1,n)}$$
P5:
$$\boldsymbol{A} \times_2^+ \boldsymbol{\mathcal{B}}^{\mathrm{T}(1,2)} = \boldsymbol{\mathcal{B}} \times_1 \boldsymbol{A}$$
P6:
$$\boldsymbol{\mathcal{A}} \times_n \boldsymbol{B} = (\boldsymbol{\mathcal{A}}^{\mathrm{T}(m,n)} \times_m \boldsymbol{B})^{\mathrm{T}(m,n)}$$
P7:
$$\boldsymbol{\mathcal{A}} \times_m \boldsymbol{B} \times_n \boldsymbol{C} = \begin{cases} \boldsymbol{\mathcal{A}} \times_n \boldsymbol{C} \times_m \boldsymbol{B} & m \neq n \\ \boldsymbol{\mathcal{A}} \times_m (\boldsymbol{C}\boldsymbol{B}) & m = n \end{cases}$$
P8:
$$\boldsymbol{\mathcal{A}} \times_m \boldsymbol{B} \times_n^+ \boldsymbol{\mathcal{C}} = \begin{cases} \boldsymbol{\mathcal{A}} \times_n^+ \boldsymbol{\mathcal{C}} \times_m \boldsymbol{B} & m \neq n \\ \boldsymbol{\mathcal{A}} \times_m (\boldsymbol{B} \times_1^+ \boldsymbol{\mathcal{C}}) & m = n \end{cases}$$
P9:
$$\boldsymbol{\mathcal{A}} \times_n \boldsymbol{B} = \boldsymbol{\mathcal{C}} \xrightarrow{\boldsymbol{B}\text{ invertible}} \boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{C}} \times_n \boldsymbol{B}^{-1}$$
P10: ($\boldsymbol{\mathcal{B}}$ is $(2+L)$-mode)
$$(\boldsymbol{A} \times_2^+ \boldsymbol{\mathcal{B}})_{\boldsymbol{\theta}} = (\boldsymbol{\mathcal{A}}_{\boldsymbol{\theta}} \times_2^+ \boldsymbol{\mathcal{B}})^{\mathrm{T}(3,L+3)} + \boldsymbol{A} \times_2^+ \boldsymbol{\mathcal{B}}_{\boldsymbol{\theta}}$$
P11:
$$\boldsymbol{\mathcal{A}}^{\mathrm{T}(m,n)} \times_p \boldsymbol{B} = \begin{cases} (\boldsymbol{\mathcal{A}} \times_p \boldsymbol{B})^{\mathrm{T}(m,n)} & p \neq m,n \\ (\boldsymbol{\mathcal{A}} \times_n \boldsymbol{B})^{\mathrm{T}(m,n)} & p = m \end{cases}$$

P12:

$$\boldsymbol{\mathcal{A}}^{\mathrm{T}(m,n)} \times_p^+ \boldsymbol{\mathcal{B}} = \begin{cases} (\boldsymbol{\mathcal{A}} \times_p^+ \boldsymbol{\mathcal{B}})^{\mathrm{T}(m,n)} & p \neq m, n \\ (\boldsymbol{\mathcal{A}} \times_n^+ \boldsymbol{\mathcal{B}})^{\mathrm{T}(m,n)} & p = m \end{cases}$$

P13:

$$\boldsymbol{\mathcal{A}} \times_n \boldsymbol{B} \times_m^+ \boldsymbol{\mathcal{C}} \overset{m \neq n}{=} \left(\boldsymbol{\mathcal{A}}^{\mathrm{T}(m,n)} \times_n^+ \boldsymbol{\mathcal{C}} \times_m \boldsymbol{B}\right)^{\mathrm{T}(m,n)}$$

P14: ($\boldsymbol{\mathcal{A}}$ is N-mode)

$$(\boldsymbol{\mathcal{A}} \bar{\times}_n \boldsymbol{b})_{\boldsymbol{\theta}} = \boldsymbol{\mathcal{A}}_{\boldsymbol{\theta}} \bar{\times}_n \boldsymbol{b} + (\boldsymbol{\mathcal{A}} \times_n \boldsymbol{B}_{\boldsymbol{\theta}}^{\mathrm{T}})^{\mathrm{T}(n,N)}$$

P15: ($\boldsymbol{C}$ does not depend on $\boldsymbol{\theta}$)

$$(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{B})_{\boldsymbol{\theta}} = (\boldsymbol{\mathcal{A}}_{\boldsymbol{\theta}} \times_1 (\boldsymbol{C}\boldsymbol{B})^{\mathrm{T}})^{\mathrm{T}(1,2)} + \boldsymbol{\mathcal{B}}_{\boldsymbol{\theta}} \times_1 (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{C})$$

P16: ($\boldsymbol{\mathcal{A}}$ is N-mode, $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{C}$ do not depend on $\boldsymbol{\theta}$)

$$(\boldsymbol{\mathcal{A}} \bar{\times}_1 (\boldsymbol{C}\boldsymbol{\theta}))_{\boldsymbol{\theta}} = (\boldsymbol{\mathcal{A}} \uparrow) \times_N \boldsymbol{C}^{\mathrm{T}}$$

P17:

$$(\boldsymbol{\mathcal{A}} \times_n \boldsymbol{B})_{\boldsymbol{\theta}} = \boldsymbol{\mathcal{A}}_{\boldsymbol{\theta}} \times_n \boldsymbol{B} + \boldsymbol{\mathcal{A}} \times_n^+ \boldsymbol{\mathcal{B}}_{\boldsymbol{\theta}}$$

P18: ($\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{B}}. \boldsymbol{\mathcal{Q}}$ are (resp.) $N, (N-1), L, (L-1)$-mode)

$$\ll \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}} \gg \circ \ll \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{Q}} \gg = \ll (\boldsymbol{\mathcal{A}} \circ \boldsymbol{\mathcal{B}})^{\mathrm{T}(2,N+1)}, \boldsymbol{\mathcal{P}} \circ \boldsymbol{\mathcal{Q}} \gg$$

III. Derivatives of Implicitly Defined Functions

Let $\boldsymbol{z} \in \mathbb{R}^M$ and $\boldsymbol{\theta} \in \mathbb{R}^K$ be two vectors, and let $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}) \in \mathbb{R}^K$ denote some continuous function thereof, three times differentiable w.r.t. both $\boldsymbol{z}$ and $\boldsymbol{\theta}$. An implicit relation between $\boldsymbol{\theta}$ and $\boldsymbol{z}$ is defined through the requirement

$$\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}) = \boldsymbol{0}, \tag{2}$$

namely, there exists some (implicitly defined) function $\boldsymbol{\theta}(\boldsymbol{z}) : \mathbb{R}^M \to \mathbb{R}^K$, such that $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}(\boldsymbol{z})) = \boldsymbol{0}$ for all $\boldsymbol{z} \in \mathbb{R}^M$ (see [19]). Additionally, assume that some $\boldsymbol{z}_0 \in \mathbb{R}^M$ and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\boldsymbol{z}_0) \in \mathbb{R}^K$ are known, such that $\boldsymbol{f}(\boldsymbol{z}_0, \boldsymbol{\theta}_0) = \boldsymbol{0}$.

Our goal is to find derivatives (up to third-order) of $\boldsymbol{\theta}(\boldsymbol{z})$ at $\boldsymbol{z}_0$. We begin with the first-order derivative matrix $\boldsymbol{\Theta}_{\boldsymbol{z}} \in \mathbb{R}^{K \times M}$ by observing the first full derivative of $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}(\boldsymbol{z}))$ w.r.t. $\boldsymbol{z}$ at $\boldsymbol{z}_0$:

$$\left.\frac{d\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}(\boldsymbol{z}))}{d\boldsymbol{z}}\right|_{\boldsymbol{z}_0} = \boldsymbol{F}_{\boldsymbol{z}} + \boldsymbol{F}_{\boldsymbol{\theta}} \boldsymbol{\Theta}_{\boldsymbol{z}} \in \mathbb{R}^{K \times M}, \tag{3}$$

where $\boldsymbol{F}_{\boldsymbol{z}} \in \mathbb{R}^{K \times M}$ and $\boldsymbol{F}_{\boldsymbol{\theta}} \in \mathbb{R}^{K \times K}$ denote the partial derivatives of $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta})$ w.r.t. $\boldsymbol{z}$ and $\boldsymbol{\theta}$ (resp.) at $\boldsymbol{z} = \boldsymbol{z}_0$, $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Since $\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta})$ is constant (zero) on the manifold $\boldsymbol{\theta}(\boldsymbol{z})$, its full derivative w.r.t. $\boldsymbol{z}$ along this manifold should equal zero as well. Thus, equating the derivative to zero we obtain

$$\boxed{\boldsymbol{\Theta}_{\boldsymbol{z}} = -\boldsymbol{F}_{\boldsymbol{\theta}}^{-1} \boldsymbol{F}_{\boldsymbol{z}} \in \mathbb{R}^{K \times M}.} \tag{4}$$

Proceeding to the second order, we take the full derivative of (3) w.r.t. $\boldsymbol{z}$, obtaining

$$\begin{aligned} \left.\frac{d^2\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}(\boldsymbol{z}))}{d\boldsymbol{z}d\boldsymbol{z}}\right|_{\boldsymbol{z}_0} &= \boldsymbol{\mathcal{F}}_{\boldsymbol{zz}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \\ &+ (\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}) \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \\ &+ \boldsymbol{F}_{\boldsymbol{\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \in \mathbb{R}^{K \times M \times M}. \end{aligned} \tag{5}$$

Equating this derivative to zero and noting that, by property P5, $\boldsymbol{F}_{\boldsymbol{\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} = \boldsymbol{\Theta}_{\boldsymbol{zz}} \times_1 \boldsymbol{F}_{\boldsymbol{\theta}}$, we get

$$\begin{aligned} \boldsymbol{\Theta}_{\boldsymbol{zz}} \times_1 \boldsymbol{F}_{\boldsymbol{\theta}} = -\big[&\boldsymbol{\mathcal{F}}_{\boldsymbol{zz}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \\ &\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\big]. \end{aligned} \tag{6}$$

By property P6, we have, using $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} = \boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta}}^{\mathrm{T}(2,3)}$:

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} = (\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}})^{\mathrm{T}(2,3)}, \tag{7}$$

therefore by property P9 we get

$$\boxed{\begin{aligned} \boldsymbol{\Theta}_{\boldsymbol{zz}} = -\Big[&\boldsymbol{\mathcal{F}}_{\boldsymbol{zz}} + (\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}})^{+[(2,3)]} \\ &+ \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\Big] \times_1 \boldsymbol{F}_{\boldsymbol{\theta}}^{-1} \in \mathbb{R}^{K \times M \times M}. \end{aligned}} \tag{8}$$

Likewise, for the third order we take the full derivative of (5) w.r.t. $\boldsymbol{z}$, obtaining

$$\begin{aligned} \left.\frac{d^3\boldsymbol{f}(\boldsymbol{z}, \boldsymbol{\theta}(\boldsymbol{z}))}{d\boldsymbol{z}d\boldsymbol{z}d\boldsymbol{z}}\right|_{\boldsymbol{z}_0} &= \boldsymbol{\mathcal{F}}_{\boldsymbol{zzz}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{zz\theta}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \\ &+ (\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta z}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta\theta}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}) \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta}} \times_3^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \\ &+ (\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z\theta}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}) \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \\ &+ (\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}) \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \\ &+ \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \\ &+ \left((\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}) \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}\right)^{\mathrm{T}(3,4)} \\ &+ \boldsymbol{F}_{\boldsymbol{\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zzz}}^{\mathrm{T}(1,2)} \end{aligned} \tag{9}$$

(where we have used P10 for the penultimate term). Expanding the brackets, we identify the following groups of like terms:

G1:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{zzz}}$
G2:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{zz\theta}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$ $\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta z}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$ $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$
G3:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta\theta}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$ $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$ $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$
G4:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$
G5:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta}} \times_3^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}$ $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}$ $\left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}\right)^{\mathrm{T}(3,4)}$
G6:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$ $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}$ $\left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}\right)^{\mathrm{T}(3,4)}$

These can all be rewritten more compactly as

G1:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{zzz}}$
G2:	$\left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,3)\\(2,4)\end{smallmatrix}\right]}$
G3:	$\left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,4)\\(3,4)\end{smallmatrix}\right]}$
G4:	$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}$
G5:	$\left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}\right)^{+\left[\begin{smallmatrix}(2,3)\\(3,4)\end{smallmatrix}\right]}$
G6:	$\left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,3)\\(3,4)\end{smallmatrix}\right]}$

where we've used the following:

- For G2: P11 and derivatives permutations in tensors, namely $\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta z}} = \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}}^{\mathrm{T}(2,3)}$ and $\boldsymbol{\mathcal{F}}_{\boldsymbol{zz\theta}} = \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}}^{\mathrm{T}(2,4)}$;
- For G3: P7, P11 and derivatives permutations;
- For G5: P12 and derivatives permutation; and
- For G6: P8, P13 and the symmetry $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}}^{\mathrm{T}(2,3)} = \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}}$.

Collecting these six terms, noting that by P5

$$\boldsymbol{F}_{\boldsymbol{\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zzz}}^{\mathrm{T}(1,2)} = \boldsymbol{\Theta}_{\boldsymbol{zzz}} \times_1 \boldsymbol{F}_{\boldsymbol{\theta}}, \tag{10}$$

and using P9 once again, we obtain, when equating this derivative to zero,

$$\boxed{\begin{aligned}
\boldsymbol{\Theta}_{\boldsymbol{zzz}} = -\Bigg[&\boldsymbol{\mathcal{F}}_{\boldsymbol{zzz}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_4 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\\
&+ \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,4)\\(3,4)\end{smallmatrix}\right]}\\
&+ \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}} \times_2 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,3)\\(2,4)\end{smallmatrix}\right]}\\
&+ \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}\right)^{+\left[\begin{smallmatrix}(2,3)\\(3,4)\end{smallmatrix}\right]}\\
&+ \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2^+ \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \times_3 \boldsymbol{\Theta}_{\boldsymbol{z}}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,3)\\(3,4)\end{smallmatrix}\right]}\Bigg] \times_1 \boldsymbol{F}_{\boldsymbol{\theta}}^{-1}\\
&\in \mathbb{R}^{K\times M\times M\times M}.
\end{aligned}} \tag{11}$$

Note that for any particular function $\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})$, the calculation of the first three derivatives of the implied $\boldsymbol{\theta}(\boldsymbol{z})$ w.r.t. $\boldsymbol{z}$ at $\boldsymbol{z}_0$, $\boldsymbol{\Theta}_{\boldsymbol{z}}$, $\boldsymbol{\Theta}_{\boldsymbol{zz}}$ and $\boldsymbol{\Theta}_{\boldsymbol{zzz}}$, can be readily obtained from (4), (8) and (11) (resp.), based on the partial derivatives (at $\boldsymbol{z}=\boldsymbol{z}_0, \boldsymbol{\theta}=\boldsymbol{\theta}_0$) of $\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})$, namely on $\boldsymbol{F}_{\boldsymbol{\theta}}, \boldsymbol{F}_{\boldsymbol{z}}, \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}}, \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}}, \boldsymbol{\mathcal{F}}_{\boldsymbol{zz}}, \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}}, \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}}$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}}$ and $\boldsymbol{\mathcal{F}}_{\boldsymbol{zzz}}$.

Once these three derivative tensors are computed, we can express the Taylor series expansion of $\boldsymbol{\theta}(\boldsymbol{z})$ in the vicinity of $(\boldsymbol{z}_0, \boldsymbol{\theta}_0)$ up to third order as

$$\begin{aligned}
\boldsymbol{\theta}(\boldsymbol{z}) = \boldsymbol{\theta}_0 + &\ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{\delta z} \gg + \frac{1}{2!} \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, \boldsymbol{\delta z} \circ \boldsymbol{\delta z} \gg\\
&+ \frac{1}{3!} \ll \boldsymbol{\Theta}_{\boldsymbol{zzz}}, \boldsymbol{\delta z} \circ \boldsymbol{\delta z} \circ \boldsymbol{\delta z} \gg + \mathcal{O}(\boldsymbol{\delta z}^4),
\end{aligned} \tag{12}$$

where $\boldsymbol{\delta z} \triangleq \boldsymbol{z} - \boldsymbol{z}_0$.

We now proceed to our particular case, in which $\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})$ is determined by the likelihood equations in a nonlinear measurements model with additive zero-mean Gaussian noise.

IV. Likelihood Equations as a Particular Case

Recalling our measurements model (1), the log-likelihood is clearly given by

$$\mathfrak{L}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{1}{2}\left[\log|2\pi\boldsymbol{\Lambda}| + (\boldsymbol{z}-\boldsymbol{g}(\boldsymbol{\theta}))^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{z}-\boldsymbol{g}(\boldsymbol{\theta}))\right], \tag{13}$$

where $|\cdot|$ denotes the determinant. We assume that $M \geq K$ (the number of measurements is not smaller than the number of unknown parameters) and, moreover, that $\boldsymbol{G}_{\boldsymbol{\theta}} \in \mathbb{R}^{M\times K}$, the first derivative matrix of $\boldsymbol{g}(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$, has full column rank. The likelihood equations, requiring a null gradient of $\mathfrak{L}(\boldsymbol{z},\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$, define our implicit function

$$\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta}) \triangleq \frac{\partial^{\mathrm{T}}}{\partial\boldsymbol{\theta}}\mathfrak{L}(\boldsymbol{z};\boldsymbol{\theta}) = \boldsymbol{G}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{z}-\boldsymbol{g}(\boldsymbol{\theta})) \in \mathbb{R}^K, \tag{14}$$

so that the MLE $\hat{\boldsymbol{\theta}}$ should satisfy $\boldsymbol{f}(\boldsymbol{z},\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$.

Note that throughout this section we wish (for convenience of notations) to distinguish the general derivatives of $\boldsymbol{g}(\boldsymbol{\theta})$ from those obtained when substituting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. We shall therefore denote the former as $\boldsymbol{G}_{\boldsymbol{\theta}}$, $\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}$ and $\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}}$, and the latter as $\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}}$, $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}}$ and $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}$, resp. For derivatives of $\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})$ we shall use the unabbreviated partial derivative notation for the general expressions, and $\boldsymbol{F}_{\boldsymbol{\theta}}$, $\boldsymbol{F}_{\boldsymbol{z}}$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}}$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}}$, etc. for the specific value obtained at $\boldsymbol{z}_0$ and $\boldsymbol{\theta}_0$, for compatibility with the previous section.

In the absence of noise we have, for each $\boldsymbol{\theta}_0$, a value $\boldsymbol{z}_0 \triangleq \boldsymbol{g}(\boldsymbol{\theta}_0)$, so that $\boldsymbol{f}(\boldsymbol{z}_0,\boldsymbol{\theta}_0) = \boldsymbol{0}$. When the noise causes perturbations in the measurements around $\boldsymbol{z}_0$, the MLE $\hat{\boldsymbol{\theta}}$ deviates accordingly from $\boldsymbol{\theta}_0$, while maintaining $\boldsymbol{f}(\boldsymbol{z}_0+\boldsymbol{v},\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$. By using our Taylor series expansion (12) we wish to approximate the resulting deviation (estimation error) $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ in terms of the noise $\boldsymbol{v}$, up to third order. We would then use the noise statistics (moments) to obtain the statistics (moments) of the approximated estimation error.

To this end, we need to calculate the required derivatives of $\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})$. Starting with the first derivative w.r.t. $\boldsymbol{\theta}$, we have

$$\frac{\partial \boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}} \bar{\times}_1 \left(\boldsymbol{\Lambda}^{-1}(\boldsymbol{z}-\boldsymbol{g}(\boldsymbol{\theta}))\right) - \boldsymbol{G}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}}. \tag{15}$$

Substituting $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\boldsymbol{z} = \boldsymbol{z}_0$ we get

$$\boldsymbol{F}_{\boldsymbol{\theta}} = -\overset{o}{\boldsymbol{G}}{}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}} \in \mathbb{R}^{K\times K}. \tag{16}$$

Differentiating (14) w.r.t. $\boldsymbol{z}$ we get

$$\frac{\partial \boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{z}} = \boldsymbol{G}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}, \tag{17}$$

so that

$$\boldsymbol{F}_{\boldsymbol{z}} = \overset{o}{\boldsymbol{G}}{}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1} \in \mathbb{R}^{K\times M}, \tag{18}$$

and, consequently,

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{zz}} = \boldsymbol{\mathcal{O}} \in \mathbb{R}^{K\times M\times M} \tag{19}$$

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{zzz}} = \boldsymbol{\mathcal{O}} \in \mathbb{R}^{K\times M\times M\times M} \tag{20}$$

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{zz\theta}} = \boldsymbol{\mathcal{O}} \in \mathbb{R}^{K\times M\times M\times K} \tag{21}$$

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{z\theta z}} = \boldsymbol{\mathcal{O}} \in \mathbb{R}^{K\times M\times K\times M} \tag{22}$$

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta zz}} = \boldsymbol{\mathcal{O}} \in \mathbb{R}^{K\times K\times M\times M}. \tag{23}$$

Proceeding to $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}}$, we further differentiate (15) w.r.t. $\boldsymbol{\theta}$, obtaining

$$\begin{aligned}\frac{\partial^2 \boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial\boldsymbol{\theta}} &= \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}} \bar{\times}_1 \left(\boldsymbol{\Lambda}^{-1}(\boldsymbol{z}-\boldsymbol{g}(\boldsymbol{\theta}))\right) - (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{\mathrm{T}})^{\mathrm{T}(1,3)} \\ &\quad - (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{\mathrm{T}})^{\mathrm{T}(1,2)} - \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{\mathrm{T}},\end{aligned} \tag{24}$$

where we have used P14 for the first two resulting terms, P15 for the last two (with $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{G}_{\boldsymbol{\theta}}$ and $\boldsymbol{C} = \boldsymbol{\Lambda}^{-1}$) and $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{\mathrm{T}}$. Substituting $\boldsymbol{\theta}_0$ and $\boldsymbol{z}_0$ we get

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} = -\left(\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}})^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(1,2)\\(1,3)\end{smallmatrix}\right]} \in \mathbb{R}^{K\times K\times K}. \tag{25}$$

For $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}}$ we differentiate (15) w.r.t. $\boldsymbol{z}$. Using P16 we get

$$\frac{\partial^2 \boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial\boldsymbol{z}} = (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}\uparrow) \times_3 \boldsymbol{\Lambda}^{-1}, \tag{26}$$

so that

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} = (\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}}\uparrow) \times_3 \boldsymbol{\Lambda}^{-1} \in \mathbb{R}^{K\times K\times M}. \tag{27}$$

Proceeding to third-order derivatives, we need to differentiate (24) w.r.t. $\boldsymbol{\theta}$:

$$\begin{aligned}&\frac{\partial^3 \boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}} \\ &\quad = \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta\theta}} \bar{\times}_1 \left(\boldsymbol{\Lambda}^{-1}(\boldsymbol{z}-\boldsymbol{g}(\boldsymbol{\theta}))\right) - (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{\mathrm{T}})^{\mathrm{T}(1,4)} \\ &\qquad - (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{\mathrm{T}} + \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}} \times_1^+ (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}^{\mathrm{T}(1,2)} \times_2 \boldsymbol{\Lambda}^{-1}))^{\mathrm{T}(1,3)} \\ &\qquad - (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{\mathrm{T}} + \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}} \times_1^+ (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}^{\mathrm{T}(1,2)} \times_2 \boldsymbol{\Lambda}^{-1}))^{\mathrm{T}(1,2)} \\ &\qquad - \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{\mathrm{T}} + \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}} \times_1^+ (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}^{\mathrm{T}(1,2)} \times_2 \boldsymbol{\Lambda}^{-1}).\end{aligned} \tag{28}$$

where for the first two resulting terms we used P14, and for subsequent terms we used P17, P15 (with $\boldsymbol{A} = \boldsymbol{G}_{\boldsymbol{\theta}}$, $\boldsymbol{C} = \boldsymbol{\Lambda}^{-1}$ and $\boldsymbol{B} = \boldsymbol{I}$, the Identity matrix) and P11. Combining like terms and substituting $\boldsymbol{\theta}_0$ and $\boldsymbol{z}_0$ we get

$$\begin{aligned}\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}} &= -\left(\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}} \times_1 (\boldsymbol{\Lambda}^{-1}\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}})^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(1,2)\\(1,3)\\(1,4)\end{smallmatrix}\right]} \\ &\quad - \left(\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}} \times_1^+ \left(\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}}{}^{\mathrm{T}(1,2)} \times_2 \boldsymbol{\Lambda}^{-1}\right)\right)^{+\left[\begin{smallmatrix}(1,2)\\(1,3)\end{smallmatrix}\right]} \in \mathbb{R}^{K\times K\times K\times K}.\end{aligned} \tag{29}$$

The last missing term is $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}}$, which we readily obtain by differentiating (24) w.r.t. $\boldsymbol{z}$ (using P16):

$$\frac{\partial^3 \boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial\boldsymbol{\theta}\partial\boldsymbol{z}} = (\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}}\uparrow) \times_4 \boldsymbol{\Lambda}^{-1}, \tag{30}$$

so that

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}} = \left(\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}\uparrow\right) \times_4 \boldsymbol{\Lambda}^{-1} \in \mathbb{R}^{K\times K\times K\times M}. \tag{31}$$

Due to some of the vanishing derivatives in this particular case (where $\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})$ is given by (14)), some of the expressions for the derivatives of $\boldsymbol{\theta}(\boldsymbol{z})$ (at $\boldsymbol{\theta}_0$, $\boldsymbol{z}_0$) derived in the previous section for a general $\boldsymbol{f}(\boldsymbol{z},\boldsymbol{\theta})$, are somewhat simplified. In addition, to simplify our notations, we may define two useful matrices:

$$\boldsymbol{B} \triangleq -\boldsymbol{F}_{\boldsymbol{\theta}}^{-1} = \left(\overset{o}{\boldsymbol{G}}{}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}}\right)^{-1} \in \mathbb{R}^{K\times K}, \tag{32}$$

which is the CRLB; and

$$\boldsymbol{H} \triangleq -\boldsymbol{F}_{\boldsymbol{\theta}}^{-1}\boldsymbol{F}_{\boldsymbol{z}} = \boldsymbol{B}\overset{o}{\boldsymbol{G}}{}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1} \in \mathbb{R}^{K\times M}. \tag{33}$$

We then obtain the following expressions for somewhat simplified versions of (4), (8) and (11), resp.:

$$\boxed{\boldsymbol{\Theta}_{\boldsymbol{z}} = \boldsymbol{H} \in \mathbb{R}^{K\times M},} \tag{34}$$

$$\boxed{\begin{aligned}\boldsymbol{\Theta}_{\boldsymbol{zz}} = &-\Big[(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2 \boldsymbol{H}^{\mathrm{T}})^{+[(2,3)]} \\ &+ \boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2 \boldsymbol{H}^{\mathrm{T}} \times_3 \boldsymbol{H}^{\mathrm{T}}\Big] \times_1 \boldsymbol{B} \in \mathbb{R}^{K\times M\times M},\end{aligned}} \tag{35}$$

$$\boxed{\begin{aligned}\boldsymbol{\Theta}_{\boldsymbol{zzz}} = &-\Bigg[\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}} \times_2 \boldsymbol{H}^{\mathrm{T}} \times_3 \boldsymbol{H}^{\mathrm{T}} \times_4 \boldsymbol{H}^{\mathrm{T}} \\ &+ \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}} \times_2 \boldsymbol{H}^{\mathrm{T}} \times_3 \boldsymbol{H}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,4)\\(3,4)\end{smallmatrix}\right]} \\ &+ \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}} \times_2^{+} \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)}\right)^{+\left[\begin{smallmatrix}(2,3)\\(3,4)\end{smallmatrix}\right]} \\ &+ \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}} \times_2^{+} \boldsymbol{\Theta}_{\boldsymbol{zz}}^{\mathrm{T}(1,2)} \times_3 \boldsymbol{H}^{\mathrm{T}}\right)^{+\left[\begin{smallmatrix}(2,3)\\(3,4)\end{smallmatrix}\right]}\Bigg] \times_1 \boldsymbol{B} \\ &\in \mathbb{R}^{K\times M\times M\times M}.\end{aligned}} \tag{36}$$

The only application-specific computations required are those of the derivatives of $\boldsymbol{g}(\boldsymbol{\theta})$, namely $\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}}$, $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}}$ and $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}$. Once obtained, they can be readily plugged into the expressions for $\boldsymbol{F}_{\boldsymbol{\theta}}$, $\boldsymbol{F}_{\boldsymbol{z}}$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta}}$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta z}}$, $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta\theta}}$ and $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta\theta z}}$ (cf. (16), (18), (25), (27), (29) and (31), resp.), leading to $\boldsymbol{\Theta}_{\boldsymbol{z}}$, $\boldsymbol{\Theta}_{\boldsymbol{zz}}$ and $\boldsymbol{\Theta}_{\boldsymbol{zzz}}$. The latter three enable (via (12)) to express the ML estimation error $\boldsymbol{\epsilon} \triangleq \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ in terms of the noise $\boldsymbol{v} = \boldsymbol{z}_0 - \boldsymbol{z}$ up to third-order terms,

$$\begin{aligned}\boldsymbol{\epsilon}(\boldsymbol{v}) = &\ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{v} \gg + \frac{1}{2!} \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, \boldsymbol{v}\circ\boldsymbol{v} \gg \\ &+ \frac{1}{3!} \ll \boldsymbol{\Theta}_{\boldsymbol{zzz}}, \boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v} \gg + \boldsymbol{\mathcal{O}}(\boldsymbol{v}^4) \in \mathbb{R}^K.\end{aligned} \tag{37}$$

Note that this expression enables the (approximated) prediction of the realization-specific error of the MLE from the specific realization of the noise, assuming that the noise is "small enough" in two respects:

- The first is the common assumption, that the fourth-order and higher terms in the Taylor series expansion are negligible.
- The second has to do with the fact that the global maximizer of the likelihood is generally not a continuous function of the noise (as the noise increases, a previous global maximum may vanish and a new global maximum may pop up at a distant location), and therefore the noise should be assumed small enough so that continuity of the minimizer w.r.t. the noise (as implied by the Taylor series) is maintained.

Note further, that since the MLE for the Gaussian case is well-known to coincide with the OWLS estimate (with the inverse noise covariance as the weight matrix), this means that the realization-specific OWLS estimation error can also be predicted from the noise realization, regardless of the specific probability distribution of the noise (be it Gaussian or not).

However, when it comes to predicting statistics of the estimation error, namely its bias and MSE, the noise distribution comes into play, as analyzed in the following section.

V. Error Statistics

Consider the model (1) and let $\hat{\boldsymbol{\theta}}(\boldsymbol{z}) : \mathbb{R}^M \to \mathbb{R}^K$ denote the ML mapping $\hat{\boldsymbol{\theta}}(\boldsymbol{z}) = \arg\max_\theta \mathfrak{L}(\boldsymbol{z}, \boldsymbol{\theta})$, where the likelihood $\mathfrak{L}(\boldsymbol{z}, \boldsymbol{\theta})$ is defined in (13). Further define $\boldsymbol{\epsilon}(\boldsymbol{v}) : \mathbb{R}^M \to \mathbb{R}^K$ as the noise to estimation-error mapping, $\boldsymbol{\epsilon}(\boldsymbol{v}) \triangleq \hat{\boldsymbol{\theta}}(\boldsymbol{z}_0 + \boldsymbol{v}) - \boldsymbol{\theta}_0$ (where $\boldsymbol{z}_0 = \boldsymbol{g}(\boldsymbol{\theta}_0)$). Let $\mathfrak{B} \subseteq \mathbb{R}^M$ denote the vicinity of $\boldsymbol{v} = \boldsymbol{0}$ in which $\boldsymbol{\epsilon}(\boldsymbol{v})$ is a continuous function of $\boldsymbol{v}$. Also denote $\rho \triangleq \Pr\{\boldsymbol{v} \notin \mathfrak{B}\}$ (the probability that the noise is strong enough to drive the global maximum away from the implied vicinity of $\boldsymbol{\theta}_0$).

Assuming that $\boldsymbol{\Lambda}$ is sufficiently small, such that $\rho \ll 1$, implying $\rho E[\boldsymbol{\epsilon}^2(\boldsymbol{v})|\boldsymbol{v} \notin \mathfrak{B}] \ll (1-\rho)E[\boldsymbol{\epsilon}^2(\boldsymbol{v})|\boldsymbol{v} \in \mathfrak{B}]$ (where $\boldsymbol{\epsilon}^2(\boldsymbol{v})$ denotes the element-wise squared value, and the inequality is interpreted element-wise), we can conclude, for the sake of statistical analysis, that (37) serves as a good approximation of $\boldsymbol{\epsilon}(\boldsymbol{v})$ for "most" values of $\boldsymbol{v}$. Thus, taking its mean, we easily obtain an approximation of the bias:

$$\begin{aligned}E[\boldsymbol{\epsilon}(\boldsymbol{v})] \approx &\ll \boldsymbol{\Theta}_{\boldsymbol{z}}, E[\boldsymbol{v}] \gg + \frac{1}{2!} \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, E[\boldsymbol{v}\circ\boldsymbol{v}] \gg \\ &+ \frac{1}{3!} \ll \boldsymbol{\Theta}_{\boldsymbol{zzz}}, E[\boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v}] \gg + \boldsymbol{\mathcal{O}}(E[\boldsymbol{v}^4]) \\ = &\frac{1}{2} \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, \boldsymbol{\Lambda} \gg + \boldsymbol{\mathcal{O}}(E[\boldsymbol{v}^4]) \in \mathbb{R}^K.\end{aligned} \tag{38}$$

Note that the "standard", classical first-order analysis, which essentially ignores (neglects) the second and higher terms above, leads to zero bias - which is generally inaccurate in a nonlinear measurement model.

For the MSE we first need to express the outer product of $\boldsymbol{\epsilon}(\boldsymbol{v})$ with itself,

$$\begin{aligned}\boldsymbol{\epsilon}(\boldsymbol{v}) \circ \boldsymbol{\epsilon}(\boldsymbol{v}) \approx &\ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{v} \gg \circ \ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{v} \gg \\ &+ \frac{1}{2!} \ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{v} \gg \circ \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, \boldsymbol{v}\circ\boldsymbol{v} \gg \\ &+ \frac{1}{2!} \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, \boldsymbol{v}\circ\boldsymbol{v} \gg \circ \ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{v} \gg \\ &+ \frac{1}{3!} \ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{v} \gg \circ \ll \boldsymbol{\Theta}_{\boldsymbol{zzz}}, \boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v} \gg \\ &+ \frac{1}{3!} \ll \boldsymbol{\Theta}_{\boldsymbol{zzz}}, \boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v} \gg \circ \ll \boldsymbol{\Theta}_{\boldsymbol{z}}, \boldsymbol{v} \gg \\ &+ \frac{1}{2!}\cdot\frac{1}{2!} \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, \boldsymbol{v}\circ\boldsymbol{v} \gg \circ \ll \boldsymbol{\Theta}_{\boldsymbol{zz}}, \boldsymbol{v}\circ\boldsymbol{v} \gg \\ &+ \boldsymbol{\mathcal{O}}(\boldsymbol{v}^5) \in \mathbb{R}^{K\times K}.\end{aligned} \tag{39}$$

Using P18 and collecting terms, this product reads

$$\begin{aligned}\boldsymbol{\epsilon}(\boldsymbol{v}) \circ \boldsymbol{\epsilon}(\boldsymbol{v}) \approx &\ll (\boldsymbol{\Theta}_{\boldsymbol{z}} \circ \boldsymbol{\Theta}_{\boldsymbol{z}})^{\mathrm{T}(2,3)}, \boldsymbol{v}\circ\boldsymbol{v} \gg \\ &+ \frac{1}{2!} \ll (\boldsymbol{\Theta}_{\boldsymbol{z}} \circ \boldsymbol{\Theta}_{\boldsymbol{zz}})^{\mathrm{T}(2,3)}, \boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v} \gg \\ &+ \frac{1}{2!} \ll (\boldsymbol{\Theta}_{\boldsymbol{zz}} \circ \boldsymbol{\Theta}_{\boldsymbol{z}})^{\mathrm{T}(2,4)}, \boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v} \gg \\ &+ \frac{1}{3!} \ll (\boldsymbol{\Theta}_{\boldsymbol{z}} \circ \boldsymbol{\Theta}_{\boldsymbol{zzz}})^{\mathrm{T}(2,3)}, \boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v}\circ\boldsymbol{v} \gg\end{aligned}$$

$$+\frac{1}{3!} \ll (\boldsymbol{\Theta}_{zzz} \circ \boldsymbol{\Theta}_{z})^{\mathrm{T}(2,5)}, \boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{v} \gg$$
$$+\frac{1}{2!} \cdot \frac{1}{2!} \ll (\boldsymbol{\Theta}_{zz} \circ \boldsymbol{\Theta}_{zz})^{\mathrm{T}(2,4)}, \boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{v} \gg$$
$$+\mathcal{O}(\boldsymbol{v}^5) \in \mathbb{R}^{K \times K}. \tag{40}$$

Taking the mean we get the MSE,

$$E[\boldsymbol{\epsilon}(\boldsymbol{v}) \circ \boldsymbol{\epsilon}(\boldsymbol{v})] \approx \ll (\boldsymbol{\Theta}_{z} \circ \boldsymbol{\Theta}_{z})^{\mathrm{T}(2,3)}, \boldsymbol{\Lambda} \gg$$
$$+\frac{1}{6} \ll (\boldsymbol{\Theta}_{z} \circ \boldsymbol{\Theta}_{zzz})^{\mathrm{T}(2,3)} + (\boldsymbol{\Theta}_{zzz} \circ \boldsymbol{\Theta}_{z})^{\mathrm{T}(2,5)}, \boldsymbol{\mathcal{K}} \gg$$
$$+\frac{1}{4} \ll (\boldsymbol{\Theta}_{zz} \circ \boldsymbol{\Theta}_{zz})^{\mathrm{T}(2,4)}, \boldsymbol{\mathcal{K}} \gg + \boldsymbol{\mathcal{O}}(\boldsymbol{v}^5) \in \mathbb{R}^{K \times K}, \tag{41}$$

where $\boldsymbol{\mathcal{K}}$ is the fourth-order moments tensor of $\boldsymbol{v}$, which can be easily shown (e.g., by Isserlis' Theorem [20] regarding high-order moments of the multivariate Gaussian distribution) to be given by

$$\boldsymbol{\mathcal{K}} \triangleq E[\boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{v} \circ \boldsymbol{v}] = (\boldsymbol{\Lambda} \circ \boldsymbol{\Lambda})^{+\left[\begin{smallmatrix}(2,3)\\(2,4)\end{smallmatrix}\right]} \in \mathbb{R}^{K \times K \times K \times K} \tag{42}$$

(for a zero-mean Gaussian noise vector with covariance $\boldsymbol{\Lambda}$).

Although derived for Gaussian noise, this expression will hold for any noise distribution that has the same $\boldsymbol{\mathcal{K}}$ as its fourth-order moments tensor, namely has a vanishing fourth-order cumulants tensor.

Once again, the "standard", classical first-order analysis essentially ignores (neglects) all but the first term in (41), $\ll (\boldsymbol{\Theta}_{z} \circ \boldsymbol{\Theta}_{z})^{\mathrm{T}(2,3)}, \boldsymbol{\Lambda} \gg$, which is simply a different notation for

$$\boldsymbol{\Theta}_{z} \boldsymbol{\Lambda} \boldsymbol{\Theta}_{z}^{\mathrm{T}} = \boldsymbol{H} \boldsymbol{\Lambda} \boldsymbol{H}^{\mathrm{T}} = \boldsymbol{B} \overset{o}{\boldsymbol{G}}{}_{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Lambda}^{-1} \overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}} \boldsymbol{B} = \boldsymbol{B}, \tag{43}$$

the CRLB. Therefore, the MSE is generally different from the CRLB in the nonlinear model (1), and expression (41) quantifies the deviation (up to fourth-order) in the additional terms.

When the noise is sufficiently small, the high-order terms in both the bias (38) and MSE (41) expressions indeed become negligible, giving rise to the well-known asymptotic efficiency of the MLE. Our more accurate expressions enable to quantify the relative weight of the residual terms and to deduce the noise levels above which the (squared) bias is no longer negligible w.r.t. the MSE (so that the MLE becomes essentially biased), and the excess terms in (41) are no longer negligible w.r.t. its first term - the CRLB (causing a significant efficiency gap).

To summarize our computational recipe, we note that for any given measurement model $\boldsymbol{g}(\boldsymbol{\theta})$, parameters values $\boldsymbol{\theta}_0$ and noise covariance $\boldsymbol{\Lambda}$, the user needs to provide the first derivative matrix $\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}}$ and the second and third derivatives tensors $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}}$ and $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}$ (resp.), all taken at $\boldsymbol{\theta}_0$. Then, these derivatives are substituted in (25), (27), (29), (31), (32) and (33). These are substituted, in turn, into (34), (35) and (36), which, together with $\boldsymbol{\Lambda}$, are substituted into (38) to compute the bias, and into (41) to compute the MSE.

A word of caution is in order regarding the resulting approximated MSE matrix: While the first-order approximation, namely the CRLB matrix, is guaranteed to always be positive (semi-)definite, there is no such guarantee for the resulting higher-order approximation. Indeed, although under common sub-asymptotic conditions the approximated MSE matrix would usually be larger than the CRLB, in some cases it may be smaller than the CRLB, and in extreme cases it may even be sign-indefinite.

VI. Examples

In this section we demonstrate the results of our performance analysis with three different examples, illustrating three different nonlinearities in the measurement models. The first example addresses a classical problem of emitter location estimation based on measurements of the ratio between the received field strength in two perpendicular directions, also known as Adcock's method ([21], [22]). The second example is geared towards a particular mathematical property (see below), but can be related to the estimation of the volume of droplets of different types from measurements of the diameter of one of the droplets and the diameter of their merged droplet. The third example addresses the classical problem of DOA estimation using a non-uniform linear array (NULA, [23]) with a single snapshot, and demonstrates the use of our results with complex-valued measurements by concatenating the real and imaginary parts of the measurement vector into a double-sized real-valued vector.

A. Adcock Emitter Localization

For simplicity we assume a two-dimensional setting, where $M = 3$ sensors are positioned at coordinates $\boldsymbol{s}_1 = [0,\ 0]^{\mathrm{T}}$, $\boldsymbol{s}_2 = [20,\ 0]^{\mathrm{T}}$ and $\boldsymbol{s}_3 = [30,\ 10]^{\mathrm{T}}$ (units are, say, in [Km]). Each sensor is an Adcock array of four antennas, grouped in pairs. Suppose that one of the pairs lies in the North-South axis and the other lies in the East-West axis. The ratio of the North-South pair difference to the East-West pair difference provides the tangent of the Direction of Arrival (DOA) of the signal from the emitter (positioned at $\boldsymbol{\theta} \in \mathbb{R}^2$) to the sensor. The elements of the nonlinear measurement model function $\boldsymbol{g}(\boldsymbol{\theta}) \in \mathbb{R}^3$ are therefore given by[1]

$$g(m) = \frac{\theta(1) - s_m(1)}{\theta(2) - s_m(2)} \triangleq \frac{\Delta(m,1)}{\Delta(m,2)}, \quad m = 1,2,3, \tag{44}$$

where for convenience we defined the matrix $\boldsymbol{\Delta} \in \mathbb{R}^{3\times 2}$ with $\Delta(m,k) = \theta(k) - s_m(k)$, $k = 1,2$ and $m = 1,2,3$. It is then easily observed that the required derivatives up to third order

[1]With a slight abuse of notations, the interpretation of the parenthesized term in $g(\cdot)$ is determined by the context, to either be the m-th element of $\boldsymbol{g}$, as in $g(m)$ (as per our convention stated in Section II above), or its functional argument as in $\boldsymbol{g}(\boldsymbol{\theta})$.

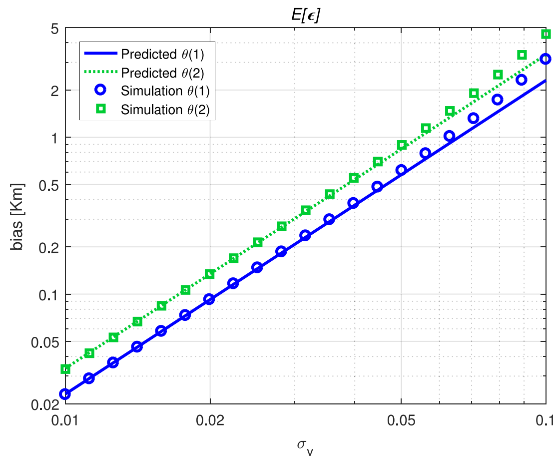


Fig. 1. Adcock emitter location bias for $\boldsymbol{s}_1 = [0\ 0]^{\mathrm{T}}$, $\boldsymbol{s}_2 = [20\ 0]^{\mathrm{T}}$, $\boldsymbol{s}_3 = [30\ 10]^{\mathrm{T}}$, and $\boldsymbol{\theta}_0 = [50\ 50]^{\mathrm{T}}$.

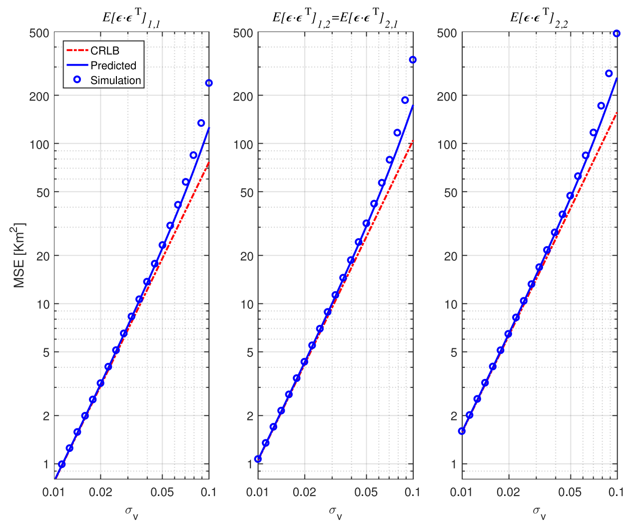


Fig. 2. Adcock emitter location MSE (same setup as in Fig. 1).

are given (for $m = 1, 2, 3$) by:

$$\frac{\partial g(m)}{\partial \theta(1)} = \frac{1}{\Delta(m,2)}, \qquad \frac{\partial g(m)}{\partial \theta(2)} = -\frac{\Delta(m,1)}{(\Delta(m,2))^2}, \tag{45}$$

$$\frac{\partial^2 g(m)}{\partial \theta(1)\partial \theta(1)} = 0 \qquad \frac{\partial^2 g(m)}{\partial \theta(1)\partial \theta(2)} = -\frac{1}{(\Delta(m,2))^2}$$
$$\frac{\partial^2 g(m)}{\partial \theta(2)\partial \theta(2)} = -\frac{2\Delta(m,1)}{(\Delta(m,2))^3} \tag{46}$$

and

$$\frac{\partial^3 g(m)}{\partial \theta(1)\partial \theta(1)\partial \theta(1)} = 0 \qquad \frac{\partial^3 g(m)}{\partial \theta(2)\partial \theta(2)\partial \theta(2)} = -\frac{6\Delta(m,1)}{(\Delta(m,2))^4}$$
$$\frac{\partial^3 g(m)}{\partial \theta(1)\partial \theta(1)\partial \theta(2)} = 0 \qquad \frac{\partial^3 g(m)}{\partial \theta(1)\partial \theta(2)\partial \theta(2)} = \frac{2}{(\Delta(m,2))^3}. \tag{47}$$

Using these expressions, and given any particular value of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for which the performance analysis is sought, the required derivative matrix $\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}} \in \mathbb{R}^{3\times 2}$ and tensors $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}} \in \mathbb{R}^{3\times 2\times 2}$ and $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}} \in \mathbb{R}^{3\times 2\times 2\times 2}$ are readily obtained, and can be substituted into the required expressions for obtaining the predicted bias vector and MSE matrix. We computed the results for $\boldsymbol{\theta}_0 = [50, 50]$ ([Km]) with $\boldsymbol{\Lambda} = \sigma_v^2 \boldsymbol{I} \in \mathbb{R}^{3\times 3}$ as the noise covariance, varying σ_v between 0.01 and 0.1. We compared our predicted performance to results that were obtained empirically in simulation, using 10^6 independent trials for each tested value of σ_v. In each trial, the noise vector $\boldsymbol{v} \in \mathbb{R}^3$ was drawn from a zero-mean Gaussian distribution with covariance $\boldsymbol{\Lambda}$ to generate the measurement vector $\boldsymbol{z} = \boldsymbol{g}(\boldsymbol{\theta}_0) + \boldsymbol{v}$, from which the MLE was obtained using gradient descent, initialized by a grid search.

Figure 1 illustrates the predicted bias in both parameters (x- and y-coordinates), showing very good agreement between the analytically predicted bias and its empirical counterpart. We remind here that with the "standard" small-errors analysis, the predicted bias would simply be zero.

Figure 2 compares the elements of our predicted 2×2 MSE matrix $E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}]$ both to the empirically obtained values and to the CRLB, given by the matrix $\boldsymbol{B}$ in (32):

$$\boldsymbol{B} = \left(\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}}{}^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}}\right)^{-1}$$
$$= \sigma_v^2 \left(\sum_{m=1}^{3} \begin{bmatrix} \frac{1}{(\Delta(m,2))^2} & \frac{-\Delta(m,1)}{(\Delta(m,2))^3} \\ \frac{-\Delta(m,1)}{(\Delta(m,2))^3} & \frac{(\Delta(m,1))^2}{(\Delta(m,2))^4} \end{bmatrix} \right)^{-1}. \tag{48}$$

The departure of the empirical values from the CRLB is clearly seen to follow the analytically predicted MSE as the noise variance is increased. Naturally, when the noise variance is further increased, the empirical errors become even larger, and reach a point where terms neglected in our high-order analysis are no longer negligible, causing departure from our predicted values. Nevertheless, our analysis not only provides an accurate prediction of the bias and MSE under non-asymptotic conditions up to a certain point, but also provides a close prediction of the "threshold" noise variance (roughly 0.05^2 in this case) above which the inefficiency of the MLE is significantly pronounced.

Additionally, in Figure 3 we show the "relative efficiency" measure, which is defined (individually for each parameter) as the ratio between the "excess MSE" and the CRLB, where the "excess MSE" is the difference between the actual MSE and the "efficient MSE" as predicted by the CRLB. Clearly, when the MLE is assumed to be efficient (based on a standard small-errors analysis), the implied relative efficiency is zero. This inefficiency measure of the MLE is also seen to be accurately predicted by our analytic results, roughly up to the above mentioned noise threshold.

In the next example we use a model for which the MLE is a cubic (third degree polynomial) function of the measurements (and therefore also of the noise), so that the good agreement

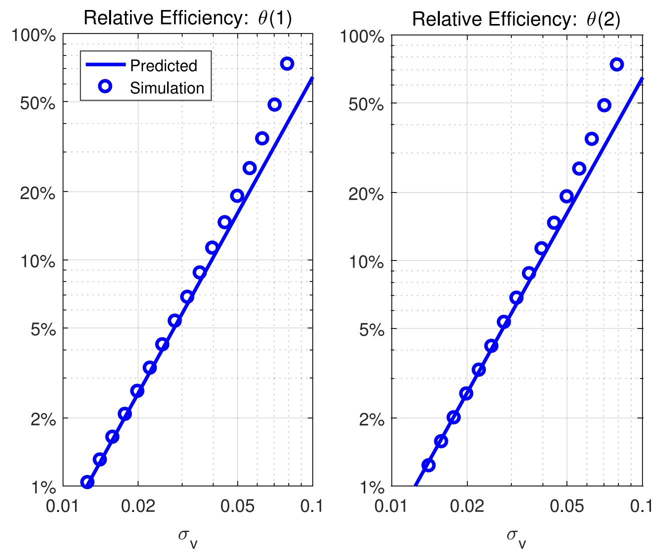


Fig. 3. Adcock emitter location relative efficiency (same setup as in Fig. 1).

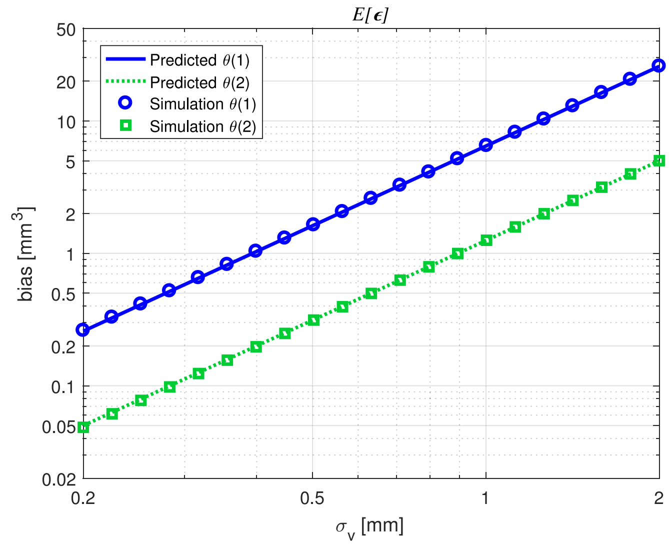


Fig. 4. Droplets volume estimation bias. The empirical results are based on 10^6 independent trials.

between our fourth-order analytical prediction and the empirical results is maintained for relatively higher noise variances.

B. Volume Estimation of Two Types of Droplets From Measurements of Merged Diameters

Our model equations in this example, with $M = 2$ measurements depending on $K = 2$ parameters, are given by

$$g(1) = (\theta(1))^{1/3} \quad , \quad g(2) = (\theta(1) + \theta(2))^{1/3}. \tag{49}$$

This model can occur, for example, in the following context. Suppose that two round (ball-shaped) droplets of two different liquid substances are submerged in a third (ambient) liquid substance. The first liquid has a distinct color, but the second liquid, as well as the ambient liquid, have a nearly transparent color. We are interested in the volumes (or masses) of the two droplets, but our instrumentation can only measure the diameter of an observed colored ball of droplet. Therefore, we first measure the diameter of the droplet of the first substance; then we somehow cause the two droplets to merge (into a single droplet which still maintains some color of the first) and measure the diameter of the merged ball (whose volume is the sum of the original volumes). Since the diameter is related to the volume through a cubic root relation, we obtain the measurement model (49). We note, however, that this description is merely a possible illustration of this mathematical model; In the MLE solution (and consequently, in our analysis) we do not constrain the estimated $\theta(1)$ and $\theta(2)$ to be positive (as droplets' volumes would be), namely, the maximization of the likelihood is assumed to be taken w.r.t. all $\boldsymbol{\theta} \in \mathbb{R}^2$.

The required derivatives are easily obtained as:

$$\begin{aligned} \boldsymbol{G}_{\boldsymbol{\theta}}(1,:) &= \tfrac{1}{3}(\theta(1))^{-\frac{2}{3}} \cdot \begin{bmatrix} 1 & 0 \end{bmatrix}, \\ \boldsymbol{G}_{\boldsymbol{\theta}}(2,:) &= \tfrac{1}{3}(\theta(1)+\theta(2))^{-\frac{2}{3}} \cdot \begin{bmatrix} 1 & 1 \end{bmatrix} \end{aligned} \tag{50}$$

$$\begin{aligned} \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}(1,:,:) &= -\tfrac{2}{9}(\theta(1))^{-\frac{5}{3}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \\ \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}(2,:,:) &= -\tfrac{2}{9}(\theta(1)+\theta(2))^{-\frac{5}{3}} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{aligned} \tag{51}$$

and

$$\begin{aligned} \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}}(1,:,:,:) &= \tfrac{10}{27}(\theta(1))^{-\frac{8}{3}} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \\ \boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}}(2,:,:,:) &= \tfrac{10}{27}(\theta(1)+\theta(2))^{-\frac{8}{3}} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \end{aligned} \tag{52}$$

where we have used standard Matlab array indexing notations in (50), (51) and (52).

For our simulation experiment we used $\boldsymbol{\theta}_0 = [10,\ 7]^{\mathrm{T}}$ (say, in [mm^3]), which we substituted into these derivative expressions to obtain $\overset{o}{\boldsymbol{G}}_{\boldsymbol{\theta}} \in \mathbb{R}^{2\times 2}$, $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}} \in \mathbb{R}^{2\times 2\times 2}$ and $\overset{o}{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}} \in \mathbb{R}^{2\times 2\times 2\times 2}$. Our noise covariance matrix was $\boldsymbol{\Lambda} = \sigma_v^2 \boldsymbol{I} \in \mathbb{R}^{2\times 2}$, varying σ_v between 0.2 [mm] and 2 [mm]. Once again, we compare our predicted performance to empirical results, obtained by averaging the MLE's bias and MSE over 10^6 independent trials. The bias is compared in Figure 4, whereas the MSE matrix is compared in Figure 5, where the CRLB, which in this case can be easily shown to be given by

$$\boldsymbol{B} = \sigma_v^2 \cdot 9 \cdot (\theta(1))^{4/3} \cdot \begin{bmatrix} 1 & -1 \\ -1 & 1 + \left(1 + \frac{\theta(2)}{\theta(1)}\right)^{4/3} \end{bmatrix} \tag{53}$$

is also shown for comparison. In Figure 6 we show the relative efficiency measure.

A good fit between the predicted performance and the empirical performance is observed, up to significant deviations from unbiasedness and from the CRLB, including an accurate indication of the "threshold" noise variance (around 0.4^2 [mm^2]) which defines the *non-asymptotic* regime.

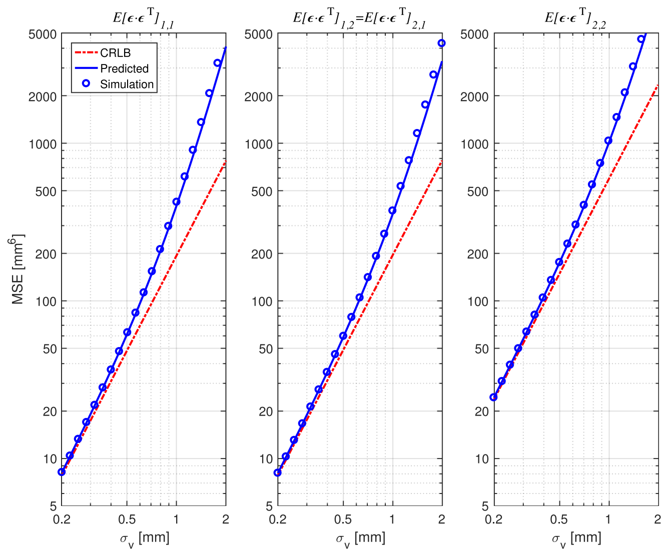


Fig. 5. Droplets volume estimation MSE (setup as in Fig. 4).

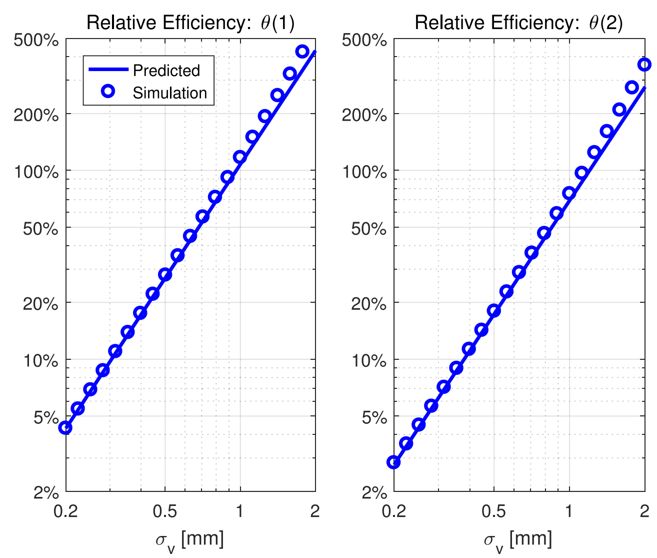


Fig. 6. Droplets volume estimation relative efficiency (setup as in Fig. 4).

C. Single-Snapshot DOA Estimation Using a NULA

In this example we consider a NULA consisting of 7 elements positioned (along the x-axis) at

$$\boldsymbol{x} = \lambda \cdot [0 \;\; 0.6 \;\; 0.9 \;\; 1.5 \;\; 1.8 \;\; 2.4 \;\; 2.7]^{\mathrm{T}}, \tag{54}$$

where λ denotes the wavelength. With a single snapshot of a wavefront with amplitude A and phase ϕ, impinging on the array from a DOA α (measured counter-clockwise relative to the y-axis), the complex-valued elements of the nonlinear measurements model function, denoted $\widetilde{\boldsymbol{g}}(\boldsymbol{\theta}) \in \mathbb{C}^7$ (with $\boldsymbol{\theta} \triangleq [\alpha \;\phi\; A]^{\mathrm{T}}$ being the unknown real-valued parameter vector) are given by

$$\widetilde{g}(m) = Ae^{j\phi} \exp\left(j2\pi \frac{x(m)}{\lambda} \sin(\alpha)\right). \tag{55}$$

Since our derivations assumed real-valued measurements, we define a real-valued model vector $\boldsymbol{g}(\boldsymbol{\theta}) \in \mathbb{R}^{14}$ consisting of the concatenation of the real and imaginary parts of $\widetilde{\boldsymbol{g}}(\boldsymbol{\theta})$. Noting that the parameter vector $\boldsymbol{\theta}$ is real-valued, the derivatives of the real and imaginary parts of $\widetilde{\boldsymbol{g}}(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$ are naturally given (resp.) by the real and imaginary parts of the respective derivatives of $\widetilde{\boldsymbol{g}}(\boldsymbol{\theta})$. Denoting the required (complex-valued) derivatives as $\widetilde{\boldsymbol{G}}_{\boldsymbol{\theta}}$, $\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}}$ and $\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}$ we have (for $m = 1, \ldots, 7$):

$$\widetilde{\boldsymbol{G}}_{\boldsymbol{\theta}}(m,:) = A\exp\left(j(d_m \sin(\alpha) + \phi)\right) \cdot [jd_m \cos(\alpha) \;\;\; j \;\;\; 1/A] \tag{56}$$

where we have defined $d_m \triangleq 2\pi x(m)/\lambda$ for convenience;

$$\begin{aligned}\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta}}(m,:,:) &= A\exp\left(j(d_m \sin(\alpha) + \phi)\right) \cdot \\ &\cdot \begin{bmatrix} -jd_m \sin(\alpha) - d_m^2 \cos(\alpha) & -d_m \cos(\alpha) & jd_m \cos(\alpha)/A \\ -d_m \cos(\alpha) & -1 & j \\ jd_m \cos(\alpha)/A & j & 0 \end{bmatrix};\end{aligned} \tag{57}$$

and

$$\begin{aligned}
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,1,1) &= jd_m(-\cos(\alpha) - 3jd_m\cos(\alpha)\sin(\alpha) \\
&\quad - d_m^2 \cos^3(\alpha))A\exp\left(j(d_m\sin(\alpha)+\phi)\right) \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,1,2) &= -d_m(-\sin(\alpha) + jd_m\cos^2(\alpha)) \\
&\quad \cdot A\exp\left(j(d_m\sin(\alpha)+\phi)\right) \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,1,3) &= jd_m(-\sin(\alpha) + jd_m\cos^2(\alpha)) \\
&\quad \cdot \exp\left(j(d_m\sin(\alpha)+\phi)\right) \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,2,2) &= -jd_m\cos(\alpha)A\exp\left(j(d_m\sin(\alpha)+\phi)\right) \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,2,3) &= -d_m\cos(\alpha)\exp\left(j(d_m\sin(\alpha)+\phi)\right) \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,3,3) &= 0 \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,2,2,2) &= -jA\exp\left(j(d_m\sin(\alpha)+\phi)\right) \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,2,2,3) &= -\exp\left(j(d_m\sin(\alpha)+\phi)\right) \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,2,3,3) &= 0 \\
\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,3,3,3) &= 0,
\end{aligned} \tag{58}$$

with symmetric completion for all other elements, i.e., $\widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,2,1,1) = \widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,2,1) = \widetilde{\boldsymbol{\mathcal{G}}}_{\boldsymbol{\theta\theta\theta}}(m,1,1,2)$, and so on. Substituting the real and imaginary parts into our computational recipe, we obtain the predicted performance.

Figure 7 presents predicted and empirical results for a range of DOAs between 60° and 80°, all at a fixed SNR: Assuming that the received complex-valued signal is contaminated by additive white complex-valued circular Gaussian noise with variance $2\sigma^2$, the real and imaginary parts of the measurements can each be assumed to be contaminated by additive white real-valued Gaussian noise with variance σ^2 (mutually independent between the two parts). In our setup we used $\sigma = 0.1$, with $A = 1$ (reflecting an SNR of 17 [dB]) and $\phi = 0$. For this experiment we only present results for the parameter of interest α - the predicted bias and root MSE (RMSE) (in [deg]), including the CRLB on the RMSE, vs. empirical results. A good fit is observed up to the higher DOA values, where due to the bounded DOA values (at 90°), the empirical MSE in the extreme cases is smaller than its predicted value.

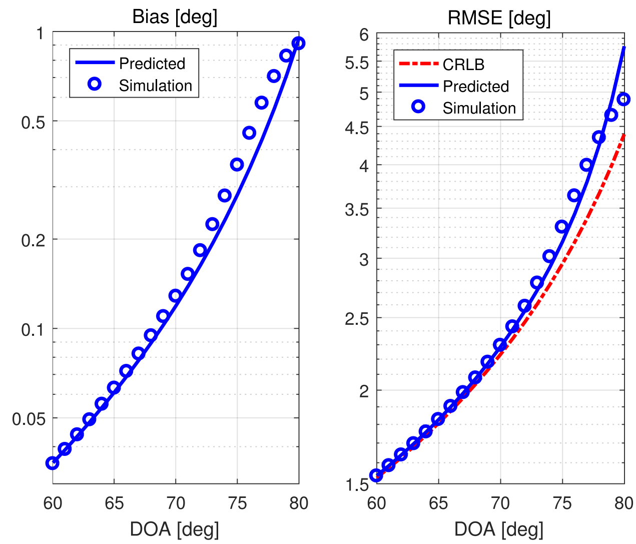


Fig. 7. DOA estimation bias and RMSE. The empirical results are based on 10^5 independent trials.

VII. Conclusion

We presented a high-order multivariate error performance analysis for ML estimation in nonlinear models with additive Gaussian noise, accounting for derivatives up to third order and noise moments up to fourth order. Our derivations lead to explicit expressions, which can be computed for any specific problem modeled by a continuous, nonlinear, three times differentiable function, by plugging in the first three derivatives tensors of the specified measurement model.

Since the MLE coincides with the OWLS estimate (when using the inverse noise covariance as the weight matrix), our results also enable the prediction of the realization-specific OWLS estimation error given the noise realization, regardless of the specific distribution of the noise. Our results can also be used to predict the performance (bias and MSE) for WLS estimation with other weight matrices, as well as with non-Gaussian noise, as long as high-order moments of the noise are known up to fourth-order.

In addition to the accurate prediction of the performance (bias and MSE) of the MLE for an extended range of noise levels, our results enable the important prediction, in the context of a given measurement model, of "threshold values" for the noise levels, at which the deviation from unbiasedness and from the CRLB becomes significant, and asymptotic conditions no longer prevail.

Extension of our analysis to higher orders (beyond third order derivatives and fourth order noise moments) is also possible using our tensor notations, which can keep the associated extensions tractable (albeit pretty long...).

Appendix A
Proof of the MLE's Inefficiency in Nonlinear Models

Theorem 1: Consider the model (1). The MLE is an efficient estimate of $\boldsymbol{\theta}$ if and only if $\boldsymbol{g}(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}$.

Proof: Recall ([11], [12]) that a necessary and sufficient condition for the existence of an efficient estimate is that the score function can be expressed in the form

$$\frac{\partial^{\mathrm{T}}}{\partial\boldsymbol{\theta}}\mathfrak{L}(\boldsymbol{z};\boldsymbol{\theta}) = \boldsymbol{C}(\boldsymbol{\theta})(\boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{\theta}), \tag{59}$$

where $\boldsymbol{C}(\boldsymbol{\theta}) \in \mathbb{R}^{K\times K}$ is a matrix depending only on $\boldsymbol{\theta} \in \mathbb{R}^K$, and $\boldsymbol{f}(\boldsymbol{z}) \in \mathbb{R}^K$ is some vector function depending only on $\boldsymbol{z} \in \mathbb{R}^M$ (this is essentially the necessary and sufficient condition for attaining equality in the Cauchy-Schwartz inequality in the classical derivation of the CRLB [12]). When this condition is fulfilled, we also get $\boldsymbol{C}(\boldsymbol{\theta}) = \boldsymbol{B}^{-1}$ (The Fisher Information Matrix, where $\boldsymbol{B}$ is defined in (32)), and then $\boldsymbol{f}(\boldsymbol{z})$ is the efficient estimate.

Now, consider the model (1). As shown in (14), the score for this model takes the form

$$\frac{\partial^{\mathrm{T}}}{\partial\boldsymbol{\theta}}\mathfrak{L}(\boldsymbol{z};\boldsymbol{\theta}) = \boldsymbol{G}_{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{z} - \boldsymbol{g}(\boldsymbol{\theta})) \triangleq \boldsymbol{C}'(\boldsymbol{\theta})(\boldsymbol{z} - \boldsymbol{g}(\boldsymbol{\theta})). \tag{60}$$

Obviously, when $\boldsymbol{g}(\cdot)$ is nonlinear, equation (60) cannot be written in the form of equation (59). To show this rigorously, assume that it can, namely that there exist some $\boldsymbol{C}(\boldsymbol{\theta})$ and $\boldsymbol{f}(\boldsymbol{z})$ such that $\boldsymbol{C}(\boldsymbol{\theta})(\boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{\theta}) = \boldsymbol{C}'(\boldsymbol{\theta})(\boldsymbol{z} - \boldsymbol{g}(\boldsymbol{\theta}))$. Since this equality has to hold for all $\boldsymbol{z}$ and $\boldsymbol{\theta}$, we necessarily have

$$\boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{z}) = \boldsymbol{C}'(\boldsymbol{\theta})\boldsymbol{z} \quad \text{and} \quad \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{\theta} = \boldsymbol{C}'(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{\theta}). \tag{61}$$

multiplying on the left with $\boldsymbol{C}^{-1}(\boldsymbol{\theta})$ (which must exist if an efficient estimate exists) and denoting $\boldsymbol{D}(\boldsymbol{\theta}) \triangleq \boldsymbol{C}^{-1}(\boldsymbol{\theta})\boldsymbol{C}'(\boldsymbol{\theta})$, we get

$$\boldsymbol{f}(\boldsymbol{z}) = \boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{z} \quad \text{and} \quad \boldsymbol{\theta} = \boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{\theta}). \tag{62}$$

Now, since $\boldsymbol{f}(\boldsymbol{z})$ is not allowed to depend on $\boldsymbol{\theta}$, we must conclude that $\boldsymbol{D}(\boldsymbol{\theta})$ cannot depend on $\boldsymbol{\theta}$ either, and must therefore be a fixed matrix $\boldsymbol{D}$, with which $\boldsymbol{\theta} = \boldsymbol{D}\boldsymbol{g}(\boldsymbol{\theta})$ cannot be satisfied if $\boldsymbol{g}(\boldsymbol{\theta})$ is nonlinear.

Therefore, we conclude that for any nonlinear model with additive Gaussian noise (not necessarily spatially and/or temporally white), the MLE is not efficient. ■

Appendix B
Matlab Code for Some Basic Tensor Operations

We provide short Matlab functions for implementing the elementary tensor operations defined in Section II. Our code assumes that the dimensions of all of the functions' arguments are compatible for the required operation, and therefore no explicit validation of the compatibility is included. Naturally, incompatibility would eventually generate a Matlab error (in most cases), but would not point directly to the incompatibility issue. It is the user's responsibility to make sure that the functions are called with compatible dimensions, or to add some explicit validation to the code.

The complete code for computing predicted results (accommodating user-supplied functions for the desired nonlinearity $\boldsymbol{g}(\boldsymbol{\theta})$ and its required derivatives $\boldsymbol{G}_{\boldsymbol{\theta}}$, $\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta}}$ and $\boldsymbol{\mathcal{G}}_{\boldsymbol{\theta\theta\theta}}$) is available as "supplemental material" for this paper.

- Outer Products:

$$\mathcal{R} = \mathcal{P} \circ \mathcal{Q}$$

```
function R=TT_outprod(P,Q)
szP=size(P);
szQ=size(Q);
qp=kron(Q(:),P(:));
R=reshape(qp,[szP szQ]);
end
```

- "n-th mode" Product with a Vector:

$$\mathcal{Q} = \mathcal{P} \bar{\times}_n \boldsymbol{a}$$

```
function Q=Ta_prod(P,n,a)
P=shiftdim(P,n-1);
szP=size(P);
NP=ndims(P);
szP2=szP(2:NP);
Pr=reshape(P,szP(1),prod(szP2));
Qr=a(:)'*Pr;
Q=reshape(Qr,szP2);
Q=shiftdim(Q,NP-n);
end
```

- "n-th mode" Product with a Matrix:

$$\mathcal{Q} = \mathcal{P} \times_n \boldsymbol{A}$$

```
function Q=TA_prod(P,n,A)
P=shiftdim(P,n-1);
szP=size(P);
NP=ndims(P);
szP2=szP(2:NP);
Pr=reshape(P,szP(1),prod(szP2));
Qr=A*Pr;
Q=reshape(Qr,[size(A,1) szP2]);
Q=shiftdim(Q,NP+1-n);
end
```

- "n-th mode" Product with a Tensor:

$$\mathcal{Q} = \mathcal{P} \times_n^+ \mathcal{A}$$

```
function Q=TT_prod(P,n,A)
szP=size(P);
szA=size(A);
NA=ndims(A);
szQ1=szP;
szQ1(n)=szA(1);
NQ1=prod(szQ1);
szQ=[szQ1 szA(3:NA)];
Q=zeros(szQ);
for mA=1:prod(szA(3:NA))
    Q((mA-1)*NQ1+(1:NQ1))=...
    TA_prod(P,n,A(:,:,mA));
end
end
```

- An Inner Product:

$$\mathcal{R} = \ll \mathcal{P}, \mathcal{Q} \gg$$

(note: a "simple" inner product $r = < \mathcal{P}, \mathcal{Q} >$ can be obtained implicitly as a particular case when the sizes of the two tensors are equal).

```
function R=TT_inprod(P,Q)
szP=size(P);
NP=ndims(P);
NQ=ndims(Q);
L=NP-NQ;
np1=prod(szP(1:L));
np2=prod(szP(L+1:NP));
Pr=reshape(P,np1,np2);
Rr=Pr*Q(:);
if L>1
    R=reshape(Rr,szP(1:L));
else
    R=Rr;
end
end
```

The transposing, and more generally, the permutation operation, can be implemented using the elementary Matlab function `permute`, so that, for example, $\mathcal{Q} = \mathcal{P}^{[3,1,2,4]}$ can be implemented as `Q = permute(P,[3,1,2,4])`. Likewise, Matlab's `shiftdim` can be used for $\mathcal{Q} = \mathcal{P} \uparrow$ as `Q = shiftdim(P,1)`.

References

[1] T. Abatzoglou, "A fast maximum likelihood algorithm for frequency estimation of a sinusoid based on Newton's method," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 1, pp. 77–89, Feb. 1985.

[2] A. K. Shaw, "Maximum likelihood estimation of multiple frequencies with constraints to guarantee unit circle roots," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 796–799, Mar. 1995.

[3] P. Stoica, A. Jakobsson, and J. Li, "Cisoid parameter estimation in the colored noise case: Asymptotic Cramer–Rao bound, maximum likelihood, and nonlinear least-squares," *IEEE Trans. Signal Process.*, vol. 45, no. 8, pp. 2048–2059, Aug. 1997.

[4] M. Djeddou, A. Belouchrani, and S. Aouada, "Maximum likelihood angle-frequency estimation in partially known correlated noise for low-elevation targets," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3057–3064, Aug. 2005.

[5] H. Minn and P. Tarasak, "Improved maximum likelihood frequency offset estimation based on likelihood metric design," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2076–2086, Jun. 2006.

[6] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.

[7] P. Stoica and A. Nehorai, "Music, maximum likelihood, and Cramer–Rao bound," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

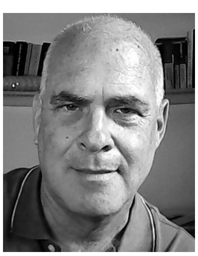
[8] P. Stoica and A. Nehorai, "Music, maximum likelihood, and Cramer–Rao bound: further results and comparisons," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.

[9] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 996–1004, Apr. 1997.

[10] J. H. Won, T. Pany, and B. Eissfeller, "Noniterative filter-based maximum likelihood estimators for GNSS signal tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1100–1114, Apr. 2012.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[12] C. R. Rao, *Linear Statistical Inference and its Applications*, 2nd ed. Hoboken, NJ, USA: Wiley, 2002.
[13] Q. Lu, Y. Bar-Shalom , P. Willett, and S. Zhou, "Nonlinear observation models with additive Gaussian noises and efficient MLEs," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 545–549, May 2017.
[14] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 4, pp. 532–544, Apr. 1988.
[15] N. D. Sidoropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.
[16] P. Comon, "Tensors : A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, May 2014.
[17] N. Vervliet, O. Debals, L. Sorber, and L. D. Lathauwer, "Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 71–79, Sep. 2014.
[18] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
[19] S. G. Kranz and H. R. Parks, *The Implicit Function Theorem: History, Theory and Applications*. Cambridge, MA, USA: Birkhäuser, 2013.
[20] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, vol. 12, no. 1–2, pp. 134–139, 1918.
[21] F. Adcock, "Improvements in means for determining the direction of a distant source of electro-magnetic radiation," Patent No. GB 130490, 1919.
[22] P. G. Redgment, W. Struszynski, and G. J. Phillips, "An analysis of the performance of multi-aerial Adcock direction-finding systems," *J. IEE - Part IIIA: Radiocommun.*, vol. 94, no. 15, pp. 751–761, 1947.
[23] A. B. Gershman and J. F. Bohme, "A note on most favorable array geometries for doa estimation and array interpolation," *IEEE Signal Process. Lett.*, vol. 4, no. 8, pp. 232–235, Aug. 1997.



Arie Yeredor (M'99–SM'02) received the B.Sc. (summa cum laude) and Ph.D. degrees in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1984 and 1997, respectively. He is currently an Associate Professor with the School of Electrical Engineering, Department of Electrical Engineering-Systems, TAU, Tel-Aviv, Israel. From 2015 to 2016, he was on sabbatical leave from TAU, as a Visiting Professor at the Georgia Institute of Technology, Atlanta, GA, USA. His research and teaching interests include statistical and digital signal processing and estimation theory. He has also held a consulting position in these research areas with NICE Systems, Inc., Ra'anana, Israel, from 1990 to 2015. He serves as the Chair of the Yitzhak and Chaya Weinstein Institute for Signal Processing Research at TAU. He was the recipient of the yearly Best Lecturer Award seven times at TAU. He was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-PART II, AND IEEE TRANSACTIONS ON SIGNAL PROCESSING, where he now serves as a Senior Area Editor. He was also a Guest Editor for the IEEE SIGNAL PROCESSING MAGAZINE'S special issue on Source Separation and Applications. He is a Member of the IEEE Signal Processing Society's Signal Processing Theory and Methods Technical Committee, and the Chair of the Signal Processing chapter of IEEE Israel Section.

Amir Weiss (S'17) received the B.Sc. (cum laude) and M.Sc. degrees in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 2013 and 2015, respectively, where he is currently working toward the Ph.D. degree with the School of Electrical Engineering, Department of Electrical Engineering-Systems. His research and teaching areas include statistical and digital signal processing and estimation theory. He has also been holding a researcher position in these research areas with Elbit Systems, EW and SIGINT Elisra Ltd., Holon, Israel, since 2013. He received the scholarship for excellent M.Sc. students from the Faculty of Engineering, TAU, in 2015. He was the recipient of the Nadav Levanon studies prize for graduate students and the scientific publication prize (twice), all from the The Yitzhak and Chaya Weinstein Research Institute for Signal Processing in 2016, 2017, and 2018, respectively.

Anthony J. Weiss (S'84–M'85–SM'86–F'97) received the B.Sc. degree from the Technion-Israel Institute of Technology, in 1973, and the M.Sc. and Ph.D. degrees from Tel Aviv University, Tel Aviv, Israel, in 1982 and 1985, all in electrical engineering. From 1973 to 1983, he was involved in research and development of numerous projects in the fields of communications, command and control, and emitter localization. In 1985, he joined the Department of Electrical Engineering-Systems, Tel Aviv University. From 1996 to 1999, he served as the Department Chairman and as the Chairman of IEEE Israel Section. From 2006 to 2011, he was the Head of the EE School, Tel Aviv University. His research interests include detection and estimation theory, signal processing, sensor array processing, and wireless networks. He currently serves as Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He also held leading scientific positions with Signal Processing Technology Ltd., Wireless on Line Ltd., and SigmaOne Ltd. He has authored or coauthored more than 200 papers in professional magazines and conferences and holds 11 U.S. patents. He was the recipient of the IEEE 1983 Acoustics, Speech, and Signal Processing Society's Senior Award and the IEEE third millennium medal.