On the efficiency of blind and non-blind estimation for coupled LL1 tensor models using the randomly-constrained Cramér-Rao bound

Clémence Prévost, Konstantin Usevich, Eric Chaumette, David Brie, Pierre Comon

On the efficiency of blind and non-blind estimation for coupled LL1 tensor models using the randomly-constrained Cramér-Rao bound

C. Prévost, *Member, IEEE*, K. Usevich, *Member, IEEE*, E. Chaumette, D. Brie, *Member, IEEE*, P. Comon, *Fellow IEEE*

***Abstract*—In this paper, we study the performance of two algorithms for tensor reconstruction in the presence of random uncertainties: the first one is the non-blind state-of-the-art, and the second one, new, is blind and is designed to deal with such uncertainties. Using coupled tensor LL1 models, we show that the proposed algorithm performs better than the state-of-the-art, which naturally raises the question of its efficiency. In that perspective, the standard approach is to resort to the usual constrained Cramér-Rao bound (CCRB), which appears to be only partially informative when addressing the asymptotic achievable performance of the considered model. Indeed, the usual CCRB is not able to account for randomness in the set of constraints. To fill this gap, we also introduce a new randomly-constrained Cramér-Rao bound and we illustrate its relevance to analyze the relative efficiency of the proposed algorithm. As a by-product, we provide closed-form expressions for the Fisher information matrices based on the LL1 model.**

***Index Terms*—Cramér-Rao bounds, random equality constraints, tensor models, low-rank approximations.**

I. Introduction

A. Background

In various engineering fields such as remote sensing or biomedical imaging, the observations often possess more than two dimensions. For instance, hyperspectral images or MRI acquisitions through time can be seen as data cubes. These high-dimensional arrays are referred to as tensors.

In this paper, we consider a specific class of tensor reconstruction problems which aim at recovering a high-resolution tensor from two observations with some lower resolutions, under the assumption that full resolution is always available in at least one dimension of each observation. Examples of such reconstruction problems can be found in, e.g., hyperspectral super-resolution [30], biomedical imaging [54], chemistry [63], spectrum cartography [55] or learning over graphs [64].

This work has been partially supported by the DGA/MRIS (2018.60.0072.00.470.75.01) and by the ANR (Agence Nationale de Recherche) under grant LeaFleT (ANR-19-CE23-0021).

This work was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Barcelona, Spain, May 2020.

C. Prévost is with University of Lille, Centrale Lille, UMR CNRS 9189 – CRIStAL, Lille, France. e-mail: firstname.lastname@univ-lille.fr.

K. Usevich and D. Brie are with Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, Vandoeuvre-lès-Nancy, France. e-mail: firstname.lastname@univ-lorraine.fr.

E. Chaumette is with ISAE-SUPAERO, Université de Toulouse,Toulouse, France. e-mail: firstname.lastname@isae-supaero.fr

P. Comon is with CNRS, GIPSA-Lab, Univ. Grenoble Alpes, Grenoble, France. e-mail: firstname.lastname@gipsa-lab.grenoble-inp.fr.

Due to its interesting uniqueness properties and possible interpretability [48], block-term modeling with ranks $(L, L, 1)$ was recently considered for spectrum cartography [55] and a number of remote sensing [40], [41], [56], [65] applications. Several algorithms based on alternating least squares (ALS) were proposed [40], [41], [56] and showed competitive performance for solving the reconstruction problems at hand.

However, some of the proposed approaches (see for instance [30], [55]) for solving the reconstruction problem share a common limitation: they assume that the observations are acquired in the same conditions. In practice, acquisition time difference, variations in the acquisition conditions [34], [35], miscalibration or aging of the sensors [43]–[46] can lead to a general variability phenomenon that encompasses unknown uncertainties in the observation model. For instance, in remote sensing, these uncertainties depict variations in the seasonal, atmospheric or illumination conditions [50]–[52]. Although of great practical interest, the variability phenomenon was only considered recently for the addressed class of problems.

Theoretical performance analysis of such tensor models has gained some interest in the signal processing community. In [26]–[28], uncoupled models admitting canonical polyadic (CP) decompositions were considered. In [31], it was proposed to explore the performance of partially-coupled CP models with (possibly) non-linear couplings. The work of [31] has been extended in other works of the authors [32], [33] for coupled CP models that characterize the reconstruction problem at hand, with specific sets of constraints describing the hyperspectral super-resolution problem. However, to the best of our knowledge, performance analysis for tensor models admitting $(L, L, 1)$ block-term decompositions (LL1-BTD) has not been addressed at the writing time of this manuscript.

Let us recall that there are two main categories of parameters. In the first case, the parameters being estimated are considered to be deterministic, whereas the second category considers the parameters as random variables with an a priori probability. This paper addresses the first category, *i.e.*, deterministic parameters. In this context, a popular way to assess the performance of coupled tensor models is to consider the Cramér-Rao bounds (CRB) (see [7, §8.4] and [8, Part III]). Historically, the deterministic CRB was introduced to investigate fundamental limits of deterministic parameters estimation or to assess the relative performance of a specific estimator (efficiency) [3]–[6]. Provided that one keeps in mind the CRB limitations [9]–[13], that is, to become overly opti-

mistic when the observation conditions degrade (low signal-to-noise ration and/or low number of snapshots), the CRB is still of great interest for system analysis and design in the asymptotic region.

Moreover, in some estimation problems, the definition, in part or totally, of the parameter space results from deterministic equality constraints, as exemplified in [14]. Hence, numerous works [15]–[20] have been devoted to extend the results introduced in [14]: 1) by providing a general reparameterization inequality and the equivalence between parameterization change and equality constraints; 2) by studying the CRB modified by constraints either required by the model or required to solve identifiability issues; 3) by investigating the use of parameters constraints from a different perspective: the value of side (*a priori*) information on estimation performance. All these works have shown the versatility of constrained Cramér-Rao bound (CCRB) for estimation performance analysis and design of a system of measurement, as highlighted in [21].

In the presence of uncertainties however, the reconstruction problem at hand fall under the scope of estimation problems [1], [2], [22]–[24] for which the probability law that governs the effect of a deterministic parameter vector value on the observation results from a two-step probabilistic mechanism involving an additional random vector. In this setting, some equality constraints on the unknown deterministic parameter vector may depend on this additional random vector yielding random equality constraints, a case which can not be tackled with the standard form of the CCRB. In the preliminary version of this work [42], the randomly constrained CRB (RCCRB) was hence introduced to take into account such random equality constraints.

In this paper, we model the reconstruction problem as a tensor LL1-BTD model accounting for random uncertainties. We study the performance of two algorithms. The first one is state-of-the-art and it is inspired by the work of [55], [57]. It is a fully-coupled (non-blind) ALS algorithm that does not account for random uncertainties. The second one was recently proposed in a technical report by the authors [56], and denoted BTD-Var. Although this new algorithm is partially-coupled (blind) only, it is designed to be able to account for the variability phenomenon.

In that perspective, the standard approach to analyse the relative efficiency of an algorithm is to resort to the usual CCRB mentioned above, which is shown to be only partially informative. Indeed, the usual CCRB is not able to account for randomness in the set of constraints. To fill this gap, we also consider the RCCRB and we illustrate its relevance for tensor reconstruction in the presence of random uncertainties. As a by-product, we provide new closed-form expressions for the Fisher information matrices (FIM) based on the LL1 model.

This paper is organized as follows. The remaining of Section I introduces the main notations and important definitions regarding the tensor framework. In Section II, we recall the theoretical results introduced in [42] regarding the RCCRB. Section III addresses standard CCRBs for coupled LL1 models and provide the closed-form expressions for the FIM. In Section IV, we conduct performance analysis of the state-of-the-art algorithm in the absence of uncertainties. Section V introduces the main model of interest, in which the constraints on the parameter involve a random variability phenomenon. The limitations of the standard tool, together with the relevance of the RCCRB, are illustrated and lead to the introduction of the new partially-coupled algorithm accounting for uncertainties. Finally, the relative efficiency of this new estimator is assessed in Section VI by using the RCCRB.

B. Definitions and notations

In this paper, we mainly follow the notations of [36], [37]. We use the following fonts: lower (a) or uppercase (A) plain font for scalars, boldface lowercase ($\mathbf{a}$) for vectors, boldface uppercase ($\mathbf{A}$) for matrices and calligraphic ($\boldsymbol{\mathcal{A}}$) for tensors. The elements of vectors, matrices and tensors are denoted as a_i, $A_{i,j}$ and $\mathcal{A}_{i_1,\ldots,i_N}$, respectively. For a matrix $\mathbf{A}$, its transpose is denoted by $\mathbf{A}^T$. We use the notation $\mathbf{I}_N$ for the $N\times N$ identity matrix and $\mathbf{0}_{L\times K}$ for the $L\times K$ matrix of zeros. The notation $\mathbf{1}_L$ denotes an all-ones column vector of size L. The symbols $\boxtimes$, $\odot$ and $\boxdot$ denote the Kronecker, Khatri-Rao and Hadamard (element-wise) products, respectively, and $o\left(.\right)$ stands for the Landau notation. We use $\mathrm{vec}\{\cdot\}$ for the standard column-major vectorization of a matrix or a tensor. For two matrices $\mathbf{A}$ and $\mathbf{B}$, the operator $\mathrm{Diag}\{\mathbf{A},\mathbf{B}\}$ produces a block-diagonal matrix whose diagonal blocks are $\mathbf{A}$ and $\mathbf{B}$, and for a vector $\mathbf{a}$, the operator $\mathrm{diag}\{\mathbf{a}\}$ produces a diagonal matrix whose diagonal entries are the elements of $\mathbf{a}$, and the other entries are zero. For matrices $\mathbf{A}_1,\ldots,\mathbf{A}_R$ having the name number of rows, the notation $[\mathbf{A}_1,\ldots,\mathbf{A}_R]$ denotes horizontal concatenation of these matrices.

Each dimension of a tensor is called a mode. A mode-p fiber of tensor $\boldsymbol{\mathcal{X}}$ is a vector of $\boldsymbol{\mathcal{X}}$ obtained by fixing all but the p-th dimension. In this paper, we restrict to the scope of three-dimensional tensors.

Definition 1: **Outer product –** The outer product between three vectors $\mathbf{a}\in\mathbb{R}^I$, $\mathbf{b}\in\mathbb{R}^J$, $\mathbf{c}\in\mathbb{R}^K$ is an order-3 tensor of rank 1 defined as $\boldsymbol{\mathcal{X}}=\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}\in\mathbb{R}^{I\times J\times K}$. Each element of $\boldsymbol{\mathcal{X}}$ is accessed as $\mathcal{X}_{i,j,k}=a_ib_jc_k$.

Definition 2: **Tensor unfoldings –** The mode-p unfolding of a tensor $\boldsymbol{\mathcal{X}}$, denoted by $\mathbf{X}^{(p)}$, is the matrix whose rows are the p-mode fibers of $\boldsymbol{\mathcal{X}}$, ordered according to the vectorization order. For a third-order tensor $\boldsymbol{\mathcal{X}}\in\mathbb{R}^{I\times J\times K}$, we have $\mathbf{X}^{(1)}\in\mathbb{R}^{JK\times I}$, $\mathbf{X}^{(2)}\in\mathbb{R}^{IK\times J}$ and $\mathbf{X}^{(3)}\in\mathbb{R}^{IJ\times K}$.

Definition 3: **Mode product –** The mode-p product between a tensor $\boldsymbol{\mathcal{X}}$ and a matrix $\mathbf{M}$ is denoted by $\boldsymbol{\mathcal{X}}\bullet_p\mathbf{M}$ and is evaluated such that each mode-p fiber of $\boldsymbol{\mathcal{X}}$ is multiplied by $\mathbf{M}$. For instance, the elements of the mode-1 product between $\boldsymbol{\mathcal{X}}\in\mathbb{R}^{I\times J\times K}$ and $\mathbf{M}\in\mathbb{R}^{L\times I}$ are accessed as $(\boldsymbol{\mathcal{X}}\bullet_1\mathbf{M})_{\ell,j,k}=\sum_i\mathcal{X}_{i,j,k}\mathbf{M}_{\ell,i}$, $\ell\in\{1,\ldots,L\}$. Moreover, it holds that $\boldsymbol{\mathcal{Y}}=\boldsymbol{\mathcal{X}}\bullet_p\mathbf{M}\Leftrightarrow\mathbf{Y}^{(p)}=\mathbf{X}^{(p)}\mathbf{M}^T$.

C. Block-term decomposition with ranks $(L,L,1)$

In this subsection, we introduce the tensor block-term decomposition with ranks $(L,L,1)$ that we will use to build our model. We also recall sufficient uniqueness conditions and useful properties.

Definition 4: **Block-term decomposition with ranks** $(L, L, 1)$ **–** A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ generally admits a block-term decomposition with ranks $(L, L, 1)$ (LL1-BTD) as

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \left(\mathbf{A}_r \mathbf{B}_r^T\right) \otimes \mathbf{c}_r, \tag{1}$$

where $\mathbf{A}_r \in \mathbb{R}^{I \times L}$, $\mathbf{B}_r \in \mathbb{R}^{J \times L}$, and $\mathbf{c}_r \in \mathbb{R}^K$, for $r \in \{1, \ldots, R\}$. Moreover, we denote the LL1 factors $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_R] \in \mathbb{R}^{I \times LR}$, $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_R] \in \mathbb{R}^{J \times LR}$ and $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_R] \in \mathbb{R}^{K \times R}$.

Definition 5: **Partition-wise Khatri-Rao product –** The partition-wise Khatri-Rao product between two partitioned matrices $\mathbf{A}$ and $\mathbf{C}$ defined as above can be expressed as

$$\mathbf{C} \odot_p \mathbf{A} = [\mathbf{c}_1 \boxtimes \mathbf{A}_1, \ldots, \mathbf{c}_R \boxtimes \mathbf{A}_R] \in \mathbb{R}^{IK \times LR}.$$

Property 1: **Tensor unfoldings and LL1-BTD –** Using the above notations, the unfoldings of a tensor $\boldsymbol{\mathcal{X}}$ admitting an LL1-BTD as above can be expressed as

$$\begin{aligned}
\mathbf{X}^{(1)} &= (\mathbf{C} \odot_p \mathbf{B})\, \mathbf{A}^T, \\
\mathbf{X}^{(2)} &= (\mathbf{C} \odot_p \mathbf{A})\, \mathbf{B}^T, \\
\mathbf{X}^{(3)} &= [(\mathbf{A}_1 \odot \mathbf{B}_1)\, \mathbf{1}_L, \ldots, (\mathbf{A}_R \odot \mathbf{B}_R)\, \mathbf{1}_L]\, \mathbf{C}^T.
\end{aligned}$$

Theorem 1: **Generic uniqueness [48, Theorem 4.7] –** Let $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ denote an LL1-BTD of a tensor $\boldsymbol{\mathcal{X}}$ as in (1). Assume that $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ are drawn from some joint absolutely continuous distributions. If $IJ \geq L^2 R$ and

$$\min\left(\lfloor \frac{I}{L} \rfloor, R\right) + \min\left(\lfloor \frac{J}{L} \rfloor, R\right) + \min(K, R) \geq 2R + 2,$$

then $\mathbf{A}_r \mathbf{B}_r^T$ and $\mathbf{c}_r$ are generically unique almost surely for $r \in \{1, \ldots, R\}$.

Under the conditions specified in Theorem 1, the LL1 factors are unique up to some trivial ambiguities [48]. Indeed, the factors in a same rank-$(L, L, 1)$ term can be arbitrarily scaled, provided that their product remains the same. Moreover, the $\mathbf{A}_r$ terms can be postmultiplied by any nonsingular matrix $\mathbf{D}_r \in \mathbb{R}^{L \times L}$, provided that the $\mathbf{B}_r$ terms are premultiplied by $\mathbf{D}_r^{-1}$ as

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \left((\mathbf{A}_r \mathbf{D}_r)\left(\mathbf{D}_r^{-1} \mathbf{B}_r\right)^T\right) \otimes \mathbf{c}_r. \tag{2}$$

When deriving Cramér-Rao bounds, a proper factor normalization[1] is required to ensure that the model is statistically identifiable (see [20] for more details). As a result, we propose a way to correct the aforementioned ambiguities by setting the first entry of the $\mathbf{c}_r$ factors to ones. Moreover, we set the first $(L \times L)$ blocks of the $\mathbf{A}_r$ factors to the identity matrix $\mathbf{I}_L$, which corresponds to setting $\mathbf{D}_r = (\mathbf{A}_r)^{-1}_{1:L,1:L}$ in (2).

[1]Permutation ambiguities in the LL1 model can be neglected in a parameter estimation framework, since the factors can be permuted after estimation.

II. CRBs with Random Equality Constraints

A. Background on standard CRBs

As introduced in [1, p53], a model of the general deterministic estimation problem has the following components: 1) a parameter space $\Theta_d \subset \mathbb{R}^P$, 2) an observation space $\mathcal{Y} \subset \mathbb{R}^M$, 3) a probabilistic mapping from parameter vector space Θ_d to observation space $\mathcal{Y}$, that is the probability law $p(\mathbf{y}; \boldsymbol{\theta})$ that governs the effect of a parameter vector value $\boldsymbol{\theta} \in \Theta_d$ on the observation $\mathbf{y} \in \mathcal{Y}$ and, 4) an estimation rule, that is the mapping of the observation space $\mathcal{Y}$ into vector parameter estimates $\widehat{\boldsymbol{\theta}} \triangleq \widehat{\boldsymbol{\theta}}(\mathbf{y})$. If a closed-form expression of $p(\mathbf{y}; \boldsymbol{\theta})$ is available, the estimation problem at hand is called a "standard" deterministic estimation problem [2]. For such problem, the mean-squared error (MSE) matrix of $\widehat{\boldsymbol{\theta}}$ is a Gram matrix [20] defined on the vector space of square integrable functions and, therefore, all known standard lower bounds (LBs) on the MSE can be formulated as the solution of a norm minimization problem under linear constraints (LCs) [10], [12]. This formulation of LBs does not only provides a straightforward understanding of the hypotheses associated with the different LBs [10], [12], but it also allows to obtain a unique formulation of each LB in terms of a unique set of linear constraints. When the lower bound is the CRB, the set of linear constraints reduces to a set of derivative constraints [20]. Indeed, the CRB is the lowest bound on the MSE of unbiased estimators, since it is derived from the *weakest* formulation of unbiasedness, *i.e.* local unbiasedness,

$$E_{\mathbf{y};\boldsymbol{\theta}+d\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta} + d\boldsymbol{\theta} + o\left(\|d\boldsymbol{\theta}\|\right), \tag{3a}$$

which means that, up to the first order and in the neighborhood of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$ remains an unbiased estimator of $\boldsymbol{\theta}$ independently of a small variation of the parameters. Interestingly, (3a) can be rewritten in terms of Taylor expansion of each side, and the uniqueness of Taylor expansion imposes that the LCs

$$E_{\mathbf{y};\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \mathbf{0},\ E_{\mathbf{y};\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right] = \mathbf{I}, \tag{3b}$$

must be statisfied by any locally unbiased estimator. Then the CRB is easily obtained by using the following lemma on the minimization of a Gram matrix (with respect to the Löwner ordering [25, §7.7]) under LCs. Let $\mathbb{U}$ be a Hilbertian vector space on the field of real numbers $\mathbb{R}$ which has a scalar product $\langle \cdot \,|\, \cdot \rangle$. Let $\mathcal{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_K)$ be a family of K linearly independent vectors and $\mathcal{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_P)$ a family of P vectors. Then

$$\mathbf{V}^T \mathbf{G}(\mathcal{C})^{-1} \mathbf{V} = \min_{\mathcal{U}} \{\mathbf{G}(\mathcal{U})\} \text{ s.t. } \langle \mathbf{u}_p \,|\, \mathbf{c}_k \rangle = \mathbf{V}_{k,p}, \tag{4}$$

where $\mathbf{G}(\mathcal{W})$ denotes the Gram matrix associated with the family of N vectors $\mathcal{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_N)$ defined as $G_{n,n'}(\mathcal{W}) = \langle \mathbf{w}_{n'} \,|\, \mathbf{w}_n \rangle$, $1 \leq n, n' \leq N$. Indeed by defining $\mathcal{U} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ and $\mathcal{C} = (1, \frac{\partial \ln p(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T})$, and by considering the scalar product $\langle f(\mathbf{y}) \,|\, g(\mathbf{y}) \rangle = E_{\mathbf{y};\boldsymbol{\theta}}\left[f(\mathbf{y})\, g(\mathbf{y})\right]$, (4) can be applied with $\mathbf{V} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix}$ (3b) and leads to

$$E_{\mathbf{y};\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right] \geq \mathbf{CRB}(\boldsymbol{\theta}) = \mathbf{F}(\boldsymbol{\theta})^{-1}, \tag{5a}$$

$$\mathbf{F}(\boldsymbol{\theta}) = E_{\mathbf{y};\boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right], \tag{5b}$$

where $\mathbf{F}(\boldsymbol{\theta})$ is the Fisher information matrix (FIM). Last, it has been shown in [20] that the CRB (5a) is also obtained if (3b) is reduced to

$$E_{\mathbf{y};\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\frac{\partial \ln p(\mathbf{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right]=\mathbf{I}. \tag{6}$$

B. Random Equality Constraints

In many estimation problems [1], [22]–[24], the probabilistic mapping mentioned above results from a two steps mechanism involving an additional random vector $\boldsymbol{\theta}_r$, $\boldsymbol{\theta}_r \in \Theta_r \subset \mathbb{R}^{P_r}$, that is i) $\boldsymbol{\theta} \to \boldsymbol{\theta}_r \sim p(\boldsymbol{\theta}_r;\boldsymbol{\theta})$, ii) $(\boldsymbol{\theta},\boldsymbol{\theta}_r) \to \mathbf{y} \sim p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})$, and leading to a compound probability distribution:

$$p(\mathbf{y},\boldsymbol{\theta}_r;\boldsymbol{\theta})=p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})\,p(\boldsymbol{\theta}_r;\boldsymbol{\theta}), \tag{7a}$$

$$p(\mathbf{y};\boldsymbol{\theta})=\int_{\Theta_r} p(\mathbf{y},\boldsymbol{\theta}_r;\boldsymbol{\theta})\,d\boldsymbol{\theta}_r, \tag{7b}$$

where $p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})$ is the conditional p.d.f. of $\mathbf{y}$ given $\boldsymbol{\theta}_r$, and $p(\boldsymbol{\theta}_r;\boldsymbol{\theta})$ is the prior p.d.f., parameterized by $\boldsymbol{\theta}$. In this setting[2],

$$\begin{aligned}&E_{\mathbf{y};\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\\&=E_{\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\right],\end{aligned} \tag{8}$$

which allows one to consider K non-redundant equality constraints on the unknown deterministic parameter vector $\boldsymbol{\theta}$ depending on the random parameter vector $\boldsymbol{\theta}_r$, that is

$$\mathbf{f}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})=\mathbf{0},\ \mathbf{f}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\in\mathbb{R}^K,\ 1\le K\le P-1, \tag{9}$$

where the matrix $\frac{\partial \mathbf{f}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\in\mathbb{R}^{K\times P}$ has full row rank.

C. CRBs with Random Equality Constraints

Since the set of K equality constraints (9) $\mathcal{C}_{\boldsymbol{\theta}_r}\subset\Theta_d$ is conditioned on the value of $\boldsymbol{\theta}_r$, it seems sensible to first look for a CR-like bound conditioned on $\boldsymbol{\theta}_r$, taking into account both local unbiasedness and equality constraints (9). Conditionally to $\boldsymbol{\theta}_r$, that is with respect to $p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})$, local unbiasedness regarding the parameter vector $\boldsymbol{\theta}$ reads

$$E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}+d\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\right]=\boldsymbol{\theta}+d\boldsymbol{\theta}+o_{\boldsymbol{\theta}_r}(\|d\boldsymbol{\theta}\|),$$

and leads (similarly to (3a) and (6)) to the LCs

$$E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right]d\boldsymbol{\theta}=\mathbf{I}d\boldsymbol{\theta}. \tag{10}$$

Moreover, if $\boldsymbol{\theta}$ and $\boldsymbol{\theta}+d\boldsymbol{\theta}$ are constrained to belong to $\mathcal{C}_{\boldsymbol{\theta}_r}$, thus, with some manipulation [20], when $\|d\boldsymbol{\theta}\|\to 0$,

$$\left\{\begin{array}{c}\mathbf{f}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})=\mathbf{0}\\ \frac{\partial \mathbf{f}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}d\boldsymbol{\theta}=\mathbf{0}\end{array}\right.\Leftrightarrow\left\{\begin{array}{c}\mathbf{0}=\mathbf{f}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\\ d\boldsymbol{\theta}=\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,d\boldsymbol{\lambda}\end{array}\right.$$

where $\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\in\mathbb{R}^{P\times(P-K)}$ is a basis of $\ker\left(\frac{\partial \mathbf{f}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right)$ and $d\boldsymbol{\lambda}\in\mathbb{R}^{P-K}$. Therefore, conditionally to $\boldsymbol{\theta}_r$, a locally unbiased estimate of $\boldsymbol{\theta}$ is now required to be locally unbiased only on $\mathcal{C}_{\boldsymbol{\theta}_r}$, what means that LCs (10) must be satisfied only when $d\boldsymbol{\theta}=\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,d\boldsymbol{\lambda}$ where $\|d\boldsymbol{\lambda}\|\to 0$, which yields the LCs

$$E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\mathbf{U}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta})\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right]=\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}). \tag{11}$$

Additionally, another desirable property is that

$$E_{\mathbf{y}|\boldsymbol{\theta}_r+d\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\right]=\boldsymbol{\theta}+o(\|d\boldsymbol{\theta}_r\|),\ \forall\boldsymbol{\theta}\in\mathcal{C}_{\boldsymbol{\theta}_r}, \tag{12a}$$

which means that, up to the first order and in the neighborhood of $\boldsymbol{\theta}_r$, $\widehat{\boldsymbol{\theta}}$ remains an unbiased estimator of $\boldsymbol{\theta}\in\mathcal{C}_{\boldsymbol{\theta}_r}$ independently of a small variation of the parameter vector $\boldsymbol{\theta}_r$. Once again, (12a) can be rewritten in terms of the following LCs

$$E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r^T}\right]=\mathbf{0}. \tag{12b}$$

Finally, conditionally to $\boldsymbol{\theta}_r$, a constrained CR-like bound fitted to the problem at hand is the lower bound associated with the LCs

$$\left\{\begin{array}{c}E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\left(\mathbf{U}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta})\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right]=\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}),\\ E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r^T}\right]=\mathbf{0},\end{array}\right. \tag{13}$$

that is, according to (4),

$$\widetilde{\mathbf{CCRB}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})= \tag{14a}$$

$$\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\left(\mathbf{U}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta})\,\mathbf{CRB}_{\boldsymbol{\theta}_r}^{-1}(\boldsymbol{\theta})\,\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right)^{-1}\mathbf{U}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta}),$$

$$\mathbf{CRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})= \tag{14b}$$

$$\left(\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})-\mathbf{F}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta}_r,\boldsymbol{\theta})\,\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)^{-1}\,\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r,\boldsymbol{\theta})\right)^{-1},$$

$$\begin{aligned}\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})&=E_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right],\\ \mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)&=E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r^T}\right],\\ \mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r,\boldsymbol{\theta})&=E_{\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right].\end{aligned}$$

Finally, if $\widehat{\boldsymbol{\theta}}\triangleq\widehat{\boldsymbol{\theta}}(\mathbf{y})$ is, conditionally to $\boldsymbol{\theta}_r$, a locally unbiased estimate belonging to a subset $\mathcal{C}_{\boldsymbol{\theta}_r}$ of the parameter space defined by K non redundant equality constraints depending on a random parameter vector $\boldsymbol{\theta}_r$, then, according to (8), its MSE matrix is lower bounded by the following randomly constrained CRB (RCCRB)

$$\begin{aligned}&E_{\mathbf{y}|\boldsymbol{\theta}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\ge\mathbf{RCCRB}(\boldsymbol{\theta}),\\ &\mathbf{RCCRB}(\boldsymbol{\theta})=E_{\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\widetilde{\mathbf{CCRB}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right].\end{aligned} \tag{15}$$

D. Further considerations

First, if no random constraints are taken into account, then $\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})=\mathbf{I}$ and $\widetilde{\mathbf{CCRB}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})=\mathbf{CRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})$ which coincides with the tighter Non-Standard CRB ($\overline{\mathbf{NSCRB}}(\boldsymbol{\theta})$) [2, (54)]. Moreover, the LCs (13) becomes equivalent to

$$E_{\mathbf{y}|\boldsymbol{\theta}_r+d\boldsymbol{\theta}_r;\boldsymbol{\theta}+d\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\right]=\boldsymbol{\theta}+d\boldsymbol{\theta}+o(\|(d\boldsymbol{\theta};d\boldsymbol{\theta}_r)\|),$$

which is the definition of a locally strict-sense unbiased estimator. This is reasonable since, as shown in [2, §IV],

[2]If only an integral form of $p(\mathbf{y};\boldsymbol{\theta})$ (7b) is available, the estimation problem at hand is so-called a "non-standard" estimation problem [2].

Non-Standard CRB are LBs on the "non-standard" MLEs (NSMLEs) defined as

$$(\widehat{\underline{\boldsymbol{\theta}}_r}, \widehat{\underline{\boldsymbol{\theta}}}) = \arg \max_{\boldsymbol{\theta}\in\Theta_d, \boldsymbol{\theta}_r\in\Theta_r} \left\{ p\left(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)\right\}, \tag{16}$$

where $\widehat{\underline{\boldsymbol{\theta}}}$ is, w.r.t. $p\left(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)$ and under reasonably general conditions, asymptotically uniformly strict-sense unbiased, Gaussian distributed and efficient when the number of independent observations tends to infinity. Therefore, it seems likely that the method of scoring with parameter constraints [19] applied to random equality constraints (9) where $\boldsymbol{\theta}_r$ is replaced with its NSMLE $\widehat{\underline{\boldsymbol{\theta}}_r}$, leads to a constrained NSMLE asymptotically efficient with respect to $\widetilde{\mathbf{CCRB}}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right)$ and hence to $\mathbf{RCCRB}\left(\boldsymbol{\theta}\right)$ — a conjecture left for future research.

Second, in general,

$$\begin{aligned}&\widetilde{\mathbf{CCRB}}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right)\\ &> \mathbf{U}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right)\left(\mathbf{U}_{\boldsymbol{\theta}_r}^T\left(\boldsymbol{\theta}\right)\mathbf{F}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right)\mathbf{U}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right)\right)^{-1}\mathbf{U}_{\boldsymbol{\theta}_r}^T\left(\boldsymbol{\theta}\right),\end{aligned}$$

which means that the RCCRB proposed (15) is tighter than the expectation of the standard CCRB parameterized by $\boldsymbol{\theta}_r$ (see (17)). However, in the case where $p\left(\mathbf{y},\boldsymbol{\theta}_r;\boldsymbol{\theta}\right) = p\left(\mathbf{y};\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)$, then

$$p\left(\mathbf{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right) = p\left(\mathbf{y};\boldsymbol{\theta}\right) \Rightarrow \mathbf{CRB}_{\boldsymbol{\theta}_r}^{-1}\left(\boldsymbol{\theta}\right) = \mathbf{F}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right) = \mathbf{F}\left(\boldsymbol{\theta}\right),$$

where $\mathbf{F}\left(\boldsymbol{\theta}\right)$ is the standard FIM (5b), leading to

$$\begin{aligned}&\mathbf{RCCRB}\left(\boldsymbol{\theta}\right) =\\ &E_{\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\mathbf{U}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right)\left(\mathbf{U}_{\boldsymbol{\theta}_r}^T\left(\boldsymbol{\theta}\right)\mathbf{F}\left(\boldsymbol{\theta}\right)\mathbf{U}_{\boldsymbol{\theta}_r}\left(\boldsymbol{\theta}\right)\right)^{-1}\mathbf{U}_{\boldsymbol{\theta}_r}^T\left(\boldsymbol{\theta}\right)\right],\end{aligned} \tag{17}$$

which reduces to the standard CCRB

$$\mathbf{CCRB}\left(\boldsymbol{\theta}\right) = \mathbf{U}\left(\boldsymbol{\theta}\right)\left(\mathbf{U}^T\left(\boldsymbol{\theta}\right)\mathbf{F}\left(\boldsymbol{\theta}\right)\mathbf{U}\left(\boldsymbol{\theta}\right)\right)^{-1}\mathbf{U}^T\left(\boldsymbol{\theta}\right),$$

if the K equality constraints (9) are non random.

III. Constrained Cramér-Rao bounds for coupled LL1 models

A. Basic observation model

We consider two tensors $\boldsymbol{\mathcal{Y}}_1 \in \mathbb{R}^{I_1\times J_1\times K_1}$ and $\boldsymbol{\mathcal{Y}}_2 \in \mathbb{R}^{I_2\times J_2\times K_2}$. The dimension in the third mode of $\boldsymbol{\mathcal{Y}}_2$ is lower than that of $\boldsymbol{\mathcal{Y}}_1$ ($K_2 \ll K_1$), while its dimensions in the first and second mode are higher ($I_2 > I_1,\ J_2 > J_1$). For the sake of simplicity, in the remainder of this paper we will adopt the following notations: $I = I_2,\ J = J_2,\ K = K_1$.

The observations $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ can be viewed as two degraded versions of the same tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$ with high dimensions in all modes. In this setting, we adopt the following tensor degradation model that can be compactly written as mode product of $\boldsymbol{\mathcal{Y}}$ with some degradation matrices $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$:

$$\begin{cases}\boldsymbol{\mathcal{Y}}_1 &= \boldsymbol{\mathcal{Y}} \bullet_1 \mathbf{P} \bullet_2 \mathbf{Q} + \boldsymbol{\mathcal{E}}_1,\\ \boldsymbol{\mathcal{Y}}_2 &= \boldsymbol{\mathcal{Y}} \bullet_3 \mathbf{R} + \boldsymbol{\mathcal{E}}_2,\end{cases} \tag{18}$$

where $\mathbf{P} \in \mathbb{R}^{I_1\times I}$, $\mathbf{Q} \in \mathbb{R}^{J_1\times J}$, and $\mathbf{R} \in \mathbb{R}^{K_2\times K}$ have full row rank. The entries of the noise terms $\boldsymbol{\mathcal{E}}_1 \sim \boldsymbol{\mathcal{N}}(\mathbf{0},\boldsymbol{\Sigma}_1)$, $\boldsymbol{\mathcal{E}}_2 \sim \boldsymbol{\mathcal{N}}(\mathbf{0},\boldsymbol{\Sigma}_2)$ are independent and identically distributed (i.i.d.) real Gaussian tensors with zero mean and variances $\boldsymbol{\Sigma}_1 = \sigma_1^2\mathbf{I}$ and $\boldsymbol{\Sigma}_2 = \sigma_2^2\mathbf{I}$, respectively.

Model (18) represents an ill-posed inverse problem, whose goal is to recover the tensor $\boldsymbol{\mathcal{Y}}$ from the observations $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$. It was used in the literature to address several reconstruction problems. For instance, in remote sensing [30], $\mathbf{P}$ and $\mathbf{Q}$ are blurring and downsampling matrices, while $\mathbf{R}$ contains the spectral response of the sensor used to acquire $\boldsymbol{\mathcal{Y}}_2$, which refers to a multispectral observation. In medical imaging [54] and spectrum cartography [55] applications, the degradation matrices select fibers of the underlying tensor $\boldsymbol{\mathcal{Y}}$ in a given mode so that the observations represent downsampled versions of an observation cube.

As in [40], [41], [55], the degradation model (18) is reformulated as a coupled LL1-BTD as

$$\begin{cases}\boldsymbol{\mathcal{Y}}_1 = \sum_{r=1}^{R}\left((\mathbf{A}_1)_r(\mathbf{B}_1)_r^T\right)\otimes(\mathbf{c}_1)_r + \boldsymbol{\mathcal{E}}_1,\\ \boldsymbol{\mathcal{Y}}_2 = \sum_{r=1}^{R}\left((\mathbf{A}_2)_r(\mathbf{B}_2)_r^T\right)\otimes(\mathbf{c}_2)_r + \boldsymbol{\mathcal{E}}_2,\end{cases} \tag{19}$$

$$\text{where } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2, \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2, \mathbf{C}_2 = \mathbf{R}\mathbf{C}_1. \tag{20}$$

The LL1 factors of the model are such that $\mathbf{A}_1 \in \mathbb{R}^{I_1\times LR}$, $\mathbf{B}_1 \in \mathbb{R}^{J_1\times LR}$, $\mathbf{C}_1 \in \mathbb{R}^{K\times R}$ and $\mathbf{A}_2 \in \mathbb{R}^{I\times LR}$, $\mathbf{B} \in \mathbb{R}^{J\times LR}$, $\mathbf{C}_2 \in \mathbb{R}^{K_2\times R}$. Since the factors are coupled in all the modes, (19)–(20) is referred to as a fully-coupled model.

Under these notations, the underlying tensor $\boldsymbol{\mathcal{Y}}$ of interest admits an LL1-BTD as

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R}\left((\mathbf{A}_2)_r(\mathbf{B}_2)_r^T\right)\otimes(\mathbf{c}_1)_r. \tag{21}$$

B. Model parameters

We consider two model parameters $\widetilde{\boldsymbol{\omega}} \in \mathbb{R}^{((I+J)L+K)R}$ and $\boldsymbol{\phi} \in \mathbb{R}^{((I_1+J_1)L+K_2)R}$ that describe (19)–(20) such that

$$\widetilde{\boldsymbol{\omega}}^T = \left[\text{vec}\{\mathbf{A}_2\}^T\ \ \text{vec}\{\mathbf{B}_2\}^T\ \ \text{vec}\{\mathbf{C}_1\}^T\right], \tag{22}$$

$$\boldsymbol{\phi}^T = \left[\text{vec}\{\mathbf{A}_1\}^T\ \ \text{vec}\{\mathbf{B}_1\}^T\ \ \text{vec}\{\mathbf{C}_2\}^T\right]. \tag{23}$$

While $\widetilde{\boldsymbol{\omega}}$ represents the LL1 factors underlying $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\phi}$ contains the entries of the factors obtained by degradation in (20). The two parameters in (22)–(23) can be stacked together into a single parameter $\widetilde{\boldsymbol{\theta}}$ defined as

$$\widetilde{\boldsymbol{\theta}}^T = \left[\widetilde{\boldsymbol{\omega}}^T\ \boldsymbol{\phi}^T\right]. \tag{24}$$

In tensor reconstruction applications, we are mostly interested in the LL1 factors underlying $\boldsymbol{\mathcal{Y}}$, contained in the parameter $\widetilde{\boldsymbol{\omega}}$. In this scenario, the fully-coupled model can be statistically identifiable even if the trivial ambiguities in $\boldsymbol{\phi}$ are neglected, see [33] for more details. The remaining ambiguities in $\widetilde{\boldsymbol{\omega}}$ are solved by setting $(\mathbf{C}_1)_{1,:} = 1$ and $(\mathbf{A}_2)_{1:L,:} = [\mathbf{I}_L\ \ldots\ \mathbf{I}_L]$ as discussed in Section I-C. As a result, we must define the reduced parameter $\boldsymbol{\omega} \in \mathbb{R}^{((I+J-L)L+(K-1))R}$ such that

$$\boldsymbol{\omega}^T = \left[\text{vec}\{(\mathbf{A}_2)_{L+1:I,:}\}^T\ \ \text{vec}\{\mathbf{B}_2\}^T\ \ \text{vec}\{(\mathbf{C}_1)_{2:K,:}\}^T\right], \tag{25}$$

that only contains the unknown entries of $\widetilde{\boldsymbol{\omega}}$. The full and reduced parameters can be linked through the relationship

$\boldsymbol{\omega} = \mathbf{M}\widetilde{\boldsymbol{\omega}}$. The matrix $\mathbf{M}$ is a selection matrix constructed from $\mathbf{I}_{((I+J)L+K)R}$ by removing the $(L^2+1)R$ rows corresponding to the known entries of $\widetilde{\boldsymbol{\omega}}$. Therefore, the global model parameter $\widetilde{\boldsymbol{\theta}}$ can be reduced to $\boldsymbol{\theta}$ defined as

$$\boldsymbol{\theta}^T = \begin{bmatrix}\boldsymbol{\omega}^T & \boldsymbol{\phi}^T\end{bmatrix}. \tag{26}$$

C. Fisher information matrix for coupled LL1 models

In this subsection, we derive a closed-form expression for the Fisher information matrix (FIM) related to model (19). We consider the random real Gaussian distributed dataset $\mathbf{x} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}}), \boldsymbol{\Sigma})$, where

$$\mathbf{x}^T = \begin{bmatrix}\mathrm{vec}\{\boldsymbol{\mathcal{Y}}_1\}^T & \mathrm{vec}\{\boldsymbol{\mathcal{Y}}_2\}^T\end{bmatrix}, \quad \boldsymbol{\Sigma} = \mathrm{Diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\},$$

$$\text{and } \boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}}) = \begin{bmatrix}\boldsymbol{\mu}_1(\widetilde{\boldsymbol{\theta}}) \\ \boldsymbol{\mu}_2(\widetilde{\boldsymbol{\theta}}).\end{bmatrix} \tag{27}$$

In (27), the mean $\boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}})$ is a block matrix and for $i = 1, 2$, its subblocks are such that

$$\boldsymbol{\mu}_i(\widetilde{\boldsymbol{\theta}}) = \mathrm{vec}\left\{\sum_{r=1}^{R}\left((\mathbf{A}_i)_r(\mathbf{B}_i)_r^T\right) \otimes (\mathbf{c}_i)_r\right\}. \tag{28}$$

The derivatives of $\boldsymbol{\mu}_i(\widetilde{\boldsymbol{\theta}})$ with respect to $\widetilde{\boldsymbol{\theta}}$ can be obtained using relationships between tensor unfoldings as

$$\begin{aligned}\boldsymbol{\mu}_i(\widetilde{\boldsymbol{\theta}}) &= [(\mathbf{C}_i \odot_p \mathbf{B}_i) \boxtimes \mathbf{I}]\,\mathrm{vec}\{\mathbf{A}_i\}, &(29)\\ &= \boldsymbol{\Pi}_i^{(2,1)}\,[(\mathbf{C}_i \odot_p \mathbf{A}_i) \boxtimes \mathbf{I}]\,\mathrm{vec}\{\mathbf{B}_i\}, &(30)\\ &= \boldsymbol{\Pi}_i^{(3,1)}\,([\ldots, ((\mathbf{A}_i)_r \odot (\mathbf{B}_i)_r)\,\mathbf{1}_L, \ldots] \boxtimes \mathbf{I})\,\mathrm{vec}\{\mathbf{C}_i\}. &(31)\end{aligned}$$

The matrices $\boldsymbol{\Pi}_i^{(2,1)}$ and $\boldsymbol{\Pi}_i^{(3,1)}$ are non-singular permutation matrices that link the entries of $\mathrm{vec}\{\mathbf{Y}_i^{(2)}\}$ (resp. $\mathrm{vec}\{\mathbf{Y}_i^{(3)}\}$) to those of $\mathrm{vec}\{\boldsymbol{\mathcal{Y}}_i\}$ $(i = 1, 2)$.

The Fisher information matrix for $\widetilde{\boldsymbol{\theta}}$ is obtained by using the Slepian-Bangs formula below [39]:

$$\mathbf{F}(\widetilde{\boldsymbol{\theta}}) = \left[\frac{\partial \boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}})}{\partial \widetilde{\boldsymbol{\theta}}^T}\right]^T \boldsymbol{\Sigma}^{-1} \left[\frac{\partial \boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}})}{\partial \widetilde{\boldsymbol{\theta}}^T}\right], \tag{32}$$

and the FIM for the parameter $\boldsymbol{\theta}$ of interest is obtained from (32) by removing the indices corresponding to known entries of $\widetilde{\boldsymbol{\theta}}$ as

$$\mathbf{F}(\boldsymbol{\theta}) = \mathrm{Diag}\{\mathbf{M}^T, \mathbf{I}\}\mathbf{F}(\widetilde{\boldsymbol{\theta}})\,\mathrm{Diag}\{\mathbf{M}, \mathbf{I}\}. \tag{33}$$

Its closed-form expression can be found in Appendix A.

D. Standard constrained Cramér-Rao bound

The deterministic constraints (20) between the LL1 factors can be expressed in terms of model parameters as

$$\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\phi} - \underbrace{\begin{bmatrix}\mathbf{I}\boxtimes\mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\boxtimes\mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\boxtimes\mathbf{R}\end{bmatrix}\mathbf{M}^T}_{\mathbf{G}}\boldsymbol{\omega}, \tag{34}$$

where $\mathbf{g}$ is a non-redundant deterministic vector function, differentiable for all $\boldsymbol{\omega}$.

Thus a basis for $\ker\left\{\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right\}$ is the matrix $\mathbf{U}$ such that

$$\mathbf{U}^T(\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{I} & \mathbf{G}^T\end{bmatrix}, \tag{35}$$

and the standard CCRB for the parameter $\boldsymbol{\theta}$ [20] is:

$$\mathbf{CCRB}(\boldsymbol{\theta}) = \mathbf{U}(\boldsymbol{\theta})\left(\mathbf{U}^T(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})\right)^{-1}\mathbf{U}^T(\boldsymbol{\theta}). \tag{36}$$

Similarly to the FIM, $\mathbf{CCRB}(\boldsymbol{\theta})$ is a block-matrix; the closed-form expressions for its subblocks are derived in Appendix A.

E. Performance bounds for tensor reconstruction

Additionally to the model parameters in (22)–(23), we also define $\mathbf{y} = \mathrm{vec}\{\boldsymbol{\mathcal{Y}}\} \in \mathbb{R}^{IJK}$, which represents the vectorized tensor $\boldsymbol{\mathcal{Y}}$ that we wish to approximate. The parameter $\mathbf{y}$ can be linked to $\boldsymbol{\omega}$ by means of a non-redundant vector function $\mathbf{h}$, differentiable for all $\boldsymbol{\omega}$ as

$$\mathbf{y} = \mathbf{h}(\boldsymbol{\omega}). \tag{37}$$

The expression of $\mathbf{h}(\boldsymbol{\omega})$ is obtained similarly to (29)–(31):

$$\begin{aligned}\mathbf{y} &= \underbrace{[(\mathbf{C}_1 \odot_p \mathbf{B}_2) \boxtimes \mathbf{I}]}_{\mathbf{S}^{(1)}}\mathrm{vec}\{\mathbf{A}_2\},\\ &= \underbrace{\boldsymbol{\Pi}^{(2,1)}\,[(\mathbf{C}_1 \odot_p \mathbf{A}_2) \boxtimes \mathbf{I}]}_{\mathbf{S}^{(2)}}\mathrm{vec}\{\mathbf{B}_2\},\\ &= \underbrace{\boldsymbol{\Pi}^{(3,1)}\,([\ldots, ((\mathbf{A}_2)_r \odot (\mathbf{B}_2)_r)\,\mathbf{1}_L, \ldots] \boxtimes \mathbf{I})}_{\mathbf{S}^{(3)}}\mathrm{vec}\{\mathbf{C}_1\}.\end{aligned}$$

As a result, we have

$$\mathbf{h}(\boldsymbol{\omega}) = \begin{bmatrix}\mathbf{S}^{(1)} & \mathbf{S}^{(2)} & \mathbf{S}^{(3)}\end{bmatrix}\mathbf{M}^T\boldsymbol{\omega}.$$

Hence we can obtain the CCRB for the parameter $\mathbf{y}$ using the following formula:

$$\mathbf{CCRB}(\mathbf{y}) = \left[\frac{\partial \mathbf{h}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T}\right]\mathbf{CCRB}(\boldsymbol{\omega})\left[\frac{\partial \mathbf{h}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T}\right]^T, \tag{38}$$

where $\mathbf{CCRB}(\boldsymbol{\omega})$ is the left upmost diagonal block of $\mathbf{CCRB}(\boldsymbol{\theta})$ and $\left[\frac{\partial \mathbf{h}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T}\right] = \begin{bmatrix}\mathbf{S}^{(1)} & \mathbf{S}^{(2)} & \mathbf{S}^{(3)}\end{bmatrix}\mathbf{M}^T$.

IV. Performance analysis in the case of deterministic constraints

A. Estimation

We now conduct performance analysis for a state-of-the-art algorithm in the case of deterministic constraints. According to the fully-coupled model (19)–(20), and since the entries of the noise terms $\boldsymbol{\mathcal{E}}_1$ and $\boldsymbol{\mathcal{E}}_2$ are i.i.d., the observations $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are distributed according to

$$\begin{cases}\mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\boldsymbol{\theta}} & = \left(2\pi\sigma_1^2\right)^{\frac{-I_1J_1K}{2}} \\ & \quad e^{\left(-\frac{1}{2\sigma_1^2}\|\boldsymbol{\mathcal{Y}}_1 - \sum_{r=1}^{R}\left(\mathbf{P}(\mathbf{A}_2)_r(\mathbf{Q}(\mathbf{B}_2)_r)^T\right)\otimes(\mathbf{c}_1)_r\|_F^2\right)}, \\ \mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\boldsymbol{\theta}} & = \left(2\pi\sigma_2^2\right)^{\frac{-IJK_2}{2}} \\ & \quad e^{\left(-\frac{1}{2\sigma_2^2}\|\boldsymbol{\mathcal{Y}}_2 - \sum_{r=1}^{R}\left((\mathbf{A}_2)_r(\mathbf{B}_2)_r^T\right)\otimes\mathbf{R}(\mathbf{c}_1)_r\|_F^2\right)}.\end{cases} \tag{39}$$

Estimation of the LL1 factors in this setting is performed by a coupled alternating least squares (ALS) algorithm inspired

by [39], [57], that we further refer to as LL1-ALS. This algorithm was used in [41] in a remote sensing context for solving the hyperspectral super-resolution problem. It minimizes the following global criterion:

$$\min_{\mathbf{A}_2,\mathbf{B}_2,\mathbf{C}_1} \|\boldsymbol{\mathcal{Y}}_1 - \sum_{r=1}^{R} \left(\mathbf{P}(\mathbf{A}_2)_r(\mathbf{Q}(\mathbf{B}_2)_r)^T\right) \otimes (\mathbf{c}_1)_r\|_F^2 + \lambda\|\boldsymbol{\mathcal{Y}}_2 - \sum_{r=1}^{R} \left((\mathbf{A}_2)_r(\mathbf{B}_2)_r^T\right) \otimes \mathbf{R}(\mathbf{c}_1)_r\|_F^2. \quad (40)$$

Under the assumption of Gaussian noise and residuals, and assuming that $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$, (40) corresponds to the maximum likelihood (ML) criterion for coupled $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ distributed as in (39).

B. Experimental setup

We consider the dimensions $I_1 = J_1 = 6$, $I = J = 24$, $K = 30$ and $K_2 = 6$. The LL1 ranks are $L = 2$ and $R = 2$. The entries of the true factors $\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1$ were generated once as i.i.d. real standard Gaussian variables, for which scaling and matrix ambiguities were corrected as in Section I-C by fixing some entries of $\mathbf{C}_1$ and $\mathbf{A}_2$ to known values. The true factors $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ were constructed according to the deterministic constraints (20).

The degradation matrices $\mathbf{P}$ and $\mathbf{Q}$ are Gaussian blurring and downsampling matrices, generated following Wald's protocol [38] with a Gaussian filter of length q and a downsampling ratio d. For the sake of simplicity but without loss of generality, we also assume that $\mathbf{P} = \mathbf{Q}$. The degradation matrix $\mathbf{R}$ is a selection and averaging matrix constructed from the Sentinel-2 spectral response function[3]. Please refer to Appendix B for more details on the construction of these matrices This specific degradation scenario was used in e.g., [30], [56] to describe the hyperspectral super-resolution problem [29].

We simulate the performance of the coupled model under additive Gaussian noise. The SNR (in dB) is defined as $SNR_i = 10\log_{10}\left(\|\boldsymbol{\mathcal{Y}}_i\|_F^2/\|\boldsymbol{\mathcal{E}}_i\|_F^2\right)$, $(i = 1, 2)$. We fix SNR_2 to 20dB while SNR_1 is a vector with values in $\{5, 60\}$ dB. In our simulations, we consider performance analysis for various values of SNR_1 and fixed SNR_2.

The model parameters are retrieved using the ML estimator LL1-ALS with random initialization; for each realization, the best out of 10 trials is picked. We evaluate the total MSE on the estimated parameters $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{y}}$ by averaging the squared errors over 500 noise realizations. The permutation ambiguities in the estimated factors are corrected by searching for the best column permutation of $\mathbf{C}_2$ with fixed $\mathbf{C}_1$ and applying that same permutation to $\mathbf{A}_2$ and $\mathbf{B}_2$. This step is performed by merely maximizing the correlation between the estimated and true LL1 factors; but it could be performed optimally using the Hungarian algorithm [53].

C. Numerical results

We compute the standard CCRB for the parameters $\boldsymbol{\theta}$ and $\mathbf{y}$. In our experiments, we consider the uniform MSE (UMSE) and uniform CCRB (UCCRB) based on the matrix traces[4], as widely studied, e.g., in [60]–[62]. Indeed, it is easy to see that if a locally unbiased estimator is uniformly efficient (*i.e.*, the UMSE reaches the corresponding uniform bound), then it is also efficient for each entry of the parameters, which is a strong result. In Figure 1, we show on a semi-log scale the UCCRB and UMSE given by LL1-ALS for the parameters $\boldsymbol{\theta}$ and $\mathbf{y}$, respectively.

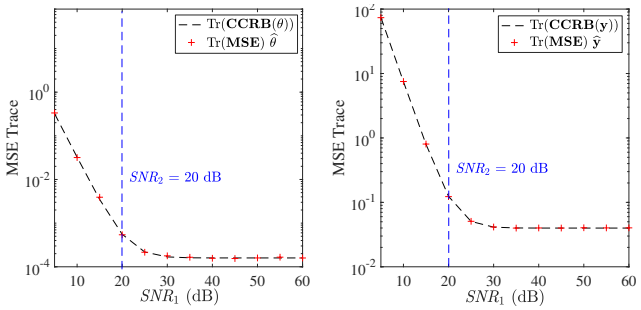

Fig. 1. **UCCRB** and **UMSE** for estimation of $\boldsymbol{\theta}$ (left) and $\mathbf{y}$ (right).

For both parameters $\boldsymbol{\theta}$ and $\mathbf{y}$, the UMSE reaches the UCCRB, showing that LL1-ALS is asymptotically efficient for LL1 factors estimation and tensor reconstruction in the presence of non-random constraints. However, this performance analysis only holds in the absence of uncertainties. Indeed, from (40), we can see that the current framework for LL1-ALS does not allow to take uncertainties into account. In the following section, we will consider a more general degradation model accounting for a random variability phenomenon.

V. Degradation model accounting for uncertainties

A. A more flexible model

The tensor degradation model (18) does not account for any uncertainties. However in practice, since $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are acquired by different sensors, they are usually obtained at different time instants. For instance, in remote sensing, the acquisition time difference can result in e.g., variations in atmospheric, seasonal or illumination conditions [52], [56], [58]. More generally, proper calibration of the sensor is crucial in order to account for the sensor specificities such as degradation of sensitivity over time [45], normalization of each channel response, or variations in observational geometry [46]. This step is usually performed before launching the sensor [49]. However, its specificities may change due to e.g., outgassing, aging of components, or misalignment, thus resulting in miscalibration. These uncertainties motivate the need for more flexible models. As a result, in the remaining of this paper, we will consider an observation model that is different from (18).

As in [56], [59], we now consider that $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are degraded versions of two distinct tensors $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$ and $\widetilde{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$, respectively. While (18) assumed that $\widetilde{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{Y}}$, this new model allows $\widetilde{\boldsymbol{\mathcal{Y}}}$ and $\boldsymbol{\mathcal{Y}}$ to be different. The tensor $\widetilde{\boldsymbol{\mathcal{Y}}}$

[3]Available for download at https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library.

[4]In fact, the bounds expressions proposed in this paper allow for calculation of the uniform CCRB and uniform RCCRB due to their synthetic form.

also admits an LL1-BTD with the same factors $\mathbf{A}_2$ and $\mathbf{B}_2$ as $\boldsymbol{\mathcal{Y}}$, but with a different factor $\widetilde{\mathbf{C}}_1 \in \mathbb{R}^{K\times R}$:

$$\widetilde{\boldsymbol{\mathcal{Y}}} = \sum_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^T\right) \otimes (\widetilde{\mathbf{c}}_1)_r. \tag{41}$$

This allows for the addition of uncertainties between the two tensors, depicted through the LL1 factor $\widetilde{\mathbf{C}}_1 \neq \mathbf{C}_1$.

We can model these uncertainties using the following multiplicative model:

$$\widetilde{\mathbf{C}}_1 = \mathbf{C}_1 \boxdot \boldsymbol{\Psi}, \tag{42}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{K\times R}$ is a matrix of scaling factors. This leads to the following model:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 &= \boldsymbol{\mathcal{Y}} \bullet_1 \mathbf{P} \bullet_2 \mathbf{Q} + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 &= \widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{R} + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{43}$$

that newly considers uncertainties in the acquisition conditions. Given (43), the coupled LL1 model accounting for uncertainties can be expressed as:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_1)_r (\mathbf{B}_1)_r^T\right) \otimes (\mathbf{c}_1)_r + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^T\right) \otimes (\mathbf{c}_2)_r + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{44}$$

$$\text{where } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2, \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2, \mathbf{C}_2 = \mathbf{R}(\mathbf{C}_1 \boxdot \boldsymbol{\Psi}). \tag{45}$$

Different from (20), the constraints (45) on the LL1 factors involve a random matrix $\boldsymbol{\Psi}$.

B. A clairvoyant algorithm

Considering model (44)–(45), we would like to assess the performance of the fully-coupled LL1-ALS algorithm in the presence of uncertainties. In the following set of experiments, we will consider the following cost function, that is a modified version of (40):

$$\begin{aligned} \min_{\mathbf{A}_2,\mathbf{B}_2,\mathbf{C}_1} \; & \|\boldsymbol{\mathcal{Y}}_1 - \sum_{r=1}^{R} \left(\mathbf{P}(\mathbf{A}_2)_r (\mathbf{Q}(\mathbf{B}_2)_r)^T\right) \otimes (\mathbf{c}_1)_r\|_F^2 \\ & + \lambda \|\boldsymbol{\mathcal{Y}}_2 - \sum_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^T\right) \otimes \mathbf{R}\left((\mathbf{c}_1)_r \boxdot \Psi_r\right)\|_F^2, \end{aligned} \tag{46}$$

where Ψ_r is the r-th column of $\boldsymbol{\Psi}$. In fact, (46) corresponds to a clairvoyant criterion in which the value of the random parameter is supposed to be known. In this setting, the algorithm minimizing (46) will be further referred to as clairvoyant LL1-ALS.

C. Standard conditional CCRB and its limitations

Let us define the random parameter vector $\boldsymbol{\theta}_r = \text{vec}\{\boldsymbol{\Psi}\} \in \mathbb{R}^{KR}$ that characterizes uncertainties. The entries of $\boldsymbol{\theta}_r$ are i.i.d. Gaussian entries with unit mean and variance σ_r^2 such that $\boldsymbol{\theta}_r \sim \boldsymbol{\mathcal{N}}(\mathbf{1}, \sigma_r^2\mathbf{I})$, thus $\mathbf{C}_1 = \lim\limits_{\sigma_r^2 \to 0} \widetilde{\mathbf{C}}_1$.

The random equality constraints (45) involve the random parameter vector $\boldsymbol{\theta}_r$ and can be expressed as

$$\mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \boldsymbol{\phi} - \underbrace{\begin{bmatrix} \mathbf{I}\boxtimes\mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\boxtimes\mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{I}\boxtimes\mathbf{R})\,\text{diag}\{\boldsymbol{\theta}_r\} \end{bmatrix}}_{\mathbf{G}_{\boldsymbol{\theta}_r}} \mathbf{M}^T \boldsymbol{\omega}. \tag{47}$$

Hence we can compute the CCRB conditional to $\boldsymbol{\theta}_r$ as

$$\mathbf{CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \left(\mathbf{U}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta}) \mathbf{F}(\boldsymbol{\theta}) \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right)^{-1} \mathbf{U}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta}), \tag{48}$$

where $\mathbf{U}_{\boldsymbol{\theta}_r}^T(\boldsymbol{\theta}) = \left[\mathbf{I}\ \mathbf{G}_{\boldsymbol{\theta}_r}^T\right]$.

In this subsection, we compare the performance of the clairvoyant algorithm to the standard CCRB conditional to the value of the random parameter. We drew two realizations of the random parameter vector with variance $\sigma_r^2 = 0.2$, namely $\boldsymbol{\theta}_r^{(1)}$ and $\boldsymbol{\theta}_r^{(2)}$. For each realization, the entries of the true LL1 factors $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ were generated according to (45) with the same dimensions as in Section IV-B. We compute the standard UCCRB conditionally to the value of the random parameter, namely $\mathbf{UCCRB}^{(i)}$ $(i = 1, 2)$.

In Figure 2, we plot on a semi-log scale the conditional UCCRB and UMSE for parameters $\boldsymbol{\omega}$ and $\mathbf{y}$. We can see that

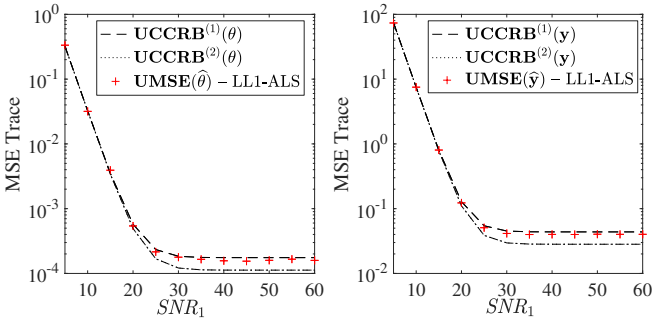

Fig. 2. Standard $\mathbf{UCCRB}^{(i)}$ $(i = 1, 2)$ and $\mathbf{UMSE}$ for $\boldsymbol{\theta}$ (left) and $\mathbf{y}$ (right).

the two UCCRB curves are almost equal for $SNR_1 \leq SNR_2$. For such SNR_1, the UMSE provided by LL1-ALS reaches the bounds. However, for $SNR_1 > SNR_2$, the discrepancy between the two UCCRB curves increases, and the UMSE can be found in-between the two bounds. In particular, Figure 2 shows that the performance of LL1-ALS can be lower than the conditional UCCRB, meaning that the standard tool becomes non-informative in the presence of random equality constraints.

Therefore we must consider a new constrained CR-type lower bound fitted to this context that is able to characterize the best achievable performance of our model, hence the introduction of the RCCRB in Section II. The relevance of this bound will be illustrated in the next subsection.

D. Usefulness of the RCCRB for the coupled LL1 model with uncertainties

1) Relevance of the RCCRB: In a preliminary version of this work [42], it is shown that the RCCRB is a lower bound on constrained parameter estimation when the constraints involve

a random vector parameter; hence it is suited for our problem. In this subsection, we illustrate the usefulness of the RCCRB for conducting performance analysis of the model (44)–(45).

The RCCRB is evaluated as in (17) by averaging the conditional CCRB over 500 realizations of the random parameter vector $\boldsymbol{\theta}_r$ with variance $\sigma_r^2 = 0.2$. We compare the RCCRB matrix trace to the performance of the clairvoyant estimator. In Figure 3, we plot on a semi-log scale the URCCRB and UMSE trace for $\boldsymbol{\theta}$ and $\mathbf{y}$.

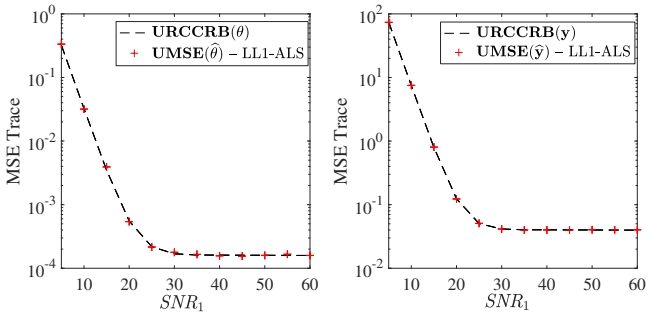

Fig. 3. **URCCRB** and **UMSE** for $\boldsymbol{\theta}$ (left) and $\mathbf{y}$ (right).

Contrary to the standard CCRB in Figure 2, the UMSE reaches the URCCRB for all considered values of SNR_1 and both parameters. Two conclusions can be drawn from Figure 2: first, that the RCCRB is indeed the right bound for assessing the performance of our model, and second, that the clairvoyant algorithm is asymptotically efficient in the presence of a random parameter vector impinging on the constraints, since it reaches the uniform RCCRB.

2) Performance loss in case of constraints mismatch: In this subsection, we investigate the performance of the clairvoyant algorithm in the case of constraints mismatch.

In the coupled model with uncertainties, $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$ are generated as in Section IV while $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ are generated according to (45) with $\boldsymbol{\theta}_r \sim \boldsymbol{\mathcal{N}}(\mathbf{1}, \sigma_r^2)$ and variance $\sigma_r^2 = 0.2$. We additionally draw two specific realizations of the random parameter vector with variance $\sigma_r^2 = 0.2$, namely $\boldsymbol{\theta}_r^{(1)}$ and $\boldsymbol{\theta}_r^{(2)}$. We run the clairvoyant LL1-ALS with a constraints mismatch, that is for $i = 1, 2$, (46) accounts for an incorrect estimation of $\boldsymbol{\theta}_r$.

We compare the UMSE obtained in this scenario, to the URCCRB and UMSE obtained by incorporating the correct $\boldsymbol{\theta}_r$ in (46). The uniform bounds and UMSE curves are shown in Figure 4.

We can see that, in the case of a constraints mismatch, the UMSE provided by LL1-ALS is higher than the URCCRB. This behaviour is particularly visible for high SNR. This means that an incorrect knowledge of the constraints leads to a loss in performance.

Thus, there are two main limitations to the use of such a clairvoyant algorithm. First, in practice, the variability phenomenon is difficult to estimate[5]; therefore, it seems to be intractable to incorporate it into the cost function (46). Then, the previous experiments showed that a wrong estimation of the random parameter can severely degrade the performance of the algorithm. Nevertheless, the clairvoyant LL1-ALS estimator and associated RCCRB remain important theoretical tools to assess the best achievable performance of the coupled model (44)–(45).

[5]In fact, [56, Theorem I.5] further indicates that it is only possible to recover a degraded version of the variability factor.

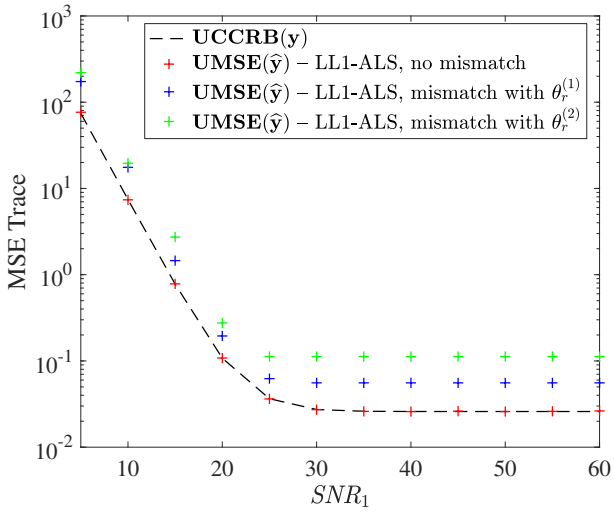

Fig. 4. URCCRB and UMSE from LL1-ALS in the case of constraints mismatch.

In the next subsection, we will introduce a more realistic estimator that is able to account for random uncertainties.

E. A blind algorithm accounting for random uncertainties

1) Blind LL1 model: Since the uncertainty phenomenon is unknown in practice, it is reasonable to consider a partially-coupled LL1 model, that we will further refer to as blind. This model considers the random part of the constraints to be unknown. It is also motivated by [56, Theorem I.5], which indicates that knowledge of the degradation between the $\mathbf{C}_1$ and $\widetilde{\mathbf{C}}_1$ factors is not needed to guarantee unique recovery of the target tensor $\boldsymbol{\mathcal{Y}}$ in the noiseless case.

Hence we propose to use a different model than (44)–(45), that supposes that only the degradation matrices $\mathbf{P}$ and $\mathbf{Q}$ are known while the matrix $\mathbf{R}$ is unknown:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 = \sum_{r=1}^{R} \left((\mathbf{A}_1)_r (\mathbf{B}_1)_r^T \right) \otimes (\mathbf{c}_1)_r + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 = \sum_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^T \right) \otimes (\mathbf{c}_2)_r + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{49}$$

$$\text{where } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2, \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2. \tag{50}$$

In (49)–(50), the $(\mathbf{c}_2)_r$ are degraded versions of $(\mathbf{c}_1)_r$ by unknown degradation and corrupted by random uncertainties.

2) A blind ALS algorithm for solving (49)–(50)*:* From [56, Theorem I.5], it is only possible to recover the $(\mathbf{c}_2)_r$ vectors from (49)–(50). Thus the design of a blind algorithm that ignores the degradation and uncertainties relationships between $\mathbf{C}_1$ and $\mathbf{C}_2$ can be envisioned. Such an algorithm was introduced in [56] in a remote sensing image fusion framework. This algorithm, denoted Block Term Decomposition -

accounting for Variability (BTD-Var), minimizes the following cost function:

$$\min_{\substack{\mathbf{A}_2,\mathbf{B}_2,\\ \mathbf{C}_1,\mathbf{C}_2}} \|\boldsymbol{\mathcal{Y}}_1 - \sum_{r=1}^{R} \left(\mathbf{P}(\mathbf{A}_2)_r(\mathbf{Q}(\mathbf{B}_2)_r)^T\right) \otimes (\mathbf{c}_1)_r\|_F^2 + \lambda\|\boldsymbol{\mathcal{Y}}_2 - \sum_{r=1}^{R} \left((\mathbf{A}_2)_r(\mathbf{B}_2)_r^T\right) \otimes (\mathbf{c}_2)_r\|_F^2, \quad (51)$$

which is the ML criterion for the blind (49)–(50) problem if $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$.

Since $\mathbf{R}$ and the random uncertainties are supposed to be unknown, the criterion (51) is partially-constrained; indeed it ignores any coupling constraint between $\mathbf{C}_1$ and $\mathbf{C}_2$. To be more precise, the LL1 factor $\mathbf{C}_2$ is still subject to some uncertainties in the blind problem, but it is considered unknown in (51). A general framework for BTD-Var is available in [56].

3) Performance bounds for the blind model: According to (49)–(50), the blind constraints on the models parameters are such that

$$\widetilde{\mathbf{g}}(\boldsymbol{\theta}) = \boldsymbol{\phi} - \underbrace{\begin{bmatrix} \mathbf{I} \boxtimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \boxtimes \mathbf{Q} \end{bmatrix} \mathbf{M}^T}_{\widetilde{\mathbf{G}}} \boldsymbol{\omega}, \quad (52)$$

which is a specific form for (47) in the case where $\mathbf{R}$ and $\boldsymbol{\Psi}$ are unknown. Nevertheless, a lower bound on parameter estimation for the blind problem still implicitly depends on unknown uncertainties.

Indeed, we can define the Blind-CCRB for the parameter $\boldsymbol{\theta}$, conditionally on the random parameter vector:

$$\mathbf{Blind\text{-}CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \widetilde{\mathbf{U}}(\boldsymbol{\theta}) \left(\widetilde{\mathbf{U}}^T(\boldsymbol{\theta})\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\widetilde{\mathbf{U}}(\boldsymbol{\theta})\right)^{-1} \widetilde{\mathbf{U}}^T(\boldsymbol{\theta}), \quad (53)$$

where $\widetilde{\mathbf{U}}^T(\boldsymbol{\theta}) = \left[\mathbf{I}\ \widetilde{\mathbf{G}}^T\right]$.

In (53), the uncertainties are solely contained in $\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \neq \mathbf{F}(\boldsymbol{\theta})$, hence the randomly-constrained bound must be considered as well to properly evaluate the performance of the blind algorithm. As a result, we define the so-called Blind-RCCRB as

$$\mathbf{Blind\text{-}RCCRB}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\mathbf{Blind\text{-}CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right]. \quad (54)$$

VI. Relative efficiency of the blind algorithm using the RCCRB

We are now ready to assess the relative efficiency of the blind algorithm BTD-Var using the appropriate randomly-constrained CRB.

A. Experiments setup

We consider the same dimensions as in Section IV. The random parameter vector $\boldsymbol{\theta}_r$ is drawn from a Gaussian distribution with unit mean and variance $\sigma_r^2 = 0.2$. The factors $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$ are generated as in Section IV while $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ are generated according to (45).

We consider different estimation scenarios and corresponding performance bounds. First, we consider the deterministic LL1-ALS algorithm, that minimizes the cost function (40). This algorithm incorrectly assumes that there is no variability phenomenon, hence it considers a false model. We compare the performance of this first algorithm to the standard CCRB.

Second, we consider the clairvoyant algorithm minimizing the cost function (46). We compare its performance to the fully-coupled RCCRB considered in Section V-D. This bound corresponds to the best performance achievable by the coupled model with uncertainties. However, since it requires the knowledge of the uncertainty phenomenon, it corresponds to a clairvoyant scenario.

Finally, we assess the performance of the blind algorithm BTD-Var, that considers the criterion (51). We compare the MSE obtained from BTD-Var to the Blind-RCCRB (54).

B. Results

In Figure 5, we plot on a semi-log scale the total bounds and UMSE for the parameter $\mathbf{y}$. For $SNR_1 \leq SNR_2$, uniform

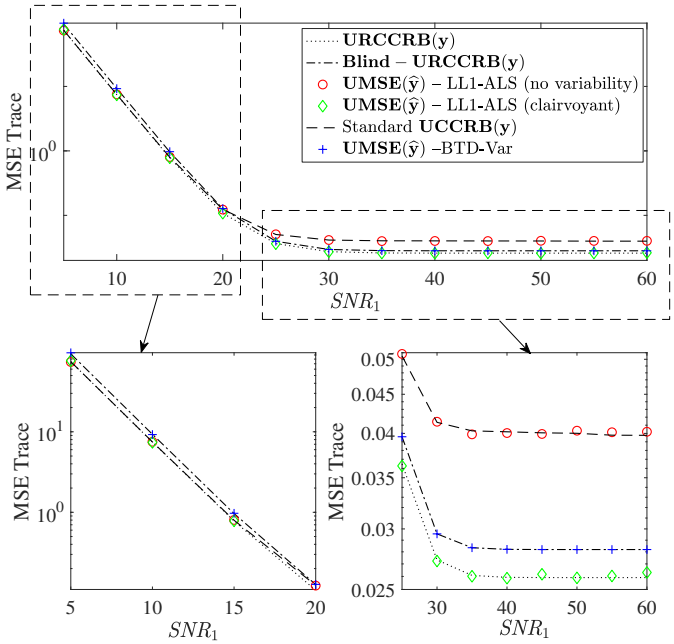

Fig. 5. Standard UCCRB, URCCRB and Blind-URCCRB, and UMSE for the parameter $\mathbf{y}$; close-ups for $SNR_1 \leq SNR_2$ (down, left) and $SNR_1 > SNR_2$ (down, right).

bounds and MSE traces are almost all equal to each other, and the MSE traces seem to reach the corresponding bounds.

For $SNR_1 > SNR_2$, different conclusions can be drawn. First, the deterministic LL1-ALS algorithm reaches the standard UCCRB. However, because this algorithm does not account for the variability phenomenon, it had the worst performance of the three estimators. In particular, some estimators even outperform the standard UCCRB, highlighting the limitations of the standard tool.

Second, the clairvoyant LL1-ALS algorithm reaches the URCCRB, thus it is uniformly asymptotically efficient for tensor reconstruction. However, it corresponds to a scenario

in which we would be able to estimate the variability phenomenon, which is often impossible in practice.

Finally, the Blind-URCCRB and UMSE obtained from BTD-Var yield slightly worse performance than in the fully-coupled case: this is reasonable since only a portion of the constraints is considered. However, the performance of BTD-Var is still better than that of the deterministic algorithm, and reaches the Blind-URCCRB. Additionally, performance analysis for BTD-Var allows to measure the loss of information induced by ignoring a portion of the constraints. Since the discrepancy between the URCCRB and Blind-URCCRB is small, it can be concluded from this experiment that considering $\mathbf{R}$ and the variability to be unknown is not critical, since it does not heavily degrade the performance.

C. Conclusion and perspectives

In this paper, we considered a coupled LL1 model accounting for uncertainties. We showed that, for such a model, it is reasonable to consider a blind (*i.e.* partially-coupled) estimator that implicitly accounts for the variability phenomenon at hand. Using the RCCRB, we demonstrated the efficiency of this new estimator, and showed that it yielded better performance than less flexible algorithms ignoring uncertainties, without the need for knowing the variability phenomenon or attached degradation. Moreover, this new algorithm only has slightly lower performance compared to the fully-coupled algorithm; therefore it is robust to the knowledge of degradation in the third mode.

Our performance analysis also indicates that it is possible to seek for an efficient estimator, similar to the clairvoyant one; to achieve the best performance, this algorithm must be able to estimate the variability phenomenon. This perspective is of great interest and is likely to be explored in future works.

APPENDIX A
LL1-BASED STANDARD FIM AND CCRB

In (32), the Fisher information matrix for $\widetilde{\boldsymbol{\theta}}$ can be viewed as a symmetric block-matrix of the form

$$\mathbf{F}(\widetilde{\boldsymbol{\theta}}) = \begin{bmatrix} \mathbf{D}_{\widetilde{\boldsymbol{\omega}},\widetilde{\boldsymbol{\omega}}} & \mathbf{D}_{\widetilde{\boldsymbol{\omega}},\boldsymbol{\phi}} \\ \mathbf{D}_{\widetilde{\boldsymbol{\omega}},\boldsymbol{\phi}}^T & \mathbf{D}_{\boldsymbol{\phi},\boldsymbol{\phi}} \end{bmatrix}. \tag{55}$$

Given (29)–(31), the subblocks in (55) are such as

$$\mathbf{D}_{\widetilde{\boldsymbol{\omega}},\widetilde{\boldsymbol{\omega}}} = \begin{bmatrix} \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^T\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^T\mathbf{S}_{B_2} & \mathbf{0} \\ \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^T\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^T\mathbf{S}_{B_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^T\mathbf{S}_{C_1} \end{bmatrix}, \tag{56}$$

$$\mathbf{D}_{\widetilde{\boldsymbol{\omega}},\boldsymbol{\phi}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^T\mathbf{S}_{C_2} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^T\mathbf{S}_{C_2} \\ \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^T\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^T\mathbf{S}_{B_1} & \mathbf{0} \end{bmatrix}, \tag{57}$$

$$\mathbf{D}_{\boldsymbol{\phi},\boldsymbol{\phi}} = \begin{bmatrix} \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^T\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^T\mathbf{S}_{B_1} & \mathbf{0} \\ \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^T\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^T\mathbf{S}_{B_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^T\mathbf{S}_{C_2} \end{bmatrix}. \tag{58}$$

Finally, the FIM for the reduced parameter $\boldsymbol{\theta}$ can be obtained as in (33).

Similarly, the standard CCRB in (36) is a block-matrix of the form

$$\mathbf{CCRB}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{CCRB}_{\widetilde{\boldsymbol{\omega}},\widetilde{\boldsymbol{\omega}}} & \mathbf{CCRB}_{\widetilde{\boldsymbol{\omega}},\boldsymbol{\phi}} \\ \mathbf{CCRB}_{\widetilde{\boldsymbol{\omega}},\boldsymbol{\phi}}^T & \mathbf{CCRB}_{\boldsymbol{\phi},\boldsymbol{\phi}} \end{bmatrix}. \tag{59}$$

Denote $\mathbf{CCRB}(\boldsymbol{\omega})$ and $\mathbf{CCRB}(\boldsymbol{\phi})$ the diagonal blocks in (59). Developing (36) using (55) yields

$$\mathbf{CCRB}(\boldsymbol{\omega}) = \left(\mathbf{U}(\boldsymbol{\theta})^T\mathbf{F}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})\right)^{-1}, \tag{60}$$

$$\mathbf{CCRB}(\boldsymbol{\phi}) = \mathbf{G}\left(\mathbf{U}(\boldsymbol{\theta})^T\mathbf{F}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})\right)^{-1}\mathbf{G}^T, \tag{61}$$

where

$$\begin{aligned}\mathbf{U}(\boldsymbol{\theta})^T\mathbf{F}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta}) = {}& \mathbf{M}^T\mathbf{D}_{\widetilde{\boldsymbol{\omega}},\widetilde{\boldsymbol{\omega}}}\mathbf{M} + \mathbf{G}^T\mathbf{D}_{\widetilde{\boldsymbol{\omega}},\boldsymbol{\phi}}^T\mathbf{M} \\ & + \mathbf{M}^T\mathbf{D}_{\widetilde{\boldsymbol{\omega}},\boldsymbol{\phi}}\mathbf{G} + \mathbf{G}^T\mathbf{D}_{\boldsymbol{\phi},\boldsymbol{\phi}}\mathbf{G}.\end{aligned} \tag{62}$$

APPENDIX B
DEGRADATION MATRICES

Here, we explain in details how the degradation matrices are constructed. As in [30], $\mathbf{P}$ is constructed as $\mathbf{P} = \mathbf{S}_1\mathbf{T}_1$, where $\mathbf{T}_1$ is a blurring Toeplitz matrix and $\mathbf{S}_1$ is a downsampling matrix.

The blurring matrix is constructed from a Gaussian blurring kernel $\boldsymbol{\phi} \in \mathbb{R}^{q\times 1}$ with a standard deviation $\sigma = \frac{q\sqrt{2\log(2)}}{4}$. For $m \in \{1,\dots,q\}$ and $m' = m - \lceil \frac{q}{2} \rceil$, we have

$$\phi(m) = \exp\left(\frac{-m'^2}{2\sigma^2}\right).$$

Thus, $\mathbf{T}_1 \in \mathbb{R}^{I\times I}$ can be seen as

$$\mathbf{T}_1 = \begin{bmatrix} \phi(\lceil \frac{q}{2} \rceil) & \dots & \phi(q) & 0 & \dots & 0 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ \phi(1) & & \ddots & & \ddots & 0 \\ 0 & \ddots & & \ddots & & \phi(q) \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ 0 & \dots & 0 & \phi(1) & \dots & \phi(\lceil \frac{q}{2} \rceil) \end{bmatrix}.$$

The downsampling matrix $\mathbf{S}_1 \in \mathbb{R}^{I_H\times I}$, with downsampling ratio d, is made of I_H independent rows such that for $i \in \{1,\dots,I_H\}$, $(\mathbf{S}_1)_{i,2+(i-1)d} = 1$ and the other coefficients are zeros.

The degradation matrix $\mathbf{R}$ is computed from the Sentinel-2 spectral response functions. In a remote sensing framework, this matrix selects the common spectral bands of the tensors $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Y}}_2$. To be more precise, we select a non-zero portion of the 3rd to 8th spectral channels, that correspond to the wavelengths 543–577nm, 650–680nm, 698–712nm, 733–747nm, 773–793nm and 785–900nm.

REFERENCES

[1] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Part 1*, New York, Wiley, 1968

[2] N. Kbayer, J. Galy, E. Chaumette, F. Vincent, A. Renaux and P. Larzabal, "On Lower Bounds for Non-Standard Deterministic Estimation", *IEEE Trans. on SP*, 65(6): 1538–1553, 2017

[3] M. Fréchet, "Sur l'extension de certaines évaluations statistiques au cas de petits échantillons", *Rev. Int. Stat*, 11: 182–205, 1943

[4] G. Darmois, "Sur les lois limites de la dispersion de certaines estimations", *Rev. Int. Stat*, 13: 9–15, 1945

[5] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946

[6] C.R. Rao, "Information and accuracy attainable in the estimation of statistical parameters", *Bull. Calcutta Math. Soc.*, 37: 81–91, 1945

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

[8] H.L. Van Trees and K. L. Bell, Eds., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, Wiley/IEEE Press, 2007

[9] R. McAulay and L.P. Seidman, "A useful form of the Barankin lower bound and its application to PPM threshold analysis", *IEEE Trans. on IT*, 15(2): 273–279, 1969

[10] F.E. Glave, "A new look at the Barankin Lower Bound", *IEEE Trans. on IT*, 18(3): 349–356, 1972

[11] J.S. Abel, "A bound on mean-square-estimate error", *IEEE Trans. on IT*, 39(5): 1675–1680, 1993

[12] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, "A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of MLEs", *IEEE Trans. on SP*, 56(11): 5319–5333, 2008

[13] K. Todros and J. Tabrikian, "General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators", *IEEE Trans. on IT*, 56(10): 5064–5082, 2010

[14] J.D. Gorman and A.O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. on IT*, 26(6): 1285–1301, 1990

[15] T.L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér–Rao bound", *IEEE Trans. on SP*, 41(6): 2247–2249, 1993

[16] P. Stoica and T.L. Marzetta, "Parameter estimation problems with singular information matrices", *IEEE Trans. on SP*, 49(1): 87–90, 2001

[17] P. Stoica and B.C. Ng, " On the Cramér–Rao bound under parametric constraints", *IEEE SP Letters*, 5(7): 177–179, 1998

[18] A.K. Jagannatham, B.D. Rao, "Cramér-Rao Lower Bound for Constrained Complex Parameters", *IEEE SP letters*, 11(11), 2004

[19] T.J. Moore, B.M. Sadler, R.J. Kozick, "Maximum-Likelihood Estimation, the Cramér-Rao Bound, and the Method of Scoring With Parameter Constraints", *IEEE Trans. on SP*, 56(3): 895–908, 2008

[20] T. Menni, E. Chaumette, P. Larzabal and J.P. Barbot, "New results on Deterministic Cramér-Rao bounds for real and complex parameters", *IEEE Trans. on SP*, 60(3): 1032–1049, 2012

[21] T. Menni, J. Galy, E. Chaumette, P. Larzabal, "Versatility of Constrained CRB for System Analysis and Design", *IEEE Trans. on AES*, 50(3): 1841–1863, 2014

[22] R.W. Miller and C.B. Chang, "A modified Cramér-Rao bound and its applications," *IEEE Trans. on IT*, 24(3): 398–400, 1978

[23] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations-part I: Far-field sources," *IEEE Trans. on ASSP*, 35(3): 286–299, 1987

[24] A.N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramér-Rao bound and its application to synchronization problems," *IEEE Trans. on Commun.*, 42(2/3/4): 1391–1399, 1994

[25] R.A. Horn, C.R. Johnson, *Matrix Analysis (2nd Ed)*. Cambridge University Press, 2013.

[26] S. Sahnoun and P. Comon, "Joint source estimation and localization", *IEEE Trans. SP*, 63(10): 2485–2595, 2015.

[27] X. Liu and N.D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensionnal arrays", *IEEE Trans. SP*, 49(9): 2074–2086, 2001.

[28] M. Boizard, R. Boyer, G. Favier, J.E. Cohen, and P. Comon, "Performance estimation for tensor CP decomposition with structured factors", in *Proc. ICASSP, 2015*.

[29] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion", *IEEE Trans. Geosci. Remote Sens.*, 50(2): 528–537, 2012.

[30] C.I. Kanatsoulis, X. Fu, N.D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach", *IEEE Trans. SP*, 66(24): 6503–6517, 2018.

[31] C. Ren, R. Cabral Farias, P-O. Amblard, and P. Comon, "Performance bounds for coupled models", in *Proc. 2016 IEEE SAM*, 2016.

[32] C. Prevost, K. Usevich, M. Haardt, P. Comon, and D. Brie, "Performance bounds for coupled CP model in the framework of hyperspectral super-resolution", in *Proc. CAMSAP, 2019*, available online at https://hal.archives-ouvertes.fr/hal-02303132.

[33] C. Prevost, K. Usevich, M. Haardt, P. Comon, and D. Brie, "Constrained Cramér-Rao lower bounds for CP-based hyperspectral super-resolution", available online at https://hal.archives-ouvertes.fr/hal-03083709v1.

[34] D. Manolakis, D. Marden and G.A. Shawn, "Hyperspectral image processing for automatic target detection applications", *Lincoln Laboratory journal*, 14(1): 79–116, 2003.

[35] A. Camacho, C.V. Correa and H. Arguello, "An analysis of spectral variability in hyperspectral imagery: a case study of stressed oil palm detection in Colombia", *International Journal of Remote Sens.*, 1–21, 2019.

[36] P. Comon, 'Tensors: A brief introduction", *IEEE Signal Process. Mag.*, 31(3): 44–53, 2014.

[37] T.G. Kolda and B.W. Bader, "Tensor decompositions and applications", *SIAM Review*, 51(3): 455–500, 2009.

[38] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images", *Photogrammetric Eng. and Remote Sens.*, 63(6): 691–699, 1997.

[39] D. Slepian, "Estimation of signal parameters in the presence of noise", *Trans. IRE Professional Group Inf. Theory*, 3(3): 68–69, 1954.

[40] M. Ding, X. Fu, K. Huang and J. Wang, "Hyperspectral Super-Resolution via Interpretable Block-Term Tensor Modeling", 2020, available at arXiv:2006.10248.

[41] G. Zhang, K. Usevich, M. Haardt, P. Comon, and D. Brie, "Hyperspectral Super-Resolution: A Coupled Nonnegative Block-Term Tensor Decomposition Approach", in *Proc. IEEE CAMSAP, 2019*, 470–474.

[42] C. Prévost, K. Usevich, E. Chaumette, D. Brie and P. Comon, "On Cramér-Rao lower bounds with random equality constraints", in *Proc. IEEE ICASSP, 2020*, 5355–5359.

[43] D.R. Hearn, "Characterization of instrument spectral resolution by the spectral modulation transfer function", in *Earth Observing Systems III*, 3439: 400–407, 1998, International Society for Optics and Photonics.

[44] P. Gege, J. Fries, P. Haschberger, P. Schötz, H. Schwarzer, P. Strobl, B. Suhr, G. Ulbrich and W.J. Vreeling, "Calibration facility for airborne imaging spectrometers", in *ISPRS J. Photogramm. Remote Sens.*, 64(4): 387–397, 2009, Elsevier.

[45] W.R. Tahnk and J.A. Coakley Jr., "Updated calibration coefficients for NOAA-14 AVHRR channels 1 and 2", in *International Journal of Remote Sens.*, 22(15): 3053–3057, 2001, Taylor & Francis.

[46] G. Gutman, A. Gruber, D. Tarpley and R. Taylor, "Application of angular models to AVHRR data for determination of the clear-sky planetary albedo over land surfaces", in *Journal of Geophysical Research: Atmospheres*, 94(D7): 9959-9970, 1989.

[47] N. Vervliet, O. Debals, L. Sorber, M. Van Barel and L. De Lathauwer, "Tensorlab 3.0", *available online, URL: www.tensorlab.net*, 2016.

[48] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness", in *SIAM J. Matrix Anal. Appl*, 30(3): 1033–1066, 2008.

[49] L. Guanter, R. Ritcher and, J. Moreno "Spectral calibration of hyperspectral imagery using atmospheric absorption features", in *Applied optics*, 45(10): 2360–2370, 2006, Optical Society of America.

[50] T. Hilker, M.A. Wulder, N.C. Coops, J. Linke, G. McDermid, J.G. Masek, F. Gao, and J.C. White, "A new data fusion model for high spatial and temporal-resolution mapping of forest disturbance based on LANDSAT and MODIS," in *Remote Sensing of Environment*, 113(8): 1613–1627, 2009.

[51] I.V. Emelyanova, T.R. McVicar, T.G. Van Niel, L.T. Li, and A.M. Van Dijk, "Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," in *Remote Sensing of Environment*, 133: 193–209, 2013.

[52] R. A. Borsoi, T. Imbiriba, J.M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, C. Zare, and C. Jutten, "Spectral variability in hyperspectral data unmixing: A comprehensive review," in *IEEE Geoscience and Remote Sensing Magazine*, 2021.

[53] J. Munkres. "Algorithms for the assignment and transportation problems." Journal of the society for industrial and applied mathematics, 5(1): 32–38.

[54] C.I. Kanatsoulis, X.Fu, N.D. Sidiropoulos, and M. Akçakaya, "Tensor completion from regular sub-Nyquist samples." *IEEE Trans. SP*, 68: 1–16, 2019.

[55] G. Zhang, X. Fu, J. Wang, X.L. Zhao and M. Hong, "Spectrum cartography via coupled block-term tensor decomposition." *IEEE Trans. SP*, 68: 3660–3675, 2020.

[56] C. Prévost, R.A. Borsoi, K. Usevich, D. Brie, J.C. Bermudez, and C. Richard, "Hyperspectral super-resolution accounting for spectral variability: LL1-based recovery and blind unmixing" (2021). available online at https://hal.archives-ouvertes.fr/hal-03158076.

[57] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms—Part III: Alternating least squares algorithms." *SIAM J. Matrix Anal. Appl.*, 30(3): 1067–1083, 2008.

[58] B. Somers, G.P. Asner, L. Tits and P.Coppin, "Endmember variability in spectral mixture analysis: A review." *Remote Sensing of Environment*, 115(7): 1603–1616, 2011.

[59] R.A. Borsoi, C. Prévost, K. Usevich, D. Brie, J.C. Bermudez, and C. Richard, "Coupled Tensor Decomposition for Hyperspectral and Multispectral Image Fusion with Inter-image Variability." *IEEE J. Sel. Topics SP*, 2021.

[60] Y.C. Eldar "Minimum variance in biased estimation: Bounds and asymptotically optimal estimators", *IEEE Trans. on SP,* 52(7): 1915–1930, 2004.

[61] A.O. Hero, "A Cramér-Rao type lower bound for essentially unbiased parameter estimation", *Lincoln laboratory journal*, 1992.

[62] A.O. Hero, J.A. Fessler and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound." *IEEE Trans. SP*, 44(8): 2026–2041, 1996.

[63] I.H. Godtliebsen, B. Thomsen, and O. Christiansen, "Tensor decomposition and vibrational coupled cluster theory." *The Journal of Physical Chemistry* 117(32): 7267–7279, 2013.

[64] C.I. Kanatsoulis and N.D. Sidiropoulos, "TeX-Graph: Coupled tensor-matrix knowledge-graph embedding for COVID-19 drug repurposing." *Proc. SIAM SDM*, 2021.

[65] F. Xiong, Y. Qian, J. Zhou, and Y.Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization." *IEEE Trans. Geosci. Remote Sens.*, 57(4): 2341–2357, 2018.