

Performances optimales pour les données multimodales partiellement couplées

Chengfang Ren, Rodrigo Cabral Farias, Pierre-Olivier Amblard, Pierre Comon

► **To cite this version:**

Chengfang Ren, Rodrigo Cabral Farias, Pierre-Olivier Amblard, Pierre Comon. Performances optimales pour les données multimodales partiellement couplées. GRETSI 2017 - XXVIème Colloque francophone de traitement du signal et des images, Sep 2017, Juan-Les-Pins, France. hal-01567124

HAL Id: hal-01567124

<https://hal.archives-ouvertes.fr/hal-01567124>

Submitted on 21 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Performances optimales pour les données multimodales partiellement couplées

Chengfang REN¹, Rodrigo CABRAL FARIAS², Pierre-Olivier AMBLARD³ et Pierre COMON^{3*}

¹SONDRA, CentraleSupélec, 91192 Gif-sur-Yvette, France

²I3S, UNS CNRS, 06900 Sophia Antipolis, France

³CNRS, GIPSA-Lab, Univ. Grenoble Alpes, F-38000 Grenoble, France

chengfang.ren@centralesupelec.fr, cabral@i3s.unice.fr,
prenom.nom@gipsa-lab.grenoble-inp.fr

Résumé – Deux modèles d’observation sont dits “couplés” si une partie non vide des paramètres d’intérêt est reliée par une équation implicite différentiable. Dans ce contexte, plusieurs stratégies d’estimation sont possibles: soit un processus d’estimation jointe dans lequel les données issues de ces deux modèles sont fusionnées, soit un processus d’estimation individuelle où les paramètres d’intérêt sont estimés à partir de leur modèle d’observation respectif. Dans ce papier, nous montrons à travers une analyse de bornes que l’estimation optimale jointe est plus précise qu’une estimation individuelle en termes d’erreur quadratique moyenne. Cette propriété est également vraie pour les paramètres qui ne sont pas directement reliés. Enfin, ce résultat est illustré à travers un problème de décompositions tensorielles couplées.

Abstract – Two models are called “coupled” when a non empty set of the underlying parameters of interest are related through a differentiable implicit function. In this context, several estimation strategies can be derived: either a joint estimation where the parameters of interest are retrieved by merging all datasets, or an individual estimation where the parameters of interest are respectively estimated from each dataset. In this paper, we show by analyzing lower bounds that the optimal process of joint estimation is always more accurate than the individual estimation process in the sense of mean square error for a general class of dataset distributions. This property is still true for the parameters that are not directly linked. Finally, we illustrate our results with the fusion of multiple tensor data.

1 Introduction

L’analyse de données multimodales suscite un intérêt croissant dans de nombreuses applications [1–3]. Elle consiste à estimer des paramètres d’intérêt à partir de données acquises par des capteurs multiples [4], parfois de nature différente, ce qui conduit à proposer des modèles paramétriques différents pour chaque type de capteurs. Néanmoins, certains paramètres d’intérêt entre ces modèles peuvent être reliés car c’est le même phénomène physique qui est observé. Dans ce papier, nous nous intéressons à un modèle de couplage partiel potentiellement non linéaire défini dans la section 2. Dans la section 3, nous proposons une borne de Cramér-Rao sous contraintes (BCRC) pour l’estimation simultanée des paramètres couplés et non couplés en nous basant sur les travaux de [5–7]. Cette borne établit une jauge de performances optimales en termes d’erreur quadratique moyenne (EQM). Nous nous intéressons en particulier au comportement de l’EQM optimale dans les cas où les paramètres d’intérêt sont estimés conjointement à partir des données fusionnées ou non. Pour illustrer la validité de la borne proposée, nous l’appliquons aux problèmes de décompositions tensorielles polyadiques canoniques (PC) complexes et

couplées dans la section 4. La décomposition PC est utile dans de nombreux domaines, par exemple en traitement d’antennes, en traitement radar et en télécommunications [8–11]. Pour les décompositions non couplées, les études de performances optimales sont faites dans [8, 12, 13]. Pour les tenseurs hétérogènes réels linéairement couplés, les bornes de performances ont déjà été proposées dans un contexte d’estimation bayésienne [14]. Notre contribution étend les précédents travaux, puisque les résultats proposés ici restent valables pour des modèles d’observation généraux et des couplages non linéaires.

2 Modèles hétérogènes couplés

Soit \mathbf{x}_1 et \mathbf{x}_2 deux ensembles de données hétérogènes couplés de manière suivante :

$$\begin{cases} \mathbf{x}_1 \sim f_{\mathbf{x}_1; \boldsymbol{\theta}_1, \boldsymbol{\phi}_1} \text{ and } \mathbf{x}_2 \sim f_{\mathbf{x}_2; \boldsymbol{\theta}_2, \boldsymbol{\phi}_2}, \\ \mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}, \end{cases} \quad (1)$$

où $\mathbf{x}_1 \in \Omega_1 \subseteq \mathbb{C}^{n_1}$ et $\mathbf{x}_2 \in \Omega_2 \subseteq \mathbb{C}^{n_2}$ sont des données distribuées respectivement suivant les densités de probabilité (ddp) $f_{\mathbf{x}_1; \boldsymbol{\theta}_1, \boldsymbol{\phi}_1}$ et $f_{\mathbf{x}_2; \boldsymbol{\theta}_2, \boldsymbol{\phi}_2}$, paramétrées respectivement par les vecteurs déterministes inconnus $(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1) \in \Theta_1 \times \Phi_1 \subseteq \mathbb{R}^{m_1} \times \mathbb{R}^{p_1}$ et $(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2) \in \Theta_2 \times \Phi_2 \subseteq \mathbb{R}^{m_2} \times \mathbb{R}^{p_2}$, n_i , m_i et p_i désignant respectivement la taille des vecteurs \mathbf{x}_i , $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$. Ce modèle est dit *partiellement couplé* car seuls les vecteurs $\boldsymbol{\theta}_1$ et $\boldsymbol{\theta}_2$ sont

*Ce projet est financé par le projet ERC "DECODA" no. 320594 dans le cadre du programme européen FP7/2007-2013.

couplés par la contrainte $\mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}$ tandis qu'il n'y a pas de dépendance entre les vecteurs $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_2$. Nous ferons les hypothèses que (i) $\mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ est une fonction vectorielle déterministe connue, *non-redondante*¹ et différentiable par rapport à $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta_1 \times \Theta_2$, (ii) la loi $f_{\mathbf{x}_{1:2}; \boldsymbol{\theta}_{1:2}, \boldsymbol{\phi}_{1:2}}$ est différentiable par rapport² à $\boldsymbol{\theta}_{1:2}$ et $\boldsymbol{\phi}_{1:2}$ et son support ne dépend pas de ses paramètres, (iii) \mathbf{x}_1 et \mathbf{x}_2 sont statistiquement indépendants, *c.-à-d.* $f_{\mathbf{x}_{1:2}; \boldsymbol{\theta}_{1:2}, \boldsymbol{\phi}_{1:2}} = f_{\mathbf{x}_1; \boldsymbol{\theta}_1, \boldsymbol{\phi}_1} f_{\mathbf{x}_2; \boldsymbol{\theta}_2, \boldsymbol{\phi}_2}$.

De plus, les ddps de \mathbf{x}_1 et \mathbf{x}_2 peuvent être totalement différentes bien que les paramètres d'intérêt sont les mêmes *c.-à-d.* $f_{\mathbf{x}_1; \boldsymbol{\theta}, \boldsymbol{\phi}} \neq f_{\mathbf{x}_2; \boldsymbol{\theta}, \boldsymbol{\phi}}$. Ceci est un modèle pour la fusion de données hétérogènes.

3 Performances optimales

Pour simplifier les notations, nous omettons la dépendance par rapport aux vraies valeurs des paramètres $\mathbf{F} \triangleq \mathbf{F}(\boldsymbol{\theta}_{1:2}, \boldsymbol{\phi}_{1:2})$, $\mathbf{G} \triangleq \mathbf{G}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ et $\mathbf{BCRC} \triangleq \mathbf{BCRC}(\boldsymbol{\theta}_{1:2}, \boldsymbol{\phi}_{1:2})$. Lorsque le modèle partiellement couplé (1) et le modèle équivalent non couplé (*c.-à-d.* le même modèle sans la contrainte $\mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}$) sont tous deux identifiables alors la BCRC est donnée par [5]

$$\mathbf{BCRC} = \mathbf{F}^{-1} - \mathbf{F}^{-1} \mathbf{G}^\top (\mathbf{G} \mathbf{F} \mathbf{G}^\top)^{-1} \mathbf{G} \mathbf{F}^{-1}, \quad (2)$$

où $\mathbf{G}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \left[\frac{\partial \mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1^\top} \mathbf{0} \quad \mathbf{0} \quad \frac{\partial \mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2^\top} \right]$ est le gradient de \mathbf{g} par rapport à $(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1, \boldsymbol{\theta}_2, \boldsymbol{\phi}_2)$ et \mathbf{F} est la matrice d'information de Fisher (FIM) pour le modèle non couplé. Or, $\mathbf{F}^{-1} = \mathbf{BCR}$, donc le gain sur la précision d'estimation est :

$$\mathbf{BCR} - \mathbf{BCRC} = \mathbf{F}^{-1} \mathbf{G}^\top (\mathbf{G} \mathbf{F}^{-1} \mathbf{G}^\top)^{-1} \mathbf{G} \mathbf{F}^{-1} \succeq \mathbf{0}, \quad (3)$$

où $\mathbf{A} \succeq \mathbf{B}$ signifie que $\mathbf{A} - \mathbf{B}$ est une matrice semi-définie positive (SDP). Cette inégalité (3) montre que la BCRC est toujours plus basse que la BCR. En effet, les modèles couplés ont en général des informations supplémentaires, qui réduisent l'espace de recherche des paramètres et donc améliorent la précision d'estimation. Dans notre cas, tous les paramètres d'intérêt ne sont pas soumis à des contraintes mais seulement une partie. Donc, nous allons étudier de plus près l'effet de ces contraintes partielles sur la précision d'estimation.

3.1 Influence du couplage partiel sur la précision d'estimation

Comme $f_{\mathbf{x}_{1:2}; \boldsymbol{\theta}_{1:2}, \boldsymbol{\phi}_{1:2}} = f_{\mathbf{x}_1; \boldsymbol{\theta}_1, \boldsymbol{\phi}_1} f_{\mathbf{x}_2; \boldsymbol{\theta}_2, \boldsymbol{\phi}_2}$, les couples de variables $(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1)$ et $(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2)$ ne sont pas reliés pour le modèle non couplé. Ainsi la FIM classique est bloc diagonale $\mathbf{F} = \text{Diag}\{\mathbf{F}_1, \mathbf{F}_2\}$ avec

$$\mathbf{F}_i = -\mathbb{E} \begin{bmatrix} \frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\theta}_i} & \frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\phi}_i} \\ \frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\phi}_i} & \frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\theta}_i} \end{bmatrix}, \text{ pour } i = 1, 2, \quad (4)$$

où $\mathcal{L}_i = \ln f_{\mathbf{x}_i; \boldsymbol{\theta}_i, \boldsymbol{\phi}_i}(\mathbf{x}_i)$ est la log vraisemblance pour \mathbf{x}_i , et $\frac{\partial^2}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top} \triangleq \frac{\partial^2}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top}$ la dérivée seconde par rapport à $\boldsymbol{\theta}_i$ et $\boldsymbol{\theta}_i^\top$. Une

1. C'est à dire que les équations lignes $g_i = 0$ ne sont pas une combinaison linéaire des autres lignes $\{g_k = 0, k \neq i\}$ pour tout $\boldsymbol{\theta}_{1:2} \in \Theta_1 \times \Theta_2$.

2. $\mathbf{x}_{1:2}$, $\boldsymbol{\theta}_{1:2}$ et $\boldsymbol{\phi}_{1:2}$ sont respectivement la concaténation des vecteurs \mathbf{x}_1 et \mathbf{x}_2 , $\boldsymbol{\theta}_1$ et $\boldsymbol{\theta}_2$, $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_2$.

inversion par bloc permet d'écrire la matrice inverse sous la forme : $\mathbf{F}^{-1} = \text{Diag}\{\mathbf{BCR}_{\boldsymbol{\theta}_1 \boldsymbol{\phi}_1}, \mathbf{BCR}_{\boldsymbol{\theta}_2 \boldsymbol{\phi}_2}\}$. Notons $\mathbf{BCR}_{\boldsymbol{\theta}_i}$ et $\mathbf{BCR}_{\boldsymbol{\phi}_i}$ les blocs diagonaux de la matrice $\mathbf{BCR}_{\boldsymbol{\theta}_i \boldsymbol{\phi}_i}$. Alors :

$$\mathbf{BCR}_{\boldsymbol{\theta}_i} = (\mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\theta}_i} - \mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\phi}_i} \mathbf{D}_{\boldsymbol{\phi}_i \boldsymbol{\phi}_i}^{-1} \mathbf{D}_{\boldsymbol{\phi}_i \boldsymbol{\theta}_i}^\top)^{-1}, \quad (5)$$

$$\mathbf{BCR}_{\boldsymbol{\phi}_i} = (\mathbf{D}_{\boldsymbol{\phi}_i \boldsymbol{\phi}_i} - \mathbf{D}_{\boldsymbol{\phi}_i \boldsymbol{\theta}_i}^\top \mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\theta}_i}^{-1} \mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\phi}_i})^{-1}, \quad (6)$$

avec $\mathbf{D}_{\boldsymbol{\alpha}_i \boldsymbol{\beta}_i} = -\mathbb{E} [\frac{\partial^2 \mathcal{L}_i}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\beta}_i}]$.

L'évaluation de la BCRC permet de quantifier le gain du couplage donné en (3). Notons $\mathbf{K}_{\boldsymbol{\theta}_1}$, $\mathbf{K}_{\boldsymbol{\theta}_2}$, $\mathbf{K}_{\boldsymbol{\phi}_1}$ et $\mathbf{K}_{\boldsymbol{\phi}_2}$ les éléments diagonaux de (3) correspondant respectivement à la différence entre la BCR et la BCRC pour les paramètres $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_2$. En substituant les équations (5) et (6) dans (3), nous obtenons pour $i = 1, 2$

$$\mathbf{K}_{\boldsymbol{\theta}_i} = \mathbf{BCR}_{\boldsymbol{\theta}_i} \frac{\partial \mathbf{g}^\top}{\partial \boldsymbol{\theta}_i} \mathbf{BCR}_g^{-1} \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}_i} \mathbf{BCR}_{\boldsymbol{\theta}_i} \quad (7)$$

$$\mathbf{K}_{\boldsymbol{\phi}_i} = \mathbf{D}_{\boldsymbol{\phi}_i \boldsymbol{\phi}_i}^{-1} \mathbf{D}_{\boldsymbol{\phi}_i \boldsymbol{\theta}_i}^\top \mathbf{K}_{\boldsymbol{\theta}_i} \mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\theta}_i}^{-1} \mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\phi}_i}, \quad (8)$$

où $\mathbf{BCR}_g = \sum_{i=1}^2 \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}_i} \mathbf{BCR}_{\boldsymbol{\theta}_i} \frac{\partial \mathbf{g}^\top}{\partial \boldsymbol{\theta}_i}$. Il est facile de vérifier que les matrices $\mathbf{K}_{\boldsymbol{\theta}_i}$ et $\mathbf{K}_{\boldsymbol{\phi}_i}$ sont SDPs. Par conséquent, l'effet de couplage entre $\boldsymbol{\theta}_1$ et $\boldsymbol{\theta}_2$ améliore aussi la précision d'estimation des vecteurs $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_2$, malgré leur indépendance.

3.2 Cas gaussien couplé à moyenne paramétrée

Soient deux ensembles de données $\mathbf{x}_1 \sim \mathcal{CN}(\mathbf{m}_1(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1), \boldsymbol{\Sigma}_1)$ et $\mathbf{x}_2 \sim \mathcal{CN}(\mathbf{m}_2(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2), \boldsymbol{\Sigma}_2)$ où $\boldsymbol{\Sigma}_1$ et $\boldsymbol{\Sigma}_2$ sont connues et les paramètres $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_2$ inconnus, réels et non aléatoires. Si les fonctions $\mathbf{m}_1(\boldsymbol{\theta}_1, \boldsymbol{\phi}_1)$ et $\mathbf{m}_2(\boldsymbol{\theta}_2, \boldsymbol{\phi}_2)$ sont différentiables par rapport à $\boldsymbol{\theta}_1, \boldsymbol{\phi}_1$ et $\boldsymbol{\theta}_2, \boldsymbol{\phi}_2$, respectivement, alors on peut montrer que $\mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\theta}_i} = 2\Re\{\frac{\partial \mathbf{m}_i^\top}{\partial \boldsymbol{\theta}_i} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \mathbf{m}_i}{\partial \boldsymbol{\theta}_i}\}$, $\mathbf{D}_{\boldsymbol{\theta}_i \boldsymbol{\phi}_i} = 2\Re\{\frac{\partial \mathbf{m}_i^\top}{\partial \boldsymbol{\theta}_i} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \mathbf{m}_i}{\partial \boldsymbol{\phi}_i}\}$ et $\mathbf{D}_{\boldsymbol{\phi}_i \boldsymbol{\phi}_i} = 2\Re\{\frac{\partial \mathbf{m}_i^\top}{\partial \boldsymbol{\phi}_i} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \mathbf{m}_i}{\partial \boldsymbol{\phi}_i}\}$, où \Re est l'opérateur partie réelle. Par conséquent, les expressions analytiques de $\mathbf{K}_{\boldsymbol{\theta}_i}$ et de $\mathbf{K}_{\boldsymbol{\phi}_i}$ sont obtenues en substituant ces trois expressions dans les équations (7) et (8). Ces gains correspondent à la modification de la formule de Slepian-Bangs [15] pour prendre en compte le couplage $\mathbf{g}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}$.

4 Décomposition tensorielle couplée

Considérons deux tenseurs d'ordre trois, $\boldsymbol{\mathcal{X}}^{(1)} \in \mathbb{C}^{I_1 \times J_1 \times K_1}$ et $\boldsymbol{\mathcal{X}}^{(2)} \in \mathbb{C}^{I_2 \times J_2 \times K_2}$, modélisés par

$$\boldsymbol{\mathcal{X}}^{(1)} = \boldsymbol{\mathcal{T}}^{(1)} + \boldsymbol{\mathcal{N}}^{(1)} \quad \text{and} \quad \boldsymbol{\mathcal{X}}^{(2)} = \boldsymbol{\mathcal{T}}^{(2)} + \boldsymbol{\mathcal{N}}^{(2)}, \quad (9)$$

où $\boldsymbol{\mathcal{T}}^{(1)}$ et $\boldsymbol{\mathcal{T}}^{(2)}$ sont respectivement deux tenseurs de rang faible R_1 et R_2 et où $\boldsymbol{\mathcal{N}}^{(1)}$ et $\boldsymbol{\mathcal{N}}^{(2)}$ sont les tenseurs bruits dont les éléments sont supposés aléatoires, indépendants et identiquement distribués (i.i.d.) suivant une loi gaussienne complexe circulaire de moyenne nulle et de variance σ_1^2 et σ_2^2 , respectivement. On suppose également l'indépendance entre les éléments des deux tenseurs bruits. Soit la décomposition PC couplée de $\boldsymbol{\mathcal{T}}^{(i)}$ (voir [16]) définie par, pour $i = 1, 2$:

$$\boldsymbol{\mathcal{T}}^{(i)} = \sum_{r=1}^{R_i} \mathbf{a}_r^{(i)} \otimes \mathbf{b}_r^{(i)} \otimes \mathbf{c}_r^{(i)} \quad \text{sous} \quad \mathbf{g}(\mathbf{C}^{(1)}, \mathbf{C}^{(2)}) = \mathbf{0}, \quad (10)$$

où l'opérateur \otimes désigne le produit tensoriel, et les matrices $\mathbf{C}^{(i)} \triangleq [\mathbf{c}_1^{(i)} \ \mathbf{c}_2^{(i)} \ \dots \ \mathbf{c}_{R_i}^{(i)}] \in \mathbb{C}^{K_i \times R_i}$. Définissons également les facteurs matriciels $\mathbf{A}^{(i)} \triangleq [\mathbf{a}_1^{(i)} \ \dots \ \mathbf{a}_{R_i}^{(i)}] \in \mathbb{C}^{I_i \times R_i}$ et $\mathbf{B}^{(i)} \triangleq [\mathbf{b}_1^{(i)} \ \dots \ \mathbf{b}_{R_i}^{(i)}] \in \mathbb{C}^{J_i \times R_i}$. Puisque $\mathcal{N}^{(1)}$ et $\mathcal{N}^{(2)}$ sont i.i.d., chaque tenseur $\mathcal{X}^{(i)}$ est distribué selon la loi $f_{\mathcal{X}^{(i)}; \boldsymbol{\theta}_i, \phi_i} = (\pi \sigma_i)^{-I_i J_i K_i} \exp(-\frac{1}{\sigma_i^2} \|\mathcal{X}^{(i)} - \mathcal{T}^{(i)}\|^2)$, où $\|\cdot\|$ est la norme de Frobenius : $\|\mathcal{X}\|^2 \triangleq \sum_{i,j,k} |x_{ijk}|^2$.

4.1 Expression analytique de la BCRC

Un dépliage astucieux de $\mathcal{X}^{(i)}$ donne une expression compacte de la BCRC à l'aide des résultats de la section 3.2. Définissons les dépliages matriciels mode- p [16, 17], $\mathbf{T}_p^{(i)}$ et $\mathbf{X}_p^{(i)}$, de $\mathcal{T}^{(i)}$ et de $\mathcal{X}^{(i)}$ respectivement, $p \in \{1, 2, 3\}$. On a :

$$\begin{aligned} \mathbf{T}_1^{(i)} &= \mathbf{A}^{(i)}(\mathbf{C}^{(i)} \odot \mathbf{B}^{(i)})^\top, \mathbf{T}_2^{(i)} = \mathbf{B}^{(i)}(\mathbf{C}^{(i)} \odot \mathbf{A}^{(i)})^\top, \\ \mathbf{T}_3^{(i)} &= \mathbf{C}^{(i)}(\mathbf{B}^{(i)} \odot \mathbf{A}^{(i)})^\top, \end{aligned} \quad (11)$$

où l'opérateur \odot est le produit de Khatri-Rao. En utilisant les résultats de [18], le dépliage de la matrice $\mathbf{T}_3^{(i)}$ en vecteur s'exprime sous les trois formes suivantes :

$$\text{vec } \mathbf{T}_3^{(i)} = \mathbf{S}_C \text{vec } \mathbf{C}^{(i)} = \mathbf{S}_A \text{vec } \mathbf{A}^{(i)} = \mathbf{S}_B \text{vec } \mathbf{B}^{(i)} \quad (12)$$

où $\mathbf{S}_C = \mathbf{B}^{(i)} \odot \mathbf{A}^{(i)} \boxtimes \mathbf{I}_{K_i}$, $\mathbf{S}_A = \mathbf{J}_{13}^{(i)}(\mathbf{C}^{(i)} \odot \mathbf{B}^{(i)}) \boxtimes \mathbf{I}_{I_i}$, $\mathbf{S}_B = \mathbf{J}_{23}^{(i)}(\mathbf{C}^{(i)} \odot \mathbf{A}^{(i)}) \boxtimes \mathbf{I}_{J_i}$, $\mathbf{J}_{23}^{(i)}$ et $\mathbf{J}_{13}^{(i)}$ sont respectivement les matrices de permutation reliant les éléments du vecteur $\text{vec } \mathbf{X}_2^{(i)}$ aux ceux du vecteur $\text{vec } \mathbf{X}_3^{(i)}$ et les éléments de $\text{vec } \mathbf{X}_1^{(i)}$ aux ceux de $\text{vec } \mathbf{X}_3^{(i)}$, l'opérateur \boxtimes est le produit de Kronecker, et \mathbf{I}_α est la matrice d'identité de taille α .

Ainsi, le problème (9) colle au modèle couplé (1) avec $\mathbf{x}_i = \text{vec } \mathbf{X}_3^{(i)} \in \mathbb{C}^{I_i J_i K_i}$, $\boldsymbol{\theta}_i = [(\Re \Im)\{\text{vec}^\top \mathbf{C}^{(i)}\}]^\top \in \mathbb{R}^{2I_i R_i}$, et $\tilde{\boldsymbol{\phi}}_i = [(\Re \Im)\{\text{vec}^\top \mathbf{A}^{(i)}\} (\Re \Im)\{\text{vec}^\top \mathbf{B}^{(i)}\}]^\top \in \mathbb{R}^{2(J_i + K_i)R_i}$. Avec les notations précédentes, on a

$$\mathbf{x}_i \sim \mathcal{CN}(\mathbf{m}_i(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\phi}}_i), \sigma_i^2 \mathbf{I}), \quad (13)$$

où $\mathbf{m}_i(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\phi}}_i) = \text{vec } \mathbf{T}_3^{(i)}$. Cependant, la décomposition PC (10) n'est unique qu'à un facteur d'échelle près [16]. Pour lever cette indétermination, une solution est de fixer les éléments de la première ligne des matrices $\mathbf{A}^{(i)}$ et $\mathbf{B}^{(i)}$ à des valeurs prédéfinies. Cela signifie que certains éléments de $\tilde{\boldsymbol{\phi}}_i$ sont connus, et donc, nous devons définir un nouveau $\boldsymbol{\phi}_i \in \mathbb{R}^{2(J_i + K_i - 2)R_i}$ contenant seulement les paramètres inconnus du vecteur $\tilde{\boldsymbol{\phi}}_i$. La relation entre $\boldsymbol{\phi}_i$ et $\tilde{\boldsymbol{\phi}}_i$ est donnée par $\boldsymbol{\phi}_i = \mathbf{M} \tilde{\boldsymbol{\phi}}_i$ où $\mathbf{M} \in \mathbb{R}^{2(J_i + K_i - 2)R_i \times 2(J_i + K_i)R_i}$ est une matrice de masquage. Elle est obtenue en retirant les $4R_i$ lignes de la matrice d'identité $\mathbf{I}_{2(J_i + K_i)R_i}$ pour les valeurs prédéfinies de $\tilde{\boldsymbol{\phi}}_i$.

En utilisant (12), on obtient facilement les dérivées partielles

$$\frac{\partial}{\partial \boldsymbol{\theta}_i^\top} \mathbf{m}_i(\boldsymbol{\theta}_i, \boldsymbol{\phi}_i) = \mathbf{S}_C [\mathbf{I}_{K_i R_i} \ j \mathbf{I}_{K_i R_i}], \quad (14)$$

$$\frac{\partial}{\partial \boldsymbol{\phi}_i^\top} \mathbf{m}_i(\boldsymbol{\theta}_i, \boldsymbol{\phi}_i) = [\mathbf{S}_A \ \mathbf{S}_B] \mathbf{P} \mathbf{M}^\top, \quad (15)$$

avec $\mathbf{P} = \begin{bmatrix} \mathbf{I}_{I_i R_i} & j \mathbf{I}_{I_i R_i} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{J_i R_i} & j \mathbf{I}_{J_i R_i} \end{bmatrix}$ et j l'unité imaginaire pure. Donc, on obtient d'après la section (3.2)

$$\mathbf{D}_{\boldsymbol{\theta}_i, \boldsymbol{\theta}_i} = \frac{2}{\sigma_i^2} \Re \left\{ [\mathbf{I}_{K_i R_i} \ j \mathbf{I}_{K_i R_i}]^\text{H} \mathbf{S}_C^\text{H} \mathbf{S}_C [\mathbf{I}_{K_i R_i} \ j \mathbf{I}_{K_i R_i}] \right\} \quad (16)$$

$$\mathbf{D}_{\boldsymbol{\theta}_i, \boldsymbol{\phi}_i} = \frac{2}{\sigma_i^2} \Re \left\{ [\mathbf{I}_{K_i R_i} \ j \mathbf{I}_{K_i R_i}]^\text{H} \mathbf{S}_C^\text{H} [\mathbf{S}_A \ \mathbf{S}_B] \mathbf{P} \mathbf{M}^\top \right\} \quad (17)$$

$$\mathbf{D}_{\boldsymbol{\phi}_i, \boldsymbol{\phi}_i} = \frac{2}{\sigma_i^2} \Re \left\{ \mathbf{M} \mathbf{P}^\text{H} \begin{bmatrix} \mathbf{S}_A^\text{H} \mathbf{S}_A & \mathbf{S}_B^\text{H} \mathbf{S}_A \\ \mathbf{S}_A^\text{H} \mathbf{S}_B & \mathbf{S}_B^\text{H} \mathbf{S}_B \end{bmatrix} \mathbf{P} \mathbf{M}^\top \right\} \quad (18)$$

avec $\mathbf{S}_C^\text{H} \mathbf{S}_C = (\mathbf{B}^{(i)\text{H}} \mathbf{B}^{(i)} \boxtimes \mathbf{A}^{(i)\text{H}} \mathbf{A}^{(i)}) \boxtimes \mathbf{I}_{K_i}$, et où l'opérateur \boxtimes est le produit d'Hadamard. Finalement, la BCRC dans (2) est obtenue en remplaçant (16), (17), (18) dans (7) et (8).

4.2 Simulations

Dans cette partie, on choisit le modèle tensoriel explicité à la section précédente et le couplage est donnée par la fonction $g(\mathbf{C}^{(1)}, \mathbf{C}^{(2)}) = \mathbf{C}^{(1)} - \mathbf{C}^{(2)}$. L'estimation des facteurs matriciels PC consiste généralement à maximiser la fonction de vraisemblance $f_{\mathcal{X}^{(1)}, \mathcal{X}^{(2)}; \boldsymbol{\theta}_{1:2}, \phi_{1:2}}$. Cependant, ce critère dépend de nombreux paramètres et il n'est pas convexe malgré sa multilinéarité. Pour réduire le coût de calcul, une approche itérative a été proposée dans la littérature appelée l'algorithme des moindres carrés alternés (MCA) [19, pp. 62-63].

MCA non couplé La mise à jour à la k -ème itération du facteur $\mathbf{C}^{(i)}$ est une solution des moindres carrés [19, pp. 62-63] :

$$\hat{\mathbf{C}}_k^{(i)} = \mathbf{X}_3^{(i)} \left(\hat{\mathbf{B}}_k^{(i)} \odot \hat{\mathbf{A}}_k^{(i)} \right)^\star \left(\mathbf{M}_k^{(i)} \right)^{-1},$$

où \star est l'opérateur complexe conjugué et $\mathbf{M}_k^{(i)} = (\hat{\mathbf{B}}_k^{(i)\top} \hat{\mathbf{B}}_k^{(i)\star}) \boxtimes (\hat{\mathbf{A}}_k^{(i)\top} \hat{\mathbf{A}}_k^{(i)\star})$. Des expressions similaires peuvent être déduites pour les autres facteurs en se basant sur le dépliage (11).

MCA couplé La mise à jour pour les facteurs non couplés ne sont pas modifiés. Celle pour les facteurs couplés se fait conjointement en estimant la matrice \mathbf{C} en une seule fois :

$$\hat{\mathbf{C}}_k = \left[\sum_{i=1}^2 \frac{1}{\sigma_i^2}, \mathbf{X}_3^{(i)} \left(\hat{\mathbf{B}}_k^{(i)} \odot \hat{\mathbf{A}}_k^{(i)} \right)^\star \right] \left[\sum_{i=1}^2 \frac{1}{\sigma_i^2} \mathbf{M}_k^{(i)} \right]^{-1}.$$

Notons qu'il faut également éliminer les ambiguïtés d'échelle et de permutation inhérentes à la décomposition PC. Pour cela, si les premières lignes des matrices $\mathbf{A}^{(i)}$ et $\mathbf{B}^{(i)}$ sont non nulles, alors l'ambiguïté d'échelle est éliminée en normalisant $\hat{a}_{1r}^{(i)} = 1$ et $\hat{b}_{1r}^{(i)} = 1$ (sinon, on peut normaliser par rapport à des éléments non nuls de $\mathbf{A}^{(i)}$ et de $\mathbf{B}^{(i)}$). L'ambiguïté de permutation est corrigée en cherchant les vecteurs colonnes de $\hat{\mathbf{C}}^{(2)}$ se rapprochant le plus des vecteurs colonnes de $\hat{\mathbf{C}}^{(1)}$.

Scénario de simulation les tenseurs $\mathcal{T}^{(1)}$ et $\mathcal{T}^{(2)}$ ont des dimensions toutes égales à 15 et les rangs $R^{(1)} = R^{(2)} = 3$.

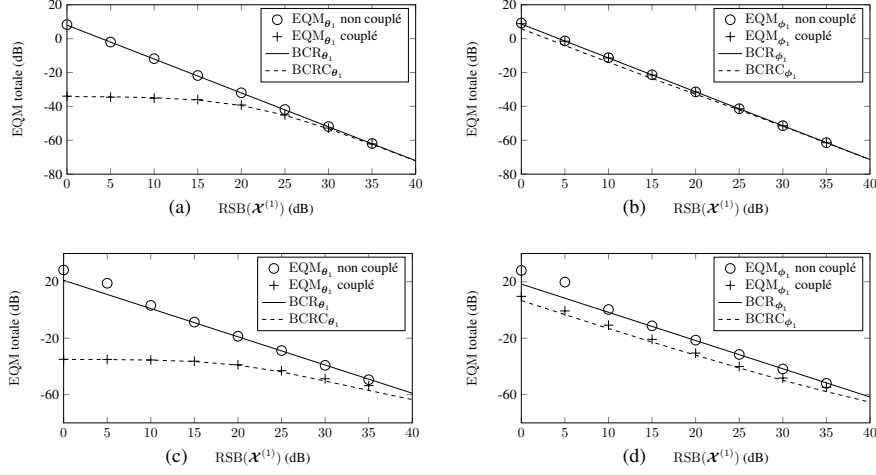


FIGURE 1 – L’EQM totale et les bornes pour les décompositions PC couplées. Figures (a) et (b) pour les éléments de $\mathbf{A}^{(1)}$ i.i.d et figures (c) et (d) pour $\mathbf{A}^{(1)}$ avec deux vecteurs colonnes quasi-colinéaires. Le RSB de $\mathbf{X}^{(2)}$ est fixé à 20dB.

Nous envisageons deux scénarios. Pour le premier, tous les facteurs matriciels sont générés aléatoirement suivant une loi gaussienne complexe circulaire. Le deuxième scénario est quasiment le même sauf pour deux vecteurs colonnes de la matrice $\mathbf{A}^{(1)}$, qui sont presque colinéaires (coefficient de corrélation $\rho = 0.99$). Les premières lignes des facteurs non couplés sont mis à 1. L’EQM totale est évaluée sur 500 réalisations de Monte-Carlo. Le RSB³ de $\mathbf{X}^{(2)}$ est fixé à 20dB tandis que celui de $\mathbf{X}^{(1)}$ varie entre 0 et 40dB.

Nous constatons un gain conséquent de la BCRC par rapport à la BCR pour l’estimation de θ_1 . Pour le paramètre ϕ_1 , la différence entre la BCR et BCRC est petite dans le cas où les éléments de $\mathbf{A}^{(1)}$ sont i.i.d., mais il y a un écart conséquent pour le cas où $\mathbf{A}^{(1)}$ contient deux vecteurs colonnes presque colinéaires. Nous pouvons également constater que l’EQM totale suit de près les performances prédites par les bornes pour θ_1 et donc les deux algorithmes MCA semblent asymptotiquement efficace. Cette application montre bien que l’estimation conjointe à partir des données fusionnées permet de diminuer l’EQM totale.

5 Conclusions

Nous avons proposé une étude d’estimation paramétrique pour la fusion de données hétérogènes partiellement couplées. Quand les relations inhérentes entre deux ensembles de données sont correctement spécifiées, les performances d’estimation jointe peuvent être bornées par la BCRC. De plus, nous avons montré dans un contexte très général que la précision d’estimation jointe à partir des données fusionnées est toujours (asymptotiquement) meilleure en termes de l’EQM que l’estimation individuelle, non couplée, même pour les paramètres qui n’ont pas de liens directs. Cette proposition est vérifiée à travers un exemple de décompositions tensorielles couplées.

3. Le RSB pour les tenseurs aléatoires complexes est donné par
$$\text{RSB}(\mathbf{X}^{(i)}) = 10 \log_{10} \left\{ \frac{\mathbb{E}[\|\mathcal{T}^{(i)}\|^2]}{\mathbb{E}[\|\mathbf{X}^{(i)} - \mathcal{T}^{(i)}\|^2]} \right\} = 10 \log_{10} \left(\frac{R_i}{\sigma_i^2} \right),$$
 où σ_i est la variance du bruit gaussien complexe.

Références

- [1] J. Sui, T. Adalı, Q. Yu, J. Chen, and V. Calhoun, “A review of multivariate methods for multimodal fusion of brain imaging data,” *J. Neurosci. Meth.*, vol. 204, pp.68–81, 2012.
- [2] E. Acar, A. Lawaetz, M. Rasmussen, and R. Bro, “Structure-revealing data fusion model with applications in metabolomics,” in *Conf. Proc. IEEE Eng. Med. Biol. Soc.* IEEE, 2013, pp. 6023–6026.
- [3] B. Ermiş, E. Acar, and A. T. Cemgil, “Link prediction in heterogeneous data via generalized coupled tensor factorization,” *Data Min. Knowl. Discov.*, vol. 29, pp.203–236, 2013.
- [4] D. Lahat, T. Adalı, and C. Jutten, “Multimodal data fusion : an overview of methods, challenges, and prospects,” *Proc. IEEE*, vol. 103, pp. 1449–1477, 2015.
- [5] J. D. Gorman and A. O. Hero, “Lower bounds for parametric estimation with constraints,” *IEEE Trans. Inf. Theory*, vol. 36, pp. 1285–1301, 1990.
- [6] P. Stoica and B. C. Ng, “On the Cramér-Rao bound under parametric constraints,” *IEEE Sig. Proc. Lett.*, vol. 5, pp. 177–179, 1998.
- [7] T. J. Moore Jr., “A theory of Cramér-Rao bounds for constrained parametric models,” Ph.D. dissertation, University of Maryland, 2010.
- [8] S. Sahnoun and P. Comon, “Joint Source Estimation and Localization,” *IEEE Trans. Sig. Proc.*, vol. 63, pp.2485–2495, 2015.
- [9] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, “Blind parafac receivers for DS-CDMA systems,” *IEEE Trans. Sig. Proc.*, vol. 48, pp.810–823, 2000.
- [10] D. Nion and N. D. Sidiropoulos, “Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar,” *IEEE Trans. Signal Proc.*, vol. 58, pp.5693–5705, 2010.
- [11] M. Sørensen and L. D. Lathauwer, “Coupled tensor decompositions for applications in array signal processing,” in *Proc. CAMSAP*, 2013, pp.228–231.
- [12] X. Liu and N. D. Sidiropoulos, “Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays,” *IEEE Trans. Signal Proc.*, vol. 49, pp.2074–2086, 2001.
- [13] 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, pp.3482–3486.
- [14] R. Cabral Farias, J. E. Cohen, and P. Comon, “Exploring multimodal data fusion through joint decompositions with flexible couplings,” *IEEE Trans. Sig. Proc.*, Sept. 2016, vol.64, pp.4830–4844. arxiv :1505.07717.
- [15] D. Slepian, “Estimation of signal parameters in the presence of noise,” *Trans. IRE Professional Group Inf. Theory*, vol. 3, pp.68–69, 1954.
- [16] P. Comon, “Tensors : a brief introduction,” *IEEE Sig. Proc. Magazine*, vol. 31, pp.44–53, May 2014, special issue on BSS. hal-00923279.
- [17] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM review*, vol. 51, pp.455–500, 2009.
- [18] J. W. Brewer, “Kronecker products and matrix calculus in system theory,” *IEEE Trans. on Circuits and Systems*, vol. 25, pp.114–122, Sep. 1978.
- [19] R. Bro, “Multi-way Analysis in the Food Industry : Models, Algorithms, and Applications,” Ph.D. dissertation, University of Amsterdam, The Netherlands, 1998.