

Avant-propos

Les décompositions tensorielles jouent un rôle sans cesse croissant pour la représentation, la compression, l'analyse, la classification, la fouille et la fusion de données massives, multidimensionnelles, multimodales, et hétérogènes. L'objectif de ce volume est de fournir une présentation à la fois synthétique, *self-contained*, et didactique des outils tensoriels en vue de faciliter leur utilisation en traitement du signal, mais aussi dans bien d'autres domaines d'application.

Notre principale motivation pour la rédaction de cet ouvrage a été d'introduire les connaissances nécessaires au développement de méthodes de traitement basées sur les tenseurs. Ces connaissances sont introduites tant d'un point de vue théorique que pratique, en s'appuyant sur la démonstration des principaux résultats fondamentaux, avec l'utilisation de multiples tableaux pour résumer ces résultats, et en fournissant de nombreux exemples pour illustrer les opérations et les décompositions tensorielles présentées. Ainsi, plusieurs illustrations sont fournies dans le chapitre introductif *via* un survol de différents problèmes pouvant être résolus à l'aide des tenseurs, avec une brève description des principales décompositions tensorielles et de quelques exemples d'application. De même, dans les chapitres 1 et 5, certaines décompositions matricielles et tensorielles sont illustrées à l'aide d'applications issues du traitement du signal. Notre motivation a été renforcée par le fait qu'il existe très peu d'ouvrages en langue française consacrés aux tenseurs et à leurs applications, si ce n'est ceux traitant des tenseurs et du calcul tensoriel en physique.

Le calcul tensoriel faisant largement appel au calcul matriciel, les deux premiers chapitres sont consacrés à des rappels sur les décompositions matricielles et sur les produits de Hadamard, Kronecker et Khatri-Rao, ces deux derniers produits intervenant dans les dépliements matriciels de modèles tensoriels.

Dans le [chapitre 1](#), une attention particulière est portée sur la SVD et quelques-unes de ses applications. En particulier, les relations entre SVD et ACP (analyse en composantes

principales) d'une part, et entre SVD et séparation aveugle de sources d'autre part, sont mises en évidence.

Dans le [chapitre 2](#), les principales propriétés des produits de Hadamard, Kronecker et Khatri-Rao sont établies en introduisant une convention indicielle qui généralise la convention de sommation d'Einstein, et qui est très utile aussi pour le calcul tensoriel. Des applications sont présentées concernant la définition et le rangement de dérivées partielles de fonctions, et la résolution d'équations matricielles telles que les équations de Sylvester et de Lyapunov. Le problème de l'estimation des facteurs matriciels de produits de Khatri-Rao et de Kronecker est résolu à l'aide d'algorithmes non itératifs, basés sur une approximation de rang un de matrices ou de tenseurs. Ces algorithmes sont très souvent utilisés pour l'estimation paramétrique de modèles tensoriels.

Après l'introduction de différents ensembles de tenseurs en lien direct avec différentes formes multilinéaires, le [chapitre 3](#) présente de façon détaillée les principales opérations avec des tenseurs. La notion de système tensoriel est aussi introduite, avec la résolution de certains systèmes à l'aide de la méthode des moindres carrés.

Dans le [chapitre 4](#), différentes définitions des valeurs propres et des valeurs singulières d'un tenseur sont données, avec l'introduction des notions de tenseur défini positif/négatif et de tenseurs orthogonalement/unitairement semblables. Le problème de la meilleure approximation de rang un d'un tenseur est également abordé.

Le [chapitre 5](#) est consacré à la présentation des décompositions tensorielles de base. Différentes notions de rang associées à ces décompositions sont introduites, et les algorithmes d'estimation standard ALS et HOSVD sont décrits. L'utilisation des décompositions tensorielles est illustrée par des exemples relatifs à la modélisation de signaux et de systèmes.

Une annexe est également fournie pour rappeler quelques résultats fondamentaux concernant les statistiques (cumulants et spectres) d'ordre élevé de signaux aléatoires, qui sont utilisées dans le [chapitre 5](#) pour les problèmes de séparation de sources et d'identification de systèmes linéaires RIF. L'identification de systèmes quadratiques homogènes à l'aide de cumulants d'ordre élevé est aussi traitée.

Les connaissances regroupées dans cet ouvrage sont présentées de façon accessible à tout lecteur non familier des tenseurs. Les prérequis nécessaires pour faciliter la lecture sont principalement les concepts de base en algèbre linéaire qui ont été introduits dans le volume 1. Chaque chapitre peut être lu indépendamment des autres, avec seulement quelques références croisées. Le lecteur ayant une bonne connaissance des outils matriciels peut donc démarrer sa lecture avec le chapitre introductif, puis lire les chapitres 3 à 5 qui constituent le cœur de l'ouvrage, résultat de recherches menées pendant près de deux

décennies, avec comme finalité l'utilisation des tenseurs en traitement du signal, plus particulièrement pour la conception de nouveaux systèmes de communication sans fil, et pour l'exploitation de décompositions tensorielles dans le contexte de l'identification de systèmes tant linéaires que non linéaires.

Avec l'objectif de couvrir de nombreux sujets, il en résulte que les connaissances fournies sont forcément incomplètes. Ceci concerne plus spécialement les modèles tensoriels de complexité réduite, avec des contraintes de couplage, de non-négativité, de parcimonie, et/ou de rang faible. C'est aussi le cas des aspects algorithmiques associés, tant pour l'estimation de paramètres de ces modèles que pour la résolution de certains problèmes fondamentaux comme, par exemple, la récupération d'informations manquantes dans des tenseurs creux de grandes dimensions, c'est-à-dire le problème de complétion tensorielle. De plus, il est important de souligner que plusieurs problèmes restent ouverts, comme le choix de la meilleure décomposition pour un jeu de données à traiter, et la détermination de l'ordre le plus approprié pour la décomposition choisie. Ces sujets seront approfondis dans le volume 3.

J'espère que cet ouvrage permettra à un large public d'étudiants, de chercheurs et d'ingénieurs de se familiariser avec les outils tensoriels, en permettant de nouvelles applications des tenseurs non seulement en traitement du signal, mais pour le traitement de données de façon plus générale.

Je tiens à exprimer mes sincères remerciements à mon collègue Vicente Zarzoso pour ses remarques judicieuses qui ont permis d'améliorer le contenu. Je serai reconnaissant aux lecteurs qui me feront parvenir leurs remarques *via* mon adresse e-mail¹.

1. favier@i3s.unice.fr.

Introduction

Le premier ouvrage de cette série a été consacré à une introduction des matrices et des tenseurs (d'ordre supérieur à deux) du point de vue de leur structure algébrique, en mettant en évidence leurs similitudes et leurs différences, ainsi que leurs liens avec la représentation des applications linéaires, bilinéaires et multilinéaires. Ce volume 2 a pour finalité d'approfondir les opérations et les décompositions tensorielles.

Dans cette introduction, notre objectif est de motiver l'utilisation des tenseurs par des réponses apportées à cinq questions qu'un utilisateur potentiel peut et doit se poser :

- quels avantages des approches tensorielles ?
- pour quelles utilisations ?
- pour quels domaines d'application ?
- avec quelles décompositions tensorielles ?
- avec quelles fonctions de coût et algorithmes d'optimisation ?

Bien que très incomplètes, ces réponses ont pour but avant tout de :

- présenter les avantages des approches tensorielles par rapport aux méthodes matricielles ;
- mettre en évidence quelques-unes des utilisations des outils tensoriels ;
- donner un aperçu de la très grande diversité des problèmes pouvant être résolus à l'aide des tenseurs, avec la description de quelques exemples d'application ;
- introduire les trois décompositions tensorielles les plus utilisées, en indiquant certaines de leurs propriétés et en comparant leur complexité paramétrique ;

- poser différents problèmes reposant sur des modèles tensoriels, en termes de fonction de coût à optimiser ;
- décrire divers types de traitement, avec un bref survol des méthodes d’optimisation pouvant être utilisées.

1.1. Quels avantages des approches tensorielles ?

Dans la plupart des applications, un tenseur \mathcal{X} d’ordre N est vu comme un tableau de nombres réels ou complexes. L’élément courant du tenseur est noté x_{i_1, \dots, i_N} , l’indice $i_n \in \langle I_n \rangle \triangleq \{1, \dots, I_n\}$, pour $n \in \langle N \rangle \triangleq \{1, \dots, N\}$, étant associé au mode n , avec I_n comme dimension¹, c’est-à-dire, le nombre d’éléments pour le mode n . L’ordre du tenseur est le nombre N d’indices, donc de modes. Les tenseurs seront écrits avec des lettres calligraphiques². Un tenseur d’ordre N et d’élément courant x_{i_1, \dots, i_N} sera noté $\mathcal{X} = [x_{i_1, \dots, i_N}] \in \mathbb{K}^{I_1 \times \dots \times I_N}$, avec $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} , selon que le tenseur est à valeurs réelles ou à valeurs complexes, $I_1 \times \dots \times I_N$ représentant la dimension de \mathcal{X} .

Chaque mode (appelé aussi voie) peut en général avoir l’une des principales interprétations suivantes : i) une source d’information (utilisateur, patient, client, essai, etc.), ii) un type d’entité attaché aux données (articles/produits, types de musique, de film, etc.), iii) un mot-clé (*tag*) permettant de caractériser un article, une musique, un film, etc., iv) une modalité associée à un enregistrement et à une diversité dans différents domaines (espace, temps, fréquence, longueur d’onde, polarisation, couleur, etc.).

Ainsi, une image numérique en couleur peut être représentée à l’aide d’un tenseur tridimensionnel de données (en pixels), avec deux modes spatiaux relatifs aux lignes (largeur) et aux colonnes (hauteur), et le mode canal (couleurs RVB). Par exemple, une image couleur peut être représentée à l’aide d’un tenseur de dimension $1\,024 \times 768 \times 3$, le troisième mode étant relatif à l’intensité des trois couleurs RVB (rouge, verte, bleue). Dans le cas d’une image volumétrique, il y a trois modes spatiaux (*largeur* \times *hauteur* \times *profondeur*) et les points de l’image sont appelés des voxels. En imagerie hyperspectrale, outre les deux dimensions spatiales, la troisième dimension est relative à la longueur d’onde d’émission, dans une bande spectrale.

Les approches tensorielles présentent les avantages suivants par rapport aux approches matricielles :

-
1. La dimension I_n correspond à la borne supérieure de l’intervalle de variation $\langle I_n \rangle$ de l’indice i_n . Cette borne sera généralement représenté par la même lettre que l’indice, mais en majuscule, comme par exemple (I, J, K) pour les indices (i, j, k) .
 2. Les scalaires, les vecteurs, et les matrices seront écrits avec des lettres minuscules, des lettres minuscules en gras, et des lettres majuscules en gras : $a, \mathbf{a}, \mathbf{A}$, respectivement.

– la propriété d’unicité essentielle³ sous de faibles conditions, pour certaines décompositions tensorielles comme PARAFAC (*parallel factors*) (Harshman 1970), unicité qui n’est satisfaite par les décompositions matricielles que sous certaines conditions contraignantes d’orthogonalité, de non-négativité, ou de structures particulières (triangulaire, Vandermonde, Toeplitz, etc.) ;

– la possibilité de résoudre certains problèmes, comme l’identification de canaux de communication, par exemple, directement à partir des signaux mesurés, sans faire appel au calcul de statistiques d’ordre élevé de ces signaux, ou à de longues séquences d’apprentissage. Les traitements déterministes et semi-aveugles qui en résultent peuvent être effectués avec des enregistrements de signaux de plus courte durée que pour les méthodes statistiques basées sur l’estimation de moments ou de cumulants d’ordre élevé. De même, pour le problème de séparation aveugle de sources, les approches tensorielles permettent de traiter le cas de systèmes sous-déterminés, c’est-à-dire, avec plus de sources que de capteurs ;

– la possibilité de compresser de grands ensembles de données *via* une tensorisation des données et l’utilisation d’une décomposition tensorielle, en particulier à l’aide d’une approximation tensorielle de rang multilinéaire faible ;

– une plus grande flexibilité pour la représentation et le traitement de données multimodales, en considérant les modalités de façon séparée au lieu d’empiler les données dans un vecteur ou une matrice. Ceci permet de préserver la structure multilinéaire des données, et donc de prendre en compte les interactions entre modes ;

– la prise en considération de davantage de modalités, donc d’informations complémentaires, *via* une représentation tensorielle des données, ce qui permet d’améliorer les performances de certains systèmes comme, par exemple, les systèmes de communication sans fil, de recommandation, d’aide au diagnostic ou de surveillance, en facilitant et en rendant plus efficaces les opérations de détection, d’analyse, d’interprétation, de reconnaissance et de classification.

Cela a débouché sur une généralisation (à des ordres supérieurs à deux) de certains algorithmes matriciels comme la SVD (*singular value decomposition*) en la MLSVD (*multilinear SVD*), appelée aussi HOSVD (*higher-order SVD*) (de Lathauwer *et al.* 2000a), et de certaines méthodes de traitement du signal comme, par exemple, l’analyse en composantes principales (ACP, ou PCA pour *principal component analysis*) en MPCA (*multilinear PCA*) (Lu *et al.* 2008) ou TRPCA (*tensor robust PCA*) (Lu *et al.* 2020), et l’analyse en composantes indépendantes (ACI, ou ICA pour *independent component analysis*) en MICA (*multilinear ICA*) (Vasilescu et Terzopoulos 2005) ou tensor PICA (*probabilistic ICA*) (Beckmann et Smith 2005).

3. L’unicité essentielle d’une décomposition signifie qu’elle est unique à des ambiguïtés près de permutation et de facteur d’échelle au niveau des colonnes de ses matrices facteurs.

Il est important de noter ici qu'avec la modélisation tensorielle d'un problème, le nombre de modalités à prendre en compte peut être augmenté soit *via* une augmentation de l'ordre du tenseur de données, soit par le couplage de décompositions tensorielles et/ou matricielles, ayant un ou plusieurs modes en commun. Dans le cas d'un couplage, on parle de fusion de données à l'aide d'une factorisation tensorielle/matricielle couplée. Deux exemples d'un tel couplage seront présentés dans cette introduction : l'un concerne le couplage de signaux EEG avec des données d'imagerie par résonance magnétique fonctionnelle (IRMf, ou fMRI pour *functional magnetic resonance imaging*) pour l'étude du fonctionnement du cerveau, tandis que l'autre est relatif à la fusion d'images hyperspectrales et multispectrales en télédétection.

Dans le volume 3 de la série, l'augmentation du nombre de modalités par une augmentation de l'ordre du tenseur de données sera illustrée, avec la présentation unifiée de différents modèles de systèmes de communication sans fil dont la conception est basée sur une approche tensorielle. Avec comme objectif d'améliorer la performance des systèmes tant en termes de transmission que de réception, l'idée est d'utiliser simultanément plusieurs diversités dans différents domaines (espace, temps, fréquence, code, etc.), chaque diversité étant associée à un mode du tenseur des signaux reçus par les antennes de réception du système. Des modèles tensoriels couplés seront aussi présentés dans le contexte des systèmes de communication coopératifs avec relais.

I.2. Pour quelles utilisations ?

À l'ère du *Big Data*, dans un monde de plus en plus digitalisé et interconnecté *via* Internet et l'Internet des objets (IdO, ou IoT pour *Internet of Things*), le traitement de l'information numérique occupe une place centrale dans de nombreux domaines applicatifs, chacun de ces domaines ayant ses spécificités et son propre besoin de compétences souvent pluridisciplinaires. La pluridisciplinarité peut être liée tant à l'aspect multimodal des données qu'aux techniques de traitement à mettre en œuvre.

Ainsi, les futurs systèmes de traitement d'information dits intelligents devront intégrer des outils de représentation tels que les tenseurs et les graphes, des méthodes de traitement du signal et des images, couplées à des techniques d'intelligence artificielle faisant appel aux réseaux de neurones artificiels et au *machine learning* pour l'apprentissage.

Les besoins de tels systèmes sont nombreux et variés, en termes de stockage, de visualisation (représentation 3D, réalité virtuelle, diffusion d'œuvres d'art), de transmission, d'imputation, de prédiction/prévision, d'analyse, de classification, de fouille et de fusion de données multimodales et hétérogènes. Consulter (Lahat *et al.* 2015 ; Papalexakis *et al.* 2016) pour une présentation de différents exemples de fusion et de fouille de données basés sur des modèles tensoriels.

Quelques-unes des principales utilisations des outils tensoriels concernent :

- la décomposition ou la séparation d'ensembles de données hétérogènes en des composantes/facteurs ou des sous-espaces, dans le but d'exploiter la structure multimodale des données et d'extraire l'information utile pour les utilisateurs, à partir de données/mesures incertaines/bruitées provenant de sources d'information différentes et/ou de différents types de capteurs. On peut ainsi extraire des caractéristiques dans différents domaines (spatio-temporel-fréquentiel) en vue d'une classification et d'une prise de décision ;

- l'imputation de données manquantes dans des bases de données incomplètes, à l'aide d'une modélisation tensorielle de rang faible, le manque de données pouvant provenir de capteurs ou de liens de communication défectueux, par exemple ; on parle alors de complétion tensorielle (*tensor completion*, et *matrix completion* dans le cas matriciel) (Candès et Recht 2009 ; Signoretto *et al.* 2011 ; Liu *et al.* 2013) ;

- la réduction de dimensionnalité de bases de données multimodales et hétérogènes de très grandes dimensions (*Big Data*) en résolvant un problème d'approximation de tenseur de rang multilinéaire faible ;

- l'acquisition comprimée multivoie (MCS pour *multiway compressed sensing*, appelé aussi *compressive sampling*) (Sidiropoulos et Kyriakidis 2012 ; Cai et Cichocki 2013). Les algorithmes de reconstruction opèrent alors simultanément vis-à-vis des différents modes d'un tenseur de données creux, *via* une modélisation CPD ou de Tucker, sans passer par une étape de vectorisation, comme c'est le cas pour les algorithmes standards tels que BP (*basis pursuit*) ou MP (*matching pursuit*). Ces méthodes MCS étendent au cas de données multimodales les méthodes CS standards qui consistent à reconstruire un signal ou une image à partir d'une acquisition comprimée dans une base préchoisie telle que l'information utile soit représentée de façon parcimonieuse dans cette base (Candès et Wakin 2008 ; Candès et Plan 2010) ;

- l'apprentissage comprimé multilinéaire (MCL pour *multilinear compressive learning*) (Tran *et al.* 2020). Ce type de technique combine une acquisition comprimée avec le *machine learning* ;

- la prévision de données à l'aide de séries temporelles tensorielles ;

- le filtrage multimodal et le débruitage de données tensorielles ;

- la fusion de données à partir de décompositions tensorielles couplées ;

- la modélisation et l'identification de systèmes dynamiques non linéaires ;

- la modélisation et la résolution de systèmes tensoriels d'équations ;

- la conception de systèmes comme, par exemple, des systèmes de communication multi-antennes (ou MIMO, pour *multiple-input multiple-output*) coopératifs ; ce type d'application qui a conduit au développement de plusieurs nouveaux modèles tensoriels, sera traité dans les deux prochains volumes de la série.

Les tenseurs peuvent aussi être utilisés avec des réseaux de neurones à couches entièrement connectées, *via* la tensorisation d'une couche à l'aide d'une décomposition polyadique canonique (CPD pour *canonical polyadic decomposition*), d'une décomposition de Tucker, ou d'un train de tenseurs (TT), voir section I.4. Cela permet de réduire considérablement la complexité paramétrique, et donc l'encombrement mémoire. Cette compression d'information dans les réseaux de neurones à couches *via* l'utilisation de décompositions tensorielles permet d'augmenter le nombre d'unités cachées (Novikov *et al.* 2015). Les tenseurs utilisés conjointement avec les réseaux de neurones du type perceptron multicouche pour des problèmes de classification d'images fournissent un taux d'erreur plus faible, avec un temps de calcul moindre que celui nécessité par les réseaux de neurones seuls (Chien et Bao 2017).

Par ailleurs, en exploitant une correspondance entre réseaux convolutionnels et décompositions tensorielles hiérarchiques (Hackbusch et Kühn 2012), Cohen *et al.* (2018) analysent différentes architectures de réseaux, avec pour perspective de concevoir le réseau en fonction d'une tâche à effectuer.

Les réseaux neuronaux peuvent également être utilisés pour l'apprentissage du rang d'un tenseur (Zhou *et al.* 2019), ou pour le calcul des valeurs propres et des valeurs singulières, et donc d'une approximation de rang un d'un tenseur (Che *et al.* 2017).

I.3. Pour quels domaines d'application ?

Parmi les très nombreux domaines d'application des tenseurs, on peut citer originellement la psychométrie et la chimiométrie, dans les années 1970 et 1990 respectivement, qui ont ouvert la voie aux applications en traitement du signal et des images, tout particulièrement pour la séparation aveugle de sources, les communications numériques et la vision par ordinateur, dans les années 1990 puis au début des années 2000.

Aujourd'hui, avec l'explosion quantitative des données massives (*big data*) en médecine, astronomie, météorologie, ou avec les communications sans fil de cinquième génération (5G), l'aide au diagnostic médical, ou encore les services *via* le Web, tels que les systèmes de recommandation (vidéo à la demande, vente en ligne, réservation de restaurants et d'hôtels, etc.), la recherche d'information dans des bases de données multimédia (textes, images, enregistrements audio et vidéo), et les réseaux sociaux, les tenseurs sont voués à un intérêt grandissant de la part de différentes communautés scientifiques et du monde industriel.

Nous présentons ci-après quelques exemples d'application des tenseurs faisant appel aux méthodes de traitement du signal et des images :

– la séparation aveugle de sources et l’identification aveugle de systèmes. Ces problèmes sont fondamentaux en traitement du signal. Ils consistent à séparer les signaux d’entrée (appelés aussi sources) et à identifier un système à partir de la seule mesure de ses signaux de sortie, et de l’exploitation de certaines hypothèses sur les signaux d’entrée comme, par exemple, l’indépendance statistique pour l’analyse en composantes indépendantes (Comon 1994), ou l’appartenance à un alphabet fini dans le contexte des communications numériques. Ce type de traitement intervient en particulier pour l’estimation du canal de communication et des symboles d’information émis par un émetteur. C’est le cas aussi des problèmes de séparation de parole et de musique, ou encore pour le traitement de signaux sismiques ;

– l’utilisation de décompositions tensorielles pour l’analyse de signaux biomédicaux (EEG, MEG, ECG, EOG⁴), dans les domaines spatio-temporel-fréquentiel, en vue d’une aide au diagnostic médical. Ainsi, par exemple, (Acar *et al.* 2007) utilisent une modélisation PARAFAC de signaux EEG pour l’analyse d’une crise d’épilepsie tandis que (Becker *et al.* 2014) font appel au même type de décomposition pour la localisation de sources dans des signaux EEG ;

– l’analyse du fonctionnement du cerveau à l’aide d’une fusion de données d’imagerie (fMRI) et de signaux biomédicaux (EEG et MEG), avec comme objectif de faire appel à des examens médicaux non invasifs, voir le tableau I.4 ;

– l’analyse et la classification de données d’imagerie hyperspectrale, utilisée dans de nombreux domaines (médecine, environnement, agriculture, surveillance, astrophysique, etc.). L’imagerie hyperspectrale qui consiste à faire l’acquisition d’une même scène dans un très grand nombre de bandes spectrales contiguës, est souvent caractérisée par une faible résolution spatiale. Afin d’améliorer cette résolution, les images hyperspectrales sont fusionnées avec des images multispectrales, avec pour objectif de reconstruire des images hyperspectrales de haute résolution (HSR pour *hyperspectral super-resolution*). Au cours de la dernière décennie, ce problème de fusion a fait l’objet de nombreux travaux, basés tout d’abord sur des factorisations matricielles puis tensorielles. Pour les méthodes tensorielles, on peut distinguer celles qui exploitent des modèles CPD couplés de faible rang (Kanatsoulis *et al.* 2018) et celles qui reposent sur des modèles de Tucker couplés de faible rang multilinéaire (Li *et al.* 2018 ; Prévost *et al.* 2020). Consulter (Kanatsoulis *et al.* 2018 ; Li *et al.* 2018) pour un survol bibliographique des méthodes de fusion HSR. Dans le tableau I.4, nous considérons la solution basée sur une décomposition de Tucker couplée dont le cœur est creux (modèle CSTF pour *coupled sparse tensor factorization*), proposée par (Li *et al.* 2018) ;

4. Signaux d’électroencéphalographie (EEG), de magnétoencéphalographie (MEG), d’électrocardiographie (ECG) et d’électro-oculographie (EOG).

– la prévision de données (trafic routier, climat, marchés boursiers, etc.), en utilisant des décompositions tensorielles ou des réseaux neuronaux tensorisés, voir tableau I.1 ;

– la modélisation et l’identification de systèmes non linéaires, *via* une représentation tensorielle des noyaux de Volterra ou des systèmes de Wiener-Hammerstein (Kibangou et Favier 2009a, 2010 ; Favier et Kibangou 2009 ; Favier et Bouilloc 2009, 2010 ; Favier *et al.* 2012a) ;

– l’identification de systèmes trilinéaires séparables, linéaires vis-à-vis du signal d’entrée et trilinéaires par rapport aux coefficients de la réponse impulsionnelle, modélisée comme un produit de Kronecker de trois réponses impulsionnelles linéaires (Elisei-iliescu *et al.* 2020). Un tel système est à comparer à un filtre de Volterra d’ordre trois, linéaire en les coefficients des noyaux et trilinéaire par rapport à l’entrée ;

– le débruitage d’images multispectrales et hyperspectrales, et de signaux sismiques, basé sur l’utilisation de méthodes de filtrage multivoie ou de décompositions tensorielles PARAFAC ou de Tucker (Muti *et al.* 2005, 2008 ; Marot *et al.* 2008 ; Liu *et al.* 2012 ; Bai *et al.* 2018 ; Wu *et al.* 2020) ;

– la reconnaissance faciale à partir de tenseurs de visage, à des fins d’authentification et d’identification pour des systèmes de vidéosurveillance. Pour la reconnaissance de visage, des photos des personnes à reconnaître sont stockées dans une base de données, pour différents éclairages, différentes expressions faciales, et selon plusieurs profils, pour chaque personne. Dans (Vasilescu et Terzopoulos 2002), le tenseur d’images de visages est d’ordre cinq et de dimension : $28 \times 5 \times 3 \times 3 \times 7\,943$, correspondant aux modes *personnes* \times *profils* \times *éclairages* \times *expressions* \times *pixels par image*. Pour un survol de différents systèmes de reconnaissance faciale, consulter (Arachchilage et Izquierdo 2020) ;

– la détection d’anomalies basée sur une modélisation tensorielle, utilisée en imagerie hyperspectrale, et dans les systèmes de surveillance (Fanaee et Gama 2016 ; Zhang *et al.* 2016) ;

– la conception de récepteurs semi-aveugles pour des systèmes de communication MIMO point à point ou coopératifs représentés à l’aide de modèles tensoriels ; consulter les articles de synthèse de (de Almeida *et al.* 2016 ; da Costa *et al.* 2018).

Dans le tableau I.1, nous présentons quelques exemples de tenseurs de signaux, de séries temporelles, et d’images, en précisant la nature des modes.

D’autres domaines d’application sont considérés dans le tableau I.2.

Ci-après, nous donnons quelques détails sur l’application relative aux systèmes de recommandation qui jouent aujourd’hui un rôle important pour de nombreux sites sur le Web. Ces systèmes ont pour finalité d’aider les utilisateurs dans leur choix d’articles (*items*) à partir de mots-clés (*tags*) affectés par les utilisateurs aux articles qui peuvent

être, par exemple, des films, des livres, des enregistrements musicaux, des pages web, des produits en vente sur des sites d'e-commerce, etc. Ainsi, un système de recommandation de base repose sur les trois modes suivants : *utilisateurs* × *articles* × *mots clés* (*users* × *items* × *tags*).

Signaux	Modes	Références
Traitement d'antenne	espace (antennes) × temps × sous-réseau de capteurs espace × temps × polarisation	(Sidiropoulos <i>et al.</i> 2000a) (Raimondi <i>et al.</i> 2017)
Communications numériques	espace (antennes) × temps × code espace × blocs × périodes symboles × code × fréquences	(Sidiropoulos <i>et al.</i> 2000b) (Favier et de Almeida 2014b)
ECG	espace (électrodes) × temps × fréquences	(Padhy <i>et al.</i> 2019)
EEG	espace (électrodes) × temps × fréquences × sujets ou essais	(Becker <i>et al.</i> 2014 ; Cong <i>et al.</i> 2015)
EEG + fMRI	sujets × électrodes × temps + sujets × voxels (modèles couplés par le mode sujets)	(Acar <i>et al.</i> 2017)
Séries temporelles	Modes	Références
Environnement/climat	latitude × longitude × altitude × température × humidité × vitesse du vent	(Rogers <i>et al.</i> 2013)
Marchés boursiers	entreprises × événements × opinions × temps	(Li <i>et al.</i> 2016)
Images	Modes	Références
Images en couleur	espace (largeur) × espace (hauteur) × canal (couleurs)	
Séquences vidéo en niveaux de gris	espace (largeur) × espace (hauteur) × temps	
Séquences vidéo en couleur	espace × espace × canal × temps	
Images hyperspectrales	espace × espace × bandes spectrales	(Makantasis <i>et al.</i> 2018)
Vision par ordinateur	personnes × profils × éclairages × expressions × pixels	(Vasilescu et Terzopoulos 2002)

Tableau I.1. *Tenseurs de signaux et d'images*

Les techniques de filtrage collaboratif utilisent les avis, ou les évaluations à l'aide d'un système de notation, d'un ensemble de personnes pour fournir une liste de recommandations à un utilisateur particulier. Ce type de filtrage est utilisé, par exemple, par un site comme Netflix pour la location de DVD. Les méthodes de filtrage collaboratif sont classées en trois catégories selon que le filtrage est basé sur a) la mémoire et une

mesure de similarité, b) un modèle reposant sur une factorisation matricielle à l'aide d'algorithmes tels que la SVD ou la NMF (*non-negative matrix factorization*), c) une combinaison des deux techniques précédentes, ce qui correspond à des approches de filtrage collaboratif hybride. Consulter (Luo *et al.* 2014 ; Bokde *et al.* 2015) pour des approches basées sur une factorisation matricielle. D'autres techniques de filtrage dit passif exploitent les données d'une matrice de relations entre articles, permettant de déduire des recommandations pour un utilisateur à partir de corrélations entre les articles et les choix déjà effectués par cet utilisateur, sans utiliser un système de notation. On parle alors d'approche basée sur le contenu.

Domaines	Modes	Références
Phonétique	sujets \times voyelles \times formants	(Harshman 1970)
Chimimétrie (fluorescence)	excitation \times émission \times échantillons (longueurs d'onde d'excitation et d'émission)	(Bro 1997, 2006 ; Smilde <i>et al.</i> 2004)
Systèmes de recommandation contextuelle	utilisateurs \times articles \times mots clés \times contexte 1 $\times \dots \times$ contexte N	(Rendle et Schmidt-Thieme 2010) (Symeonidis et Zioupos 2016) (Frolov et Oseledets 2017)
Trafic routier (mesures de vitesse)	espace (capteurs) \times temps (jours) \times temps (semaines) (périodes de 15s et de 24h)	(Goulart <i>et al.</i> 2017) (Tan <i>et al.</i> 2013 ; Ran <i>et al.</i> 2016)
Musique	types de musique \times fréquences \times fréquences utilisateurs \times mots clés \times chansons enregistrements \times caractéristiques (audio) \times segments	(Panagakis <i>et al.</i> 2010) (Nanopoulos <i>et al.</i> 2010) (Benetos et Kotropoulos 2008)
Bioinformatique	médicaments \times cibles \times maladies	(Wang <i>et al.</i> 2019)

Tableau I.2. *Autres domaines d'application*

Les systèmes de recommandation peuvent utiliser également des informations sur les utilisateurs (âge, nationalité, localisation géographique, participation à des réseaux sociaux, etc.), ainsi que sur les articles (types de musique, de film, classes d'hôtel, etc.). On parle alors d'informations contextuelles. La prise en compte de ces informations supplémentaires permet d'améliorer la pertinence des recommandations, au prix d'une augmentation de la dimensionnalité et de la complexité du modèle de représentation des données et donc des algorithmes de traitement. Ceci explique l'importance du rôle joué aujourd'hui par les approches tensorielles pour ce type d'application.

Noter que pour les systèmes de recommandation, les tenseurs de données sont creux. Certains mots clés peuvent être générés automatiquement par le système à

partir de mesures de similarité entre articles. C'est le cas, par exemple, pour les recommandations de musiques en exploitant des caractéristiques acoustiques de chansons (Nanopoulos *et al.* 2010). Des systèmes de recommandation personnalisés prenant en compte le profil, les préférences, et les intérêts de l'utilisateur, peuvent aussi être utilisés par celui-ci pour l'aider à choisir des mots clés existants et à créer de nouveaux mots clés (Rendle et Schmidt-Thieme 2010).

Les articles de (Bobadilla *et al.* 2013 ; Frolov et Oseledets 2017) présentent différents systèmes de recommandation, avec de nombreuses références bibliographiques. Sur le même principe que les systèmes de recommandation, les sites de réseaux sociaux comme Wikipédia, Facebook ou Twitter permettent d'échanger et de partager différents types de données, de produire des contenus, et des mises en relation.

I.4. Avec quelles décompositions tensorielles ?

Il est important de noter que, pour un tenseur $\mathcal{X} \in \mathbb{K}^{I_1 \times \dots \times I_N}$ d'ordre N , le nombre d'éléments à mémoriser est égal à $\prod_{n=1}^N I_n$. En supposant que $I_n = I$ pour $n \in \langle N \rangle$, ce nombre devient I^N , ce qui induit une croissance exponentielle avec l'ordre N du tenseur. Cette explosion du nombre de données avec N et I est appelée la « malédiction » de la dimensionnalité (*curse of dimensionality*) par (Hackbusch et Kühn 2009 ; Oseledets et Tyrtshnikov 2009).

Dans le cas de gros tenseurs de données, les décompositions tensorielles jouent un rôle fondamental pour atténuer ce problème, du fait que le nombre de paramètres qui caractérisent les décompositions est beaucoup plus petit que le nombre d'éléments contenus dans le tenseur original.

Nous allons maintenant introduire trois décompositions basiques : PARAFAC/CANDECOMP/CPD, TD et TTD⁵. Les deux premières seront étudiées de façon approfondie dans le chapitre 5, tandis que la troisième, brièvement introduite dans ce même chapitre, sera considérée de façon plus détaillée dans le volume 3.

Dans le tableau I.3, nous donnons l'expression de l'élément x_{i_1, \dots, i_N} d'un tenseur $\mathcal{X} \in \mathbb{K}^{I_1 \times \dots \times I_N}$, d'ordre N et de dimension $I_1 \times \dots \times I_N$, réel ($\mathbb{K} = \mathbb{R}$) ou complexe ($\mathbb{K} = \mathbb{C}$), pour chacune des trois décompositions mentionnées ci-avant. Leurs paramètres caractéristiques sont mis en évidence, et leur complexité paramétrique est comparée en fonction des dimensions de chaque facteur matriciel et tensoriel, en supposant que $I_n = I$ et $R_n = R$ pour $\forall n \in \langle N \rangle$.

Dans ce qui suit, nous mettons en avant les principaux avantages et inconvénients de chacune de ces décompositions.

5. PARAFAC pour *parallel factors* ; CANDECOMP pour *canonical decomposition* ; CPD pour *canonical polyadic decomposition* ; TD pour *Tucker decomposition* ; TTD pour *tensor train decomposition*.

Décompositions	Notations	Élément x_{i_1, \dots, i_N}
PARAFAC/CPD	$\llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}; R \rrbracket$	$\sum_{r=1}^R \prod_{n=1}^N a_{i_n, r}^{(n)}$
TD	$\llbracket \mathcal{G}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket$	$\sum_{r_1=1}^{R_1} \dots \sum_{r_N=1}^{R_N} g_{r_1, \dots, r_N} \prod_{n=1}^N a_{i_n, r_n}^{(n)}$
TTD	$\llbracket \mathbf{A}^{(1)}, \mathcal{A}^{(2)}, \dots, \mathcal{A}^{(N-1)}, \mathbf{A}^{(N)} \rrbracket$	$\sum_{r_1=1}^{R_1} \dots \sum_{r_{N-1}=1}^{R_{N-1}} \prod_{n=1}^N a_{r_{n-1}, i_n, r_n}^{(n)}$
Décompositions	Paramètres	Complexité
CPD	$\mathbf{A}^{(n)} \in \mathbb{K}^{I_n \times R}, \forall n \in \langle N \rangle$	$O(NIR)$
TD	$\mathcal{G} \in \mathbb{K}^{R_1 \times \dots \times R_N}$ $\mathbf{A}^{(n)} \in \mathbb{K}^{I_n \times R_n}, \forall n \in \langle N \rangle$	$O(NIR + R^N)$
TTD	$\mathbf{A}^{(1)} \in \mathbb{K}^{I_1 \times R_1}, \mathbf{A}^{(N)} \in \mathbb{K}^{R_{N-1} \times I_N}; \mathcal{A}^{(n)} \in \mathbb{K}^{R_{n-1} \times I_n \times R_n}$ $\forall n \in \{2, 3, \dots, N-1\}$	$O(2IR + (N-2)IR^2)$

Tableau I.3. Complexité paramétrique des décompositions CPD, TD et TTD

I.4.1. Décomposition PARAFAC

La décomposition PARAFAC (Harshman 1970), appelée aussi CANDECOMP (Carroll et Chang 1970) et CPD (Hitchcock 1927), d'un tenseur $\mathcal{X} \in \mathbb{K}^{I_1 \times \dots \times I_N}$ d'ordre N consiste en une somme de R tenseurs de rang un, chacun de ces tenseurs étant défini comme le produit extérieur d'une colonne de chacune des N matrices facteurs $\mathbf{A}^{(n)} \in \mathbb{K}^{I_n \times R}$. Quand R est minimal, il correspond au rang (canonique) du tenseur.

Cette décomposition peut être vue comme une extension tensorielle de la SVD matricielle, sans la contrainte d'orthogonalité des matrices facteurs.

Sous certaines conditions à satisfaire par les matrices facteurs, cette décomposition possède la propriété d'unicité essentielle. Outre cette propriété d'unicité, le principal avantage de la décomposition PARAFAC réside dans sa simplicité liée à une faible complexité paramétrique, et, d'autre part, à l'existence d'un algorithme d'estimation facile à mettre en œuvre, à savoir l'algorithme des moindres carrés alternés (ALS pour *alternating least squares*).

De façon optimale, l'estimation paramétrique d'une décomposition PARAFAC nécessite d'utiliser un algorithme d'optimisation non linéaire pour l'estimation des

facteurs matriciels. Toutefois, dans la plupart des applications, il est fait appel à l'algorithme ALS. Cet algorithme consiste à remplacer un problème d'optimisation non linéaire par une séquence de problèmes linéaires résolus par la méthode des moindres carrés. Il a l'avantage d'être très facile à implémenter, chaque facteur matriciel étant estimé de façon alternée. Il a pour inconvénient d'être sous-optimal, avec la possibilité de converger vers des minima locaux, la convergence dépendant de l'initialisation. L'algorithme ALS sera présenté dans la section 5.2.5.7.

Un autre inconvénient du modèle CPD concerne l'estimation du rang R pour laquelle il n'existe pas de solution optimale. De plus, une meilleure estimation de rang R , dans le format CPD, n'existe pas toujours.

La figure I.1 représente graphiquement le modèle PARAFAC $\llbracket \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}; R \rrbracket$ pour un tenseur $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ d'ordre trois, où les facteurs $\mathbf{A}^{(n)} \triangleq [\mathbf{a}_1^{(n)}, \dots, \mathbf{a}_R^{(n)}]$ sont définis à l'aide de leurs colonnes, pour $n \in \{1, 2, 3\}$. Une présentation détaillée du modèle PARAFAC est faite dans la section 5.2.5.

$$\mathcal{X} = \begin{array}{c} \begin{array}{l} I_3 \\ I_2 \\ I_1 \end{array} \begin{array}{l} \mathbf{a}_1^{(3)} \\ \mathbf{a}_1^{(2)} \\ \mathbf{a}_1^{(1)} \end{array} \\ \end{array} + \dots + \begin{array}{c} \begin{array}{l} I_3 \\ I_2 \\ I_1 \end{array} \begin{array}{l} \mathbf{a}_R^{(3)} \\ \mathbf{a}_R^{(2)} \\ \mathbf{a}_R^{(1)} \end{array} \\ \end{array}$$

Figure I.1. Modèle PARAFAC d'ordre trois

I.4.2. Décomposition de Tucker

La décomposition de Tucker (Tucker 1966), notée TD, peut être vue comme une généralisation de la décomposition PARAFAC par la prise en compte d'interactions entre toutes les colonnes des matrices facteurs $\mathbf{A}^{(n)} \in \mathbb{K}^{I_n \times R_n}$ via l'introduction d'un tenseur cœur $\mathcal{G} \in \mathbb{K}^{R_1 \times \dots \times R_N}$ non diagonal. Cette décomposition n'est pas unique en général.

Noter que si $R_n \leq I_n$ pour $\forall n \in \langle N \rangle$, alors le tenseur cœur \mathcal{G} fournit une forme compressée de \mathcal{X} . En choisissant R_n , pour $n \in \langle N \rangle$, comme le rang du dépliement matriciel⁶ mode- n de \mathcal{X} , alors le N -uplet (R_1, \dots, R_N) est minimal, et il est appelé le rang multilinéaire du tenseur.

6. Voir la définition [3.41] du dépliement matriciel mode- n d'un tenseur $\mathcal{X} \in \mathbb{K}^{I_1 \times \dots \times I_N}$, dont les colonnes sont les vecteurs mode- n obtenus en fixant tous les indices à l'exception de i_n .

Quoique le modèle de Tucker ne soit pas unique en général, il peut être déterminé à l'aide de l'algorithme HOSVD (*high-order SVD*), en imposant une contrainte d'orthogonalité sur les facteurs matriciels $\mathbf{A}^{(n)}$. Une forme tronquée, appelée THOSVD (*truncated HOSVD*), fournit une solution approchée au problème de meilleure approximation de rang multilinéaire faible (de Lathauwer *et al.* 2000a). Cette solution qui repose sur un choix *a priori* des dimensions R_n du tenseur cœur, est à rapprocher de la SVD tronquée du cas matriciel, même si elle n'en possède pas la propriété d'optimalité. Elle est très utilisée dans les applications pour réduire la complexité paramétrique des tenseurs de données. L'algorithme HOSVD est décrit dans la section 5.2.1.8.

Le modèle TD a l'avantage, par rapport au modèle CPD, de bénéficier d'un algorithme d'estimation numériquement stable, car basé sur un calcul de SVDs, permettant de déterminer le rang multilinéaire. L'inconvénient de cette décomposition réside dans sa complexité paramétrique, du fait que le nombre de paramètres du tenseur cœur croît exponentiellement avec l'ordre N du tenseur (voir tableau I.3).

La CPD et la HOSVD peuvent être vues comme des généralisations de la SVD matricielle, sans la contrainte d'orthogonalité pour les matrices facteurs de la CPD et sans la contrainte d'un tenseur cœur diagonal pour la HOSVD.

Les décompositions du type CPD sont plutôt utilisées pour l'estimation de paramètres cachés (*latent parameters*) tandis que la décomposition de Tucker est souvent utilisée pour l'estimation des sous-espaces modaux et la réduction de dimensionnalité *via* une approximation de rang (multilinéaire) faible et l'utilisation d'une HOSVD tronquée.

La figure I.2 représente graphiquement le modèle TD $\llbracket \mathcal{G}; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)} \rrbracket$ pour un tenseur $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ d'ordre trois, voir la section 5.2.1 pour une présentation détaillée de ce modèle.

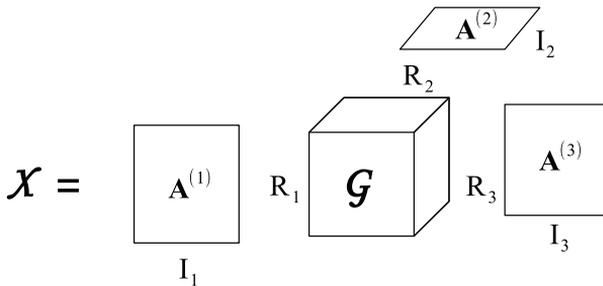


Figure I.2. Modèle de Tucker d'ordre trois

1.4.3. Décomposition TT

La décomposition TT (Oseledets 2011) est constituée d'un train de tenseurs d'ordre trois $\mathcal{A}^{(n)} \in \mathbb{K}^{R_{n-1} \times I_n \times R_n}$, pour $n \in \{2, 3, \dots, N-1\}$, le premier et le dernier wagons du train étant des matrices $\mathbf{A}^{(1)} \in \mathbb{K}^{I_1 \times R_1}$ et $\mathbf{A}^{(N)} \in \mathbb{K}^{R_{N-1} \times I_N}$, ce qui induit $R_0 = R_N = 1$, et donc $a_{r_0, i_1, r_1}^{(1)} = a_{i_1, r_1}^{(1)}$ et $a_{r_{N-1}, i_N, r_N}^{(N)} = a_{r_{N-1}, i_N}^{(N)}$. Les dimensions R_n , pour $n \in \langle N-1 \rangle$, sont appelées les rangs TT. Ils sont égaux aux rangs de certains dépliements matriciels du tenseur original (voir l'équation [5.114]).

Une description du modèle TTD, ainsi que de deux généralisations appelées GTTD (*generalized TTD*, (Oseledets et Dolgov 2012)) et CTTD (*circular TTD*) ou TR (*tensor ring*, (Zhao *et al.* 2017)), est faite dans la section 5.2.8.

Le modèle GTTD étend le modèle TTD au cas des tenseurs d'ordre pair par paire d'indices, introduits dans la section 3.2. En ce qui concerne le modèle CTTD, il généralise le modèle TTD en remplaçant les matrices $\mathbf{A}^{(1)}$ et $\mathbf{A}^{(N)}$ par des tenseurs d'ordre trois $\mathcal{A}^{(1)} \in \mathbb{K}^{R_0 \times I_1 \times R_1}$ et $\mathcal{A}^{(N)} \in \mathbb{K}^{R_{N-1} \times I_N \times R_0}$, avec $R_0 = R_N \neq 1$. Ceci induit une contraction de ces deux tenseurs sur leur mode commun associé à l'indice r_0 , conduisant ainsi à une décomposition en un anneau de tenseurs (*tensor ring*) d'ordre trois, comme cela est illustré par la figure 5.2 pour un tenseur d'ordre quatre.

Noter que des algorithmes spécifiques basés sur un calcul de SVDs, appelés TT-SVD et TR-SVD, ont été proposés par (Oseledets 2011) et (Zhao *et al.* 2017), respectivement, pour estimer les modèles TTD et TR. Un autre algorithme basé sur la SVD, dénoté TT-HSVD, a été récemment développé pour calculer un modèle TTD à l'aide d'une procédure hiérarchique et parallèle (Zniyed *et al.* 2020b). Comme la décomposition de Tucker, la décomposition TTD n'est pas unique (voir l'équation [5.119]).

De même que la décomposition de Tucker hiérarchique (Hackbush et Kühn 2009 ; Grasedyck et Hackbush 2011 ; Ballani *et al.* 2013), la décomposition TTD est un cas particulier de réseaux tensoriels (Cichocki 2014).

La figure I.3 représente graphiquement le modèle TTD $[\mathbf{A}^{(1)}, \mathcal{A}^{(2)}, \mathcal{A}^{(3)}, \mathbf{A}^{(4)}]$ pour un tenseur $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times I_3 \times I_4}$ d'ordre quatre.

La décomposition TTD a été utilisée pour la résolution de systèmes d'équations linéaires (Oseledets et Dolgov 2012), pour le problème de complétion tensorielle (Grasedyck *et al.* 2015 ; Bengua *et al.* 2017), pour la classification de données vidéo (Yang *et al.* 2017), pour la reconnaissance faciale (Brandoni et Simoncini 2020), pour le développement d'une méthode PCA tensorielle robuste (Yang *et al.* 2020), et pour la modélisation de canaux de communication MIMO (Zniyed *et al.* 2020), parmi bien d'autres applications.

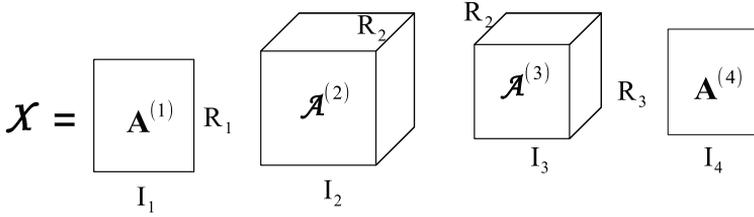


Figure I.3. Modèle TTD d'ordre quatre

À partir du tableau I.3, on note que le nombre de paramètres caractéristiques des décompositions CPD et TTD dépend linéairement de N , l'ordre du tenseur, tandis que la complexité paramétrique de la décomposition de Tucker croît exponentiellement avec N . Ceci justifie l'intérêt tout particulier porté aux deux premières décompositions pour des problèmes de grande dimension.

De ce qui précède, on peut conclure que le modèle TTD combine les avantages des deux autres décompositions, à savoir une faible complexité paramétrique, avec un nombre de paramètres linéairement proportionnel à l'ordre du tenseur, comme le modèle CPD, et l'existence d'un algorithme d'estimation stable, basé sur un calcul de SVDs, comme le modèle TD.

D'autre part, de cette brève description des trois modèles tensoriels, on peut remarquer que, contrairement aux matrices, la notion de rang n'est pas unique pour les tenseurs d'ordre supérieur à deux, puisqu'elle dépend de la décomposition utilisée. En effet, comme mentionné précédemment, on définit le rang canonique (appelé aussi rang de Kruskal) associé à la décomposition PARAFAC, le rang multilinéaire lié à la décomposition de Tucker, et les rangs TT déduits de la décomposition TT.

Un avantage des deux dernières décompositions par rapport à la première réside dans la possibilité d'estimer le rang multilinéaire et les rangs TT à l'aide de calculs de SVDs. D'autres notions de rang existent pour les tenseurs réels non-négatifs et les tenseurs symétriques (voir la section 5.2.5.2).

I.4.4. Illustration de la décomposition PARAFAC

Pour illustrer l'utilisation de la décomposition PARAFAC, nous considérons le cas des communications mobiles multi-utilisateurs, avec un système de codage CDMA (*code-division multiple access*). La technique d'accès multiple signifie que plusieurs émetteurs peuvent transmettre des informations simultanément sur un même canal de communication, avec un code attribué à chaque émetteur.

L'information est transmise sous forme de symboles $s_{n,m}$, avec $n \in \langle N \rangle$ et $m \in \langle M \rangle$, où N et M sont le nombre de slots temporels de transmission, c'est-à-dire, le nombre de périodes symboles, et le nombre d'antennes émettrices, respectivement. Les symboles appartiennent à un alphabet fini dépendant de la modulation appliquée aux signaux transmis. Ils sont codés à l'aide d'un codage spatio-temporel qui introduit une diversité de code du fait de la répétition P fois de chaque symbole, *via* un code $c_{p,m}$, attribué à la m -ième antenne émettrice, avec $p \in \langle P \rangle$, où P dénote la longueur d'étalement du code.

Considérons le cas sans bruit, et supposons que le canal de communication est à évanouissement plat (*flat fading*), de coefficients constants durant la transmission de N symboles. Le signal reçu par la k -ième antenne réceptrice, pendant la n -ième période symbole et la p -ième période *chip*, est une combinaison linéaire des symboles codés et transmis par les M antennes émettrices :

$$x_{k,n,p} = \sum_{m=1}^M h_{k,m} s_{n,m} c_{p,m} \quad [\text{I.1}]$$

où $h_{k,m}$ est le coefficient d'atténuation du canal entre les antennes réceptrice k et émettrice m . Les signaux reçus, à valeurs complexes, forment donc un tenseur d'ordre trois $\mathcal{X} \in \mathbb{C}^{K \times N \times P}$, avec pour modes : *espace* \times *temps* \times *code*, associés aux indices (k, n, p) , avec $k \in \langle K \rangle, n \in \langle N \rangle, p \in \langle P \rangle$. Ce tenseur de signaux satisfait une décomposition PARAFAC $[[\mathbf{H}, \mathbf{S}, \mathbf{C}; M]]$, dont le rang est égal au nombre M des antennes émettrices, et les facteurs matriciels sont le canal ($\mathbf{H} \in \mathbb{C}^{K \times M}$), la matrice de symboles transmis ($\mathbf{S} \in \mathbb{C}^{N \times M}$) et la matrice de codage ($\mathbf{C} \in \mathbb{C}^{P \times M}$).

Sous certaines conditions liées au critère de Kruskal (voir [5.86]), le modèle du système est essentiellement unique. En supposant les paramètres (K, N, P, M) connus, les matrices $(\mathbf{H}, \mathbf{S}, \mathbf{C})$ peuvent être estimées en utilisant l'algorithme ALS. Cet exemple constitue une forme simplifiée du système DS-CDMA (*direct-sequence CDMA*) proposé par (Sidiropoulos *et al.* 2000b).

1.5. Avec quelles fonctions de coût et quels algorithmes d'optimisation ?

Dans cette section, nous introduisons tout d'abord les principaux types de traitement à mettre en œuvre avec des tenseurs. Puis, nous présentons quelques exemples de fonctions de coût à minimiser et décrivons brièvement les principaux algorithmes d'optimisation utilisés.

1.5.1. Prétraitements et traitements tensoriels

Il est important tout d'abord de lister les prétraitements à effectuer. Il s'agit principalement d'opérations de centrage des données (élimination d'offsets), de mise à

l'échelle (*scaling*) de données non homogènes, de suppression de données aberrantes (*outliers*) et d'artefacts, d'ajustement (luminosité, contraste, alignement, etc.) d'image, de débruitage, de transformation des signaux à l'aide de certaines transformées (ondelettes, Fourier, etc.), ou encore du calcul éventuel de statistiques des signaux à traiter.

Les prétraitements sont fondamentaux tant pour améliorer la qualité des modèles estimés et donc des traitements qui en découlent, que pour éviter des problèmes numériques au niveau des algorithmes d'optimisation, comme par exemple des problèmes de conditionnement pouvant entraîner des difficultés de convergence des algorithmes.

La complexité des prétraitements de centrage et de mise à l'échelle résulte de leur interdépendance, avec différentes manières de les combiner. De plus, avec des données manquantes, le centrage peut entraîner une réduction du rang du modèle tensoriel. Pour une description détaillée de ces prétraitements, consulter (Smilde *et al.* 2004).

En ce qui concerne les traitements proprement dits, on peut distinguer différentes classes de traitement :

- supervisé/non supervisé (aveugle ou semi-aveugle), c'est-à-dire, avec l'utilisation ou non de données d'apprentissage, comme c'est le cas par exemple pour la résolution d'un problème de classification, ou d'information *a priori* à transmettre au récepteur, appelée séquence pilote, pour l'identification d'un canal de communication ;
- temps réel (en ligne)/temps différé (hors ligne) ;
- centralisé/distribué ;
- adaptatif/par bloc (vis-à-vis des données) ;
- avec/sans couplage de modèles tensoriels et/ou matriciels ;
- avec/sans données manquantes.

Il est important de distinguer les traitements en temps différé pour une analyse de données enregistrées sous forme de signaux et d'images, des traitements en temps réel tels que ceux nécessités par les systèmes de communication sans fil, les systèmes de recommandation, les recherches sur le Web, ou encore les réseaux sociaux. Pour les applications en temps réel, la dimensionalité des modèles et la complexité algorithmique sont des facteurs prépondérants. Les signaux reçus par des antennes, les échanges d'information entre système et utilisateurs, de même que les échanges de messages entre personnes d'un réseau social sont dépendants du temps, les données pouvant être stockées sous forme de graphe de connaissances.

Ainsi, par exemple, un système de recommandation interagit en temps réel avec les utilisateurs, *via* l'extension d'une base de données existante en utilisant des techniques

de *machine learning*. Pour une description de différentes applications de tenseurs couplés avec le *machine learning*, consulter (Anandkumar *et al.* 2014 ; Sidiropoulos *et al.* 2017).

1.5.2. Fonctions de coût et algorithmes d'optimisation

Les traitements mentionnés ci-avant induisent divers types d'algorithme d'optimisation :

- avec ou sans contraintes ;
- itératif/non itératif, ou forme fermée (*closed form*) ;
- alterné/global ;
- séquentiel/parallèle.

Selon l'information dont on dispose *a priori*, différents types de contraintes peuvent être pris en compte dans la fonction de coût à optimiser : rang faible, parcimonie, non-négativité, orthogonalité, différentiabilité/régularité (*smoothness*).

Dans le cas d'une optimisation avec contraintes, des pondérations doivent être choisies dans la fonction de coût à optimiser selon l'importance relative donnée à chaque contrainte, et selon la qualité de l'information *a priori* dont on dispose.

Données		$\mathcal{X} \in \mathbb{K}^{I \times J \times K}, \mathbf{Y} \in \mathbb{K}^{I \times M}$
Problèmes	Estimation	Fonctions de coût
	CPD	$f(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \ \mathcal{X} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C}; R \rrbracket\ _F^2$
	TD	$f(\mathcal{G}, \mathbf{A}, \mathbf{B}, \mathbf{C}) = \ \mathcal{X} - \llbracket \mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\ _F^2$
	CMTucker	$f(\mathcal{G}, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{U}) = \ \mathcal{X} - \llbracket \mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\ _F^2 + \ \mathbf{Y} - \mathbf{A}\mathbf{U}^T\ _F^2$
	CSTF	$f(\mathcal{G}, \mathbf{W}, \mathbf{H}, \mathbf{S}) = \ \mathcal{Y} - \llbracket \mathcal{G}; \mathbf{W}^*, \mathbf{H}^*, \mathbf{S} \rrbracket\ _F^2 + \ \mathcal{Z} - \llbracket \mathcal{G}; \mathbf{W}, \mathbf{H}, \mathbf{S}^* \rrbracket\ _F^2 + \lambda \ \mathcal{G}\ _1$
	Imputation	Fonctions de coût
	CPD	$f_{\mathcal{W}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \ \mathcal{W} \odot (\mathcal{X} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C}; R \rrbracket)\ _F^2$
	TD	$f_{\mathcal{W}}(\mathcal{G}, \mathbf{A}, \mathbf{B}, \mathbf{C}) = \ \mathcal{W} \odot (\mathcal{X} - \llbracket \mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket)\ _F^2$
	Imputation avec contrainte de rang faible	Fonctions de coût
	CPD	$f_{\mathcal{W}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \ \mathcal{W} \odot (\mathcal{X} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C}; R \rrbracket)\ _F^2 + \lambda \ \mathcal{X}\ _*$
	TD	$f_{\mathcal{W}}(\mathcal{G}, \mathbf{A}, \mathbf{B}, \mathbf{C}) = \ \mathcal{W} \odot (\mathcal{X} - \llbracket \mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket)\ _F^2 + \lambda \ \mathcal{X}\ _*$

Tableau I.4. Fonctions de coût pour l'estimation de modèles et la restauration de données manquantes

Dans le tableau I.4, nous présentons quelques exemples de fonctions de coût à minimiser pour l'estimation paramétrique de certains modèles tensoriels d'ordre trois (CPD, TD, CMTucker (*coupled matrix-Tucker*) et CSTF (*coupled sparse tensor factorization*)), pour l'imputation de données manquantes dans un tenseur de données, et pour l'estimation d'un tenseur creux de données, avec une contrainte du type rang faible, prise en compte sous la forme de la norme nucléaire du tenseur de données.

On peut faire les remarques suivantes.

- Les fonctions de coût présentées dans le tableau I.4 correspondent à des critères d'ajustement aux données (on dit aussi d'attache aux données). Ces critères, exprimés en termes de normes de Frobenius tensorielle et matricielle ($\|\cdot\|_F$), sont quadratiques par rapport aux écarts entre le tenseur de données \mathcal{X} et la sortie des modèles CPD et TD, d'une part, et entre la matrice de données \mathbf{Y} et le modèle de factorisation matricielle, dans le cas du modèle CMTucker, d'autre part. Ils sont trilineaire et quadrilinéaire par rapport aux paramètres des modèles CPD et TD à estimer, respectivement, et bilinéaire par rapport aux paramètres du modèle de factorisation matricielle.

- Dans le cas du modèle CMTucker, le couplage considéré ici intervient au niveau des premiers modes du tenseur \mathcal{X} et de la matrice \mathbf{Y} de données, *via* le facteur matriciel commun \mathbf{A} .

- Des modèles de factorisations matricielle et tensorielle couplés (CMTF pour *coupled matrix and tensor factorization*) ont été introduits par (Acar *et al.* 2011b) en couplant un modèle CPD avec une factorisation matricielle, et en proposant l'algorithme du gradient pour l'estimation des paramètres. Ce type de modèle a été exploité par (Acar *et al.* 2017) pour fusionner des données EEG et fMRI dans le but d'analyser l'activité du cerveau. Les signaux EEG sont modélisés à l'aide d'un modèle CPD normalisé (voir la section 5.2.5.5), tandis que les données fMRI sont modélisées à l'aide d'une factorisation matricielle, ces données étant couplées par le mode *subjects* (voir le tableau I.1).

La fonction de coût à minimiser est alors donnée par :

$$f(\mathbf{g}, \Sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{U}) = \left\| \mathcal{X} - \llbracket \mathbf{g}; \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket \right\|_F^2 + \left\| \mathbf{Y} - \mathbf{A} \Sigma \mathbf{U}^T \right\|_F^2 + \alpha \|\mathbf{g}\|_1 + \alpha \|\boldsymbol{\sigma}\|_1 \quad [\text{I.2}]$$

où les vecteurs colonnes des matrices facteurs (\mathbf{A} , \mathbf{B} , \mathbf{C}) sont de norme unité, Σ est une matrice diagonale dont les éléments diagonaux sont les coefficients du vecteur $\boldsymbol{\sigma}$, et $\alpha > 0$ est un paramètre de pénalité permettant de donner plus ou moins d'importance aux contraintes de parcimonie sur les vecteurs de poids (\mathbf{g} , $\boldsymbol{\sigma}$) modélisées à l'aide de la norme l_1 .

L'intérêt de fusionner des données EEG et fMRI *via* le critère [I.2] réside dans la complémentarité des méthodes d'acquisition et d'observation en termes de résolution, les signaux EEG étant caractérisés par une grande résolution temporelle mais une faible résolution spatiale, tandis que l'imagerie fMRI fournit une haute résolution spatiale.

– Dans le cas du modèle CSTF (*coupled sparse tensor factorization*) de factorisations tensorielles couplées parcimonieuses (Li *et al.* 2018), le tenseur d'images hyperspectrales haute résolution (HR-HSIs pour *high-resolution hyperspectral images*) est représenté à l'aide d'un modèle de Tucker d'ordre trois dont le cœur est creux ($\mathcal{X} = \mathcal{G} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S} \in \mathbb{R}^{M_w \times M_h \times M_s}$), avec pour modes : *espace (width)* \times *espace (height)* \times *bandes spectrales*. Les matrices $\mathbf{W} \in \mathbb{R}^{M_w \times n_w}$, $\mathbf{H} \in \mathbb{R}^{M_h \times n_h}$ et $\mathbf{S} \in \mathbb{R}^{M_s \times n_s}$ dénotent les dictionnaires relatifs aux modes largeur (w), hauteur (h) et spectre (s), composés de n_w , n_h et n_s atomes, respectivement.

Le tenseur cœur $\mathcal{G} \in \mathbb{R}^{n_w \times n_h \times n_s}$ contient les coefficients relativement aux trois dictionnaires. Le tenseur $\mathcal{Y} \in \mathbb{R}^{m_w \times m_h \times m_s}$ représente l'image hyperspectrale enregistrée, et correspond à une version spatialement sous-échantillonnée de \mathcal{X} , c'est-à-dire, telle que $m_w \ll M_w$ et $m_h \ll M_h$. Le tenseur $\mathcal{Z} \in \mathbb{R}^{M_w \times M_h \times m_s}$ représente l'image multispectrale de la même scène que \mathcal{Y} , et correspond à une version spectralement sous-échantillonnée de \mathcal{X} , c'est-à-dire, telle que $m_s \ll M_s$.

Les matrices \mathbf{W}^* , \mathbf{H}^* et \mathbf{S}^* sont des versions sous-échantillonnées spatialement et spectralement de \mathbf{W} , \mathbf{H} et \mathbf{S} , respectivement. λ est un paramètre de régularisation vis-à-vis de la contrainte de parcimonie sur le tenseur cœur, exprimée en termes de la norme l_1 de \mathcal{G} . L'objectif de la fusion est d'estimer \mathcal{X} à partir des images observées \mathcal{Y} et \mathcal{Z} .

– Pour le problème d'imputation de données manquantes à l'aide d'un modèle du type CPD ou TD, le tenseur binaire \mathcal{W} , de même dimension que \mathcal{X} , est défini comme :

$$w_{ijk} = \begin{cases} 1 & \text{si } x_{ijk} \text{ est connu} \\ 0 & \text{si } x_{ijk} \text{ est manquant} \end{cases} \quad [\text{I.3}]$$

Le produit de Hadamard, noté \odot , de \mathcal{W} avec l'écart entre \mathcal{X} et la sortie des modèles CPD et TD, a pour but d'ajuster le modèle aux seules données disponibles, et donc d'ignorer les données manquantes pour l'estimation du modèle.

Ce problème d'imputation, appelé complétion tensorielle, a été originellement traité par (Tomasi et Bro 2005 ; Acar *et al.* 2011a) en utilisant un modèle CPD, puis par (Filipovic et Jukic 2015) à l'aide d'un modèle TD. De nombreux articles traitent ce problème dans le contexte de différentes applications. Un survol bibliographique sera effectué dans le volume 3.

– Pour le problème d'imputation avec la contrainte de rang faible, la norme nucléaire de \mathcal{X} , dénotée $\|\mathcal{X}\|_*$, remplace la contrainte sur le rang dans la fonction de coût, car la fonction $\text{rang}(\mathcal{X})$ n'est pas convexe, et la norme nucléaire⁷ est l'approximation convexe la plus proche du rang du tenseur. Dans (Liu *et al.* 2013), $\|\mathcal{X}\|_*$ est remplacé par $\sum_{n=1}^3 \lambda_n \|\mathbf{X}_n\|_*$ où \mathbf{X}_n est le dépliement mode- n de \mathcal{X} .

La minimisation des critères donnés dans le tableau I.4 peut être réalisée globalement à l'aide d'une méthode d'optimisation non linéaire comme, par exemple, un algorithme de descente de type gradient (à pas fixe ou optimal), ou les algorithmes de Gauss-Newton ou de Levenberg-Marquardt, ce dernier algorithme pouvant être vu comme une forme régularisée de l'algorithme de Gauss-Newton. Dans le cas d'une optimisation avec contraintes, la méthode du Lagrangien augmenté est très souvent utilisée, car elle permet de transformer le problème d'optimisation avec contraintes en une suite de problèmes d'optimisation sans contrainte.

Ces méthodes d'optimisation présentent les inconvénients d'une lenteur de convergence pour les algorithmes de type gradient, et d'une grande complexité numérique pour les algorithmes de Gauss-Newton et de Levenberg-Marquardt, en raison de la nécessité de calculer la matrice jacobienne du critère par rapport aux paramètres à estimer et l'inverse d'une matrice de grande dimension.

Pour ces raisons, il est très souvent fait appel à des méthodes d'optimisation alternée consistant à remplacer une optimisation globale vis-à-vis de l'ensemble des facteurs matriciels et tensoriels à estimer, par une séquence d'optimisations séparées de critères linéaires par rapport à chaque facteur inconnu, les autres facteurs étant fixés avec leurs valeurs estimées dans les itérations précédentes. C'est le cas de l'algorithme ALS (*alternating least-squares*) qui sera détaillé dans le chapitre 5 pour l'estimation d'un modèle PARAFAC. Dans le cas d'une optimisation sous contrainte, la méthode couramment utilisée est l'algorithme des directions alternées (ADA, ou ADMM pour *alternating direction method of multipliers*) (Boyd *et al.* 2011).

I.6. Brève description du contenu

Pour compléter cette introduction, on peut mettre en avant les principales connaissances de base à acquérir en vue d'utiliser les outils tensoriels, des connaissances qui font en grande partie l'objet de ce volume 2 :

7. Voir les équations [1.65] et [1.67] de la norme de Frobenius et de la norme nucléaire d'une matrice ; pour un tenseur, voir la section 3.16.

- les opérations de rangement, dites aussi de reformatage (*reshaping*), des données d'un tenseur sous forme d'un vecteur (vectorisation), d'une matrice (matricisation), ou d'un tenseur d'ordre réduit, *via* des combinaisons de modes ;
- les opérations de tensorisation et d'Hankelisation qui permettent de construire des tenseurs à partir de données contenues dans des vecteurs ou des matrices de grande dimension ;
- les opérations tensorielles telles que transposition, symétrisation, multiplication, produits de Hadamard et de Kronecker, inversion, et pseudo-inversion ;
- les notions de valeur propre et de valeur singulière d'un tenseur ;
- les décompositions et modèles tensoriels, avec leurs propriétés d'unicité ;
- l'algorithmique pour la résolution des problèmes de réduction de dimensionnalité et donc de meilleure approximation de rang faible, d'estimation paramétrique, et d'imputation de données manquantes. Cet aspect algorithmique lié aux tenseurs sera approfondi dans le volume 3.

Les opérations et les décompositions tensorielles faisant largement appel aux outils matriciels, nous commencerons par rappeler quelques décompositions matricielles dans le chapitre 1, en détaillant plus particulièrement les décompositions à l'aide des valeurs propres et des valeurs singulières (EVD et SVD pour *eigenvalue decomposition* et *singular value decomposition*), ainsi que quelques-unes de leurs applications.

Les produits matriciels de Hadamard, de Kronecker et de Khatri-Rao qui jouent un rôle très important pour l'exploitation des décompositions tensorielles, seront présentés de façon détaillée dans le chapitre 2. De nombreuses propriétés de chacun de ces produits ainsi que certaines relations les liant seront mises en évidence.

Ces opérations seront illustrées *via* leur utilisation pour la représentation de dérivées partielles du premier ordre d'une fonction et pour la résolution de certaines équations matricielles telles que, par exemple, les équations de Sylvester et de Lyapunov.

L'estimation des matrices facteurs de produits de Khatri-Rao et de Kronecker, simples puis multiples, sera résolue à l'aide de solutions en forme fermée (*closed form*), c'est-à-dire, non itératives, basées sur des approximations de rang un de matrices et de tenseurs d'ordre élevé.

Dans ce même chapitre, nous introduirons une convention indiciaire très utile pour le calcul tensoriel. Cette convention qui généralise la convention de sommation d'Einstein (Pollock 2011), sera utilisée pour représenter différents produits matriciels, pour démontrer des formules de vectorisation de produits matriciels ainsi que différentes relations entre les produits de Kronecker, de Khatri-Rao et de Hadamard. Dans

le chapitre 3, elle sera utilisée de façon originale pour la matricisation et la vectorisation d'un tenseur, et dans le chapitre 5, pour l'établissement des formes matricisées des décompositions de Tucker et PARAFAC.

Dans le chapitre 3, nous présenterons différents ensembles de tenseurs, avant d'introduire les notions de tranches matricielles et tensorielles, et de combinaison de modes, à la base des opérations de rangement des données. Puis, les principales opérations tensorielles listées ci-avant seront décrites. Des liens entre produits tensoriels et systèmes d'équations tensorielles seront aussi mis en évidence, avec la résolution de certains systèmes à l'aide de la méthode des moindres carrés.

Le chapitre 4 sera consacré à une introduction des notions de valeur propre et de valeur singulière pour les tenseurs d'ordre élevé. Les tenseurs définis positifs/négatifs et les tenseurs orthogonalement/unitairement semblables seront introduits. Le problème de la meilleure approximation de rang un d'un tenseur sera aussi considéré.

Dans le chapitre 5, nous ferons une présentation détaillée de différentes décompositions tensorielles, avec un poids tout particulier donné aux décompositions basiques de Tucker et CPD, qui peuvent être vues comme des généralisations de la SVD matricielle aux tenseurs d'ordre supérieur à deux. Différentes manières d'écrire ces décompositions seront considérées (formes scalaires, matricisées, vectorisées, à l'aide de produits mode- n et de produits extérieurs), et un poids tout particulier sera donné aux tenseurs d'ordre trois. Des modèles tensoriels en blocs et des modèles tensoriels avec contraintes seront aussi décrits, comme les décompositions CONFAC (*constrained factorization*) et BTD (*block term decomposition*).

Le modèle TTD dont le but est de réduire la complexité paramétrique de la représentation tensorielle à partir de la contraction deux à deux d'une séquence de tenseurs d'ordres deux et trois, sera aussi présenté. Deux extensions du modèle TTD correspondant aux modèles GTTD et CTTD, appelé *tensor ring* par (Zhou *et al.* 2017), seront également décrites.

Les algorithmes HOSVD et ALS seront détaillés pour l'estimation d'un modèle de Tucker et d'un modèle CPD, respectivement. Les propriétés d'unicité des décompositions de Tucker et CPD, ainsi que différentes notions de rang d'un tenseur, seront présentées. Ce chapitre sera conclu par des illustrations des décompositions BTD et CPD pour la modélisation tensorielle d'harmoniques multidimensionnelles, la séparation de sources dans un mélange linéaire instantané, ainsi que pour la modélisation et l'estimation d'un système linéaire à réponse impulsionnelle finie (RIF), à l'aide d'un tenseur de cumulants d'ordre quatre de la sortie du système.

Les cumulants d'ordre élevé de signaux aléatoires, qui peuvent être vus comme des tenseurs, jouent un rôle central dans de nombreuses applications de traitement du

signal, comme cela sera illustré dans le chapitre 5. Ceci nous a motivé à écrire une annexe présentant un bref survol de quelques résultats fondamentaux concernant les statistiques d'ordre élevé (HOS pour *higher-order statistics*) de signaux aléatoires, avec deux applications des HOS pour l'estimation d'un système linéaire invariant dans le temps, et d'un système quadratique homogène.