Tensor Decomposition Methods for Denoising Dynamic Contrast-Enhanced Ultrasound

Çalış, M.

DOI
10.4233/uuid:07c735ac-c0ac-4057-8d97-b2873a465686

Publication date
2026

Document Version
Final published version

Citation (APA)
Çalış, M. (2026). *Tensor Decomposition Methods for Denoising Dynamic Contrast-Enhanced Ultrasound*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:07c735ac-c0ac-4057-8d97-b2873a465686

Tensor Decomposition Methods for Denoising Contrast-Enhanced Ultrasound

METİN ÇALIŞ

Tensor Decomposition Methods for Denoising Dynamic Contrast-Enhanced Ultrasound

Tensor Decomposition Methods for Denoising Dynamic Contrast-Enhanced Ultrasound

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Monday 12 January 2026 at 15:00 o'clock

by

Metin ÇALIŞ

Master of Science in Electrical Engineering,
Delft University of Technology, Netherlands
born in Ankara, Türkiye

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. Dr. Ir. A.J. van der Veen,	Delft University of Technology, *promotor*
Prof. Dr. M. Mischi,	Eindhoven University of Technology, *promotor*
Dr. B. Hunyadi,	Delft University of Technology, *promotor*

Independent members:

Prof. Dr. L. De Lathauwer,	KU Leuven
Prof. Dr. Ir. G.J.T. Leus,	Delft University of Technology
Prof. Dr. Ir. H. Wijkstra,	Eindhoven University of Technology
Dr. Ir. K. Batselier,	Delft University of Technology
Prof. Dr. P.J. French,	Delft University of Technology, reserve member

The research in this dissertation was in part funded by Angiogenesis Analytics and by Holland High Tech with a Publieke-Privaat Samenwerking [Public-Private Partnership (PPS)] supplement for Research and Development in the Topsector High Tech Systemen en Materialen [High Tech Systems and Materials (HTSM)] under Project 20.091. Angiogenesis Analytics also provided in-kind resources and datasets.

Keywords: Dynamic Contrast-Enhanced Ultrasound, Tensor Decomposition, Despeckling, Multilinear Singular Value Decomposition, Tensor Singular Value Decomposition, Rank estimation, Performance Bounds

ISBN 978-94-6518-143-1

An electronic copy of this dissertation is available at
https://repository.tudelft.nl/.

To my family

CONTENTS

SUMMARY

Prostate cancer is the most common malignancy among men. To confirm an initial detection by a prostate-specific antigen test, magnetic resonance imaging (MRI) is used, but MRI is costly. Ultrasound is a cost-effective imaging modality and shows promising results in diagnosing prostate cancer, especially the dynamic contrast-enhanced ultrasound. Dynamic contrast-enhanced ultrasound (DCEUS) is an imaging modality that allows the imaging of the injected microbubbles by exploiting their non-linear acoustic scatter. Because of their size, comparable to red blood cells, ultrasound contrast agents can flow through the vascular tree down to the microvessels, enabling the visualization and, possibly, quantification of the angiogenic processes associated with cancer growth. Although several techniques are applied to DCEUS to reduce noise, speckle noise still exists.

Speckle noise occurs due to the coherent imaging of many microbubbles in one resolution cell. In existing methods, a low-rank matrix decomposition is applied to the DCEUS acquisitions using singular value decomposition, and the despeckling is done by keeping the highest few singular vectors and values. The DCEUS acquisitions come in a tensor format, rich with higher-order structure. The application of the matrix-based denoising technique does not utilize the original tensor structure.

This dissertation focuses on despeckling through higher-order tensor decomposition methods. We tackle the following research question: "How can low-rank tensor decomposition methods be leveraged to effectively denoise DCEUS acquisitions of the prostate for improved prostate cancer detection?" We apply tensor decomposition methods that utilize orthogonal factors. In the spatial domain, the orthogonality allows the separation of the malignant and benign regions and the separation of the tissue and the vasculature. In the time domain, the orthogonality allows for capturing the components that correspond with the bubble movement and rejects the components related to the noise.

Despeckling of DCEUS through low-rank tensor decomposition has not been conducted before, and we propose tensor estimation algorithms for this application by utilizing established tensor decomposition frameworks. We start our research by modeling speckle noise as white Gaussian noise (WGN) with sparse outliers. We assess the performance of convex tensor estimation algorithms through simulations. We propose a novel weighting scheme for the soft-thresholding of the singular values. Instead of iterative thresholding, we can truncate the tensor and de-

speckle the DCEUS acquisitions. We propose a rank estimation method for DCEUS acquisitions. Instead of modeling speckle noise as WGN with sparse outliers, we minimize its negative log-likelihood and propose a gradient-based denoising algorithm.

Next, we investigate the classification performance of prostate cancer by comparing the proposed algorithms with the literature. We use the area under the receiver-operator characteristic curve (ROC-AUC) metric to assess the classification performance. For the voxel-based cancer diagnosis of 94 prostate cancer patients, truncated multilinear singular value decomposition has a better performance for the majority of the prostate cancer markers when the ROC-AUC metric is used. A rank estimation technique incorporating WGN with sparse outliers, followed by truncated multilinear singular value decomposition (tr-MLSVD), is the best-performing denoising method for DCEUS. In the context of the main research question, the cancer diagnosis performance of DCEUS acquisitions improves the majority of the time when a tensor-based denoising technique is used. On average, the tensor-based denoising techniques yield approximately a 1.6% relative improvement in the ROC-AUC metric compared to the literature. This translates to billions of additional correct voxel-level malignancy discriminations in our clinical study, which may significantly impact downstream classification and localization performance.

We conclude with a theoretical study on the lower bound of the tensor decomposition method that performs the best for despeckling DCEUS. We calculate a lower bound for estimating the components of MLSVD when the ranks are known. In general, the CCRB that lower bounds the variance of the unbiased estimates of the components of MLSVD does not exist due to the non-uniqueness of the decomposition. However, when the mode-n singular values are unique, the CCRB exists. Additionally, when the multilinear ranks are high and modal singular values are well-separated, it is a tight bound. Such cases do not typically occur with real data such as DCEUS, highlighting the limited modeling capability of the CCRB.

SAMENVATTING

Prostaatkanker is de meest voorkomende vorm van kanker bij mannen. Om een eerste detectie via een prostaat-specifieke antigeentest te bevestigen, wordt vaak gebruikgemaakt van magnetische resonantiebeeldvorming (MRI), maar MRI is duur. Echografie is een kosteneffectieve beeldvormingstechniek en toont veelbelovende resultaten bij de diagnose van prostaatkanker, met name de dynamisch contrast-versterkte echografie. Dynamisch contrast-versterkte echografie (DCEUS) is een beeldvormingstechniek waarmee geïnjecteerde microbubbels kunnen worden afgebeeld door gebruik te maken van hun niet-lineaire akoestische verstrooiing. Door hun grootte, vergelijkbaar met rode bloedcellen, kunnen contrastmiddelen voor echografie door de bloedvaten tot in de microvaten stromen. Dit maakt het mogelijk om de angiogeneseprocessen die met kanker gepaard gaan, zichtbaar te maken en mogelijk ook te kwantificeren. Hoewel verschillende technieken worden toegepast om ruis in DCEUS te verminderen, blijft speckleruis aanwezig.

Speckleruis ontstaat door de coherente beeldvorming van veel microbubbels binnen één resolutiecel. Huidige technieken passen een matrix-decompositie met lage rang wordt toegepast op de DCEUS-opnamen via de singuliere-waarde-decompositie (SVD) toe, waarbij het ont-specklen gebeurt door alleen de hoogste singuliere waarden en vectoren te behouden. De DCEUS-data worden vastgelegd in tensorformaat, rijk aan hogere-orde structuren. Het toepassen van de op matrices gebaseerde denoising-techniek benut deze oorspronkelijke tensorstructuur niet.

Dit proefschrift richt zich op het ont-specklen van DCEUS via hogere-orde tensor-decompositie methoden. We behandelen de volgende onderzoeksvraag: "Hoe kunnen lage-rang tensor-decompositiemethoden worden ingezet om DCEUS-opnamen van de prostaat effectief te denoisen ter verbetering van de prostaatkankerdiagnose?"We gebruiken tensor-decompositiemethoden die orthogonale factoren benutten. In het ruimtelijke domein maakt orthogonaliteit het mogelijk om maligne en benigne gebieden te scheiden, evenals het scheiden van weefsel en vasculatuur. In het temporele domein maakt orthogonaliteit het mogelijk om de componenten te isoleren die verband houden met de beweging van de microbubbels, terwijl de componenten die verband houden met ruis worden onderdrukt.

Ont-specklen van DCEUS met behulp van lage-rang tensor-decompositie is niet eerder uitgevoerd, en wij stellen tensor-schatting algoritmes voor deze toepassing voor, gebaseerd op bestaande tensor-decompositie raam-

werken. We beginnen ons onderzoek door speckleruis te modelleren als witte Gaussische ruis (WGN) met sporadische uitschieters. De prestaties van convex tensor-schatting algoritmes worden geëvalueerd via simulaties. We stellen een nieuw weging-schema voor de soft-thresholding van de singuliere waarden voor. In plaats van iteratief thresholden, kunnen we de tensor afkappen en de DCEUS-opnamen direct ont-specklen. Ook stellen we een rangschattingsmethode voor DCEUS-opnamen voor. In plaats van speckleruis als WGN met sporadische uitschieters te modelleren, minimaliseren we de negatieve log-likelihood en stellen we een op gradiënt gebaseerde denoising-algoritme voor.

Vervolgens onderzoeken we de classificatieprestaties voor prostaatkanker door de voorgestelde algoritmes te vergelijken met methoden uit de literatuur. We gebruiken de area under the receiver-operator characteristic curve (ROC-AUC) als maatstaf voor de classificatieprestaties. Voor de voxel-gebaseerde kankerdiagnose van 94 prostaatkankerpatiënten levert de getrunceerde multilineaire singuliere-waarde-decompositie betere prestaties voor de meerderheid van de prostaatkanker-biomarkers wanneer de ROC-AUC-metriek wordt toegepast. Een rangschattingsmethode die WGN met sporadische uitschieters combineert, gevolgd door getrunceerde multilineaire SVD (tr-MLSVD), blijkt de best presterende denoising-methode voor DCEUS. In de context van de hoofdonderzoeksvraag blijkt dat de diagnostische prestaties van DCEUS-opnamen meestal verbeteren wanneer een tensor-gebaseerde denoising-techniek wordt toegepast. Gemiddeld gezien leveren de tensor gebaseerde denoising technieken een relatieve verbetering van ongeveer 1.6% op in de ROC AUC metriek ten opzichte van bestaande methoden uit de literatuur. Dit komt overeen met tientallen miljarden extra correcte discriminaties op voxel niveau van maligniteiten in onze klinische studie, wat mogelijk leidt tot significante verbeteringen in de daaropvolgende classificatie en localisatie prestaties.

We sluiten af met een theoretische studie naar de ondergrens van de tensor-decompositiemethoden die het beste presteren voor het ontspecklen van DCEUS. We berekenen een ondergrens voor het schatten van de componenten van MLSVD wanneer de rangen bekend zijn. In het algemeen bestaat de constrained Cramér-Rao Bound (CCRB), die de variantie van de onvertekende schattingen van de componenten van MLSVD begrenst, niet vanwege de niet-uniciteit van de decompositie. Echter, wanneer de mode-n singuliere waarden uniek zijn, bestaat de CCRB wel. Bovendien is deze begrenzing strak wanneer de multilineaire rangen hoog zijn en de singuliere waarden goed gescheiden zijn. Dergelijke gevallen doen zich echter zelden voor in echte data zoals DCEUS, wat de beperkte modelleringscapaciteit van de CCRB benadrukt.

1

INTRODUCTION

1.1. RESEARCH MOTIVATION

Prostate cancer is the most commonly diagnosed cancer among men and presents a growing public health concern worldwide. Early and accurate diagnosis is critical in improving patient outcomes. Conventional diagnostic approaches, such as MRI or biopsy, are either expensive, invasive, or uncomfortable for the patient. This has led to a growing interest in using ultrasound imaging as a diagnostic modality due to its real-time capability, non-invasive nature, and cost-effectiveness.

Dynamic Contrast-Enhanced Ultrasound (DCEUS) is a promising imaging technique that enhances vascular visualization through the intravenous injection of microbubble contrast agents. These microbubbles interact non-linearly with the incoming ultrasound waves, producing harmonic signals that can be selectively captured. The result is a dynamic, 4th-order dataset that captures the time evolution of a volume with valuable information about microbubble perfusion, dispersion, and corresponding vascular characteristics, crucial markers in distinguishing between malignant and benign prostate regions.

However, DCEUS is susceptible to speckle noise, which complicates image interpretation and impairs diagnostic performance. Traditional denoising approaches often flatten the data onto a spatio-temporal matrix, discarding the rich temporal and spatial structure present in the full acquisition. Tensor-based methods offer a promising alternative, as they preserve the higher-order structure of DCEUS data and allow for more nuanced low-rank approximations. This dissertation is motivated by the need to develop such advanced, structure-aware denoising techniques to enhance DCEUS utility in prostate cancer diagnosis. This motivation stems from the initial application of matrix-based orthogonal decompositions to flattened DCEUS acquisitions. In this dissertation, we extend this approach by exploring despeckling through tensor-based orthogonal decompositions, which better preserve the multidimensional structure of

the data.

1.2. RESEARCH OBJECTIVES

This dissertation aims to improve prostate cancer diagnosis by enhancing the quality of DCEUS acquisitions through tensorial structure-preserving denoising techniques. The main question driving this research is "*How can low-rank tensor decomposition methods be leveraged to effectively denoise dynamic contrast-enhanced ultrasound recordings of the prostate for improved prostate cancer detection?*" We focused on orthogonal decompositions. Multiple tensor decomposition frameworks can extract orthonormal factor matrices with their corresponding singular values. In addition, these decompositions do not readily incorporate different noise types. Such distinctions create the themes that this dissertation aims to tackle. The first theme is the analysis of different orthogonal tensor decomposition frameworks for denoising speckle noise. The other theme is investigating ways to incorporate the probability distribution of the noise type into the tensor decomposition. Specifically, the research aims to answer the following research questions to answer the overarching question:

Q1: Under which signal-to-noise ratios can the speckle noise be modeled as white Gaussian noise with sparse outliers?

The success of denoising algorithms hinges on accurately modeling the noise. Identifying the regimes in which speckle noise resembles a mixture of white Gaussian noise and sparse outliers enables the design of more effective denoising techniques.

Q2: Can we improve the prostate cancer classification performance using tensor-based denoising techniques applied to DCEUS?

Tensor rank can be defined in multiple ways, each affecting the structure and performance of decomposition algorithms. Determining the most suitable definition for DCEUS helps ensure that denoising retains diagnostically relevant features.

Q3: Can a lower bound on the variance of locally unbiased estimators be derived in the context of low multilinear rank tensor estimation?

Establishing theoretical lower bounds, such as constrained Cramér-Rao bounds, provides valuable insights into the ultimate limits of denoising accuracy, serving as benchmarks for algorithm performance. Further, whether DCEUS naturally exhibits these characteristics determines whether the CCRB result can be used.

To better position this work, we present the bigger picture in prostate cancer diagnosis. Fig. 1.1 shows the clinical trial path and the algorithmic

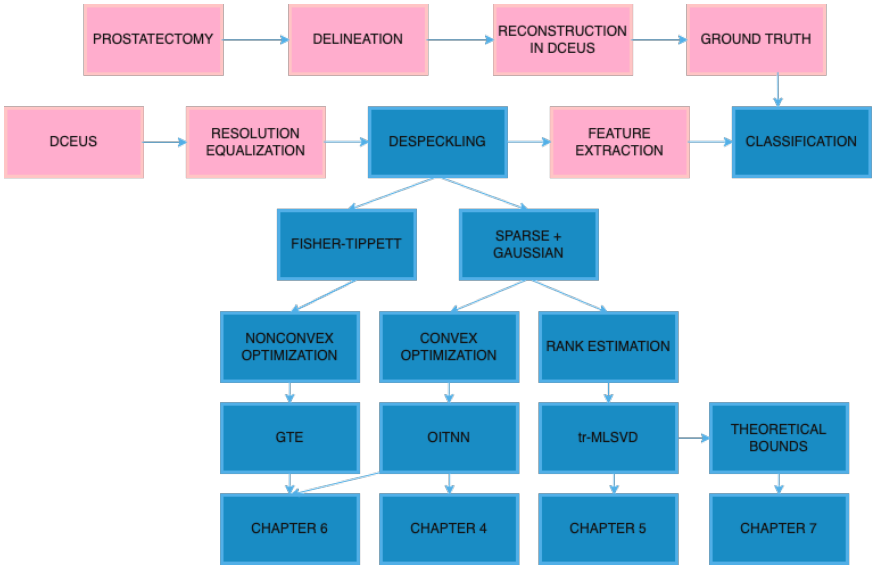


Figure 1.1: The clinical trial path and the algorithmic process related to DCEUS. The scope of this dissertation is shown in blue.

process related to DCEUS. The clinical trial path is related to obtaining a prostate cancer ground truth that matches the patient's DCEUS acquisitions. The algorithmic process is related to the feature extraction from DCEUS. The clinical trial path starts with the urologists and the cancer patients. After obtaining the ethical board approvals, the patients undergo radical prostatectomy. The pathologists delineate the cancerous and benign regions, along with their grade of malignancy. The prostate's shape outside the body and inside the body is not the same due to reasons such as shrinkage and deformation. Through a reconstruction step, the shapes are matched and the ground truth is obtained. The algorithmic process for the DCEUS acquisitions is as follows. The microbubbles are injected intravenously into the patient's arm, and a two-minute recording is done. The resolution of the recording changes in the plane orthogonal to the direction of propagation. This resolution is equalized across space through a deconvolution. The despeckling, which is the focus of this dissertation, is applied after the resolution equalization. Following the despeckling, the markers of prostate cancer are extracted from the DCEUS acquisitions.

In Fig. 1.1, the scope of this dissertation is shown in blue colors. We propose tensor-based algorithms to denoise DCEUS acquisitions and use the classification performance of prostate cancer as a metric to assess their quality. The algorithms analyzed under the Despeckling category in Fig. 1.1 are based on orthogonal tensor decompositions, employing

different probabilistic assumptions to model the speckle noise. These probability distributions are the Fisher-Tippett distribution and a relaxation based on Gaussian noise with sparse outliers. We analyzed two convex relaxations of low-rankness based on two frameworks described in Section 2.3. For ease of reference, the abbreviations of the methods discussed in the corresponding chapters, namely, General Tensor Estimation (GTE), Orientation Invariant Tensor Nuclear Norm (OITNN), and Multilinear Singular Value Decomposition (MLSVD), are presented herein.

1.3. OUTLINE OF THE DISSERTATION

The dissertation is organized as follows.

CHAPTER 2: TENSOR PRELIMINARIES

We introduce the tensor and matrix algebra that is used in the following chapters. In addition, we introduce the two tensor decomposition frameworks that explain the tensor's low-rankness and compare them.

CHAPTER 3: ULTRASOUND FOR PROSTATE CANCER

We introduce the application domain, including the formation of dynamic contrast-enhanced ultrasound recordings, the segmentation of the prostate in the recording domain, and the annotations provided by urologists. A brief introduction to prostate cancer markers is also provided, which are used in subsequent chapters to assess the performance of denoising methods. We assume that an effective denoising method will improve the separability of benign and malignant voxels. We introduce the speckle noise and motivate the usage of the proposed tensor decomposition frameworks described in Chapter 2 to DCEUS acquisitions.

CHAPTER 4: COMPARISONS OF CONVEX OPTIMIZATION ALGORITHMS FOR SPECKLE DENOISING: A SIMULATION STUDY

The log-transformed speckle noise can be modeled as WGN with sparse outliers. We investigate the denoising performance of low-rank tensor decomposition methods for denoising WGN and WGN with sparse outliers under various SNR conditions. This chapter compares various low-rank decomposition methods to identify which algorithms perform best in high and low SNR scenarios. We formulate several optimization algorithms that incorporate the tensor low-rankness and sparsity of the noise through their convex relaxation. For the tensor low-rankness, we introduce the sum of the nuclear norm (SNN) and tensor nuclear norm (TNN). In addition, the sparsity of the noise is represented through its convex relaxation, the L1 norm. Through the convex relaxation of the tensor

low-rankness and the sparsity of the noise, we can iteratively solve the optimization algorithms using proximity operators.

This chapter answers the first research question. Here, we run simulations by incorporating the knowledge about the true tensor, i.e., the ranks and the infinity norm. In DCEUS acquisitions, such knowledge about the true tensor is not available. They need to be estimated. In the following chapters, we address these questions and report the prostate cancer classification results.

CHAPTER 5: DENOISING OF DYNAMIC CONTRAST-ENHANCED ULTRASOUND USING MULTILINEAR SINGULAR VALUE DECOMPOSITION

We introduce an algorithm to estimate the multilinear ranks of a DCEUS acquisition and truncate it to reconstruct a signal subspace that represents microbubble dynamics. Through simulation, we examine how well the ranks are estimated and compare denoising performance with the existing literature. We also present classification results on six DCEUS acquisitions. This chapter partially answers the second research question. The truncated multilinear singular value decomposition (tr-MLSVD) is found to perform well in high SNR scenarios in Chapter 4. The tr-MLSVD is better suited for WGN. In the next chapter, we aim to extend the denoising performance by including the prior probability distribution of speckles.

CHAPTER 6: SPECKLE DENOISING OF DYNAMIC CONTRAST-ENHANCED ULTRASOUND: A COMPARISON OF NOISE MODELS AND ALGORITHMS

We introduce two algorithms that incorporate the probability distribution of speckle into tensor decomposition. The first method is a gradient-descent-based approach that minimizes the negative log-likelihood of the data parametrized by the distribution of the speckle noise. The second method, OITNN, is investigated in Chapter 4 through simulation, and is found to be performing well in low SNR scenarios. We test the performance on a larger patient cohort compared to Chapter 5 and compare it with various other tensor and non-tensor-based denoising techniques. This chapter partially answers the second research question.

CHAPTER 7: CONSTRAINED CRAMÉR-RAO BOUND OF MULTILINEAR SINGULAR VALUE DECOMPOSITION

The Cramér-Rao bound is commonly used in estimation problems to calculate an asymptotic lower bound on the variance of locally unbiased estimates. This chapter introduces a lower bound on the variance of locally unbiased estimates of the components of a unique multilinear sin-

gular value decomposition (MLSVD). The properties of MLSVD, such as the orthonormal factor matrices and the all-orthogonal core tensor, are incorporated as additional constraints into the Cramér-Rao bound formulation, and the resulting constrained Cramér-Rao bound is derived under the assumption of white Gaussian noise. This chapter answers the third research question.

CHAPTER 8: CONCLUSION AND FUTURE WORK

We conclude the dissertation by answering the overarching research question, "How can low-rank tensor decomposition methods be leveraged to effectively denoise DCEUS?" through the investigation of the research objectives provided in Section 1.2. We express the limitations of the dissertation and propose future work.

1.4. RESEARCH CONTEXT: ANGIOGENESIS ANALYTICS

This research is, in part, funded by Angiogenesis Analytics, which is a start-up specializing in prostate cancer diagnosis using DCEUS. In a joint collaboration with the Technical University of Eindhoven, the company conducted clinical trials in Amsterdam Medical Center, Netherlands Cancer Institute, Antoni van Leeuwenhoek, Andros Clinics, Canisius Wilhelmina Hospital. Their product uses AI algorithms to analyze 3D B-mode, 3D elastography, and 4D DCEUS acquisitions, produces a 3D heatmap indicating the potential presence of prostate cancer, assists the urologist in diagnosing, and facilitates target prostate biopsy. This dissertation contributes to the preprocessing efforts aligned with the company's overarching goal of advancing prostate cancer diagnosis using ultrasound. They provided the dataset used in this dissertation, as well as access to the servers on which the algorithms were executed.

1.5. LIST OF PUBLICATIONS

In the context of this dissertation, the following papers have been published or submitted:

JOURNAL

1. M. Calis, M. Mischi, A. van der Veen, and B. Hunyadi, "Speckle Denoising of Dynamic Contrast-enhanced Ultrasound using Low-rank Tensor Decomposition" In: Transactions in Medical Imaging (TMI), 2025
2. M. Calis, R.T. Rajan, M. Mischi, A. van der Veen, and B. Hunyadi, "Constrained Cramér-Rao Bound of Higher-order Singular Value Decomposition" In: IEEE Open Journal of Signal Processing (IEEE-OJSP), 2025

CONFERENCES

1. M. Calis, M. Mischi, A. van der Veen, and B. Hunyadi. “Denoising of dynamic contrast-enhanced ultrasound sequences: a multilinear approach.” In: *BIOSIGNALS*. 2022, pp. 192–199
2. P. Chen, M. Calis, H. Wijkstra, P. Huang, B. Hunyadi, and M. Mischi. “Multiparametric ultrasound and machine learning for prostate cancer localization”. In: *European Signal Processing Conference*. 2022, pp. 907–911
3. M. Calis and B. Hunyadi. “Denoising of the speckle noise by robust low-rank tensor decomposition”. In: *2024 32nd European Signal Processing Conference (EUSIPCO)*. 2024

SYMPOSIUM

1. M.Calis, A. van der Veen, M. Mischi, B. Hunyadi, "Outlier Resistant Tensor Decompositions for Denoising Contrast-Enhanced Ultrasound" In: 9, BME Dutch Conference for Biomedical Engineering (2023)

2

TENSOR PRELIMINARIES

Conventionally, we explain the physical domain with three Cartesian coordinates. A single observation of something in this physical domain is a 3rd-order snapshot. If we want to explain the change of such a physical domain over time, we can stack the snapshots and obtain 4th-order data. Such higher-order data is called tensors. We can consider linear transformations as matrices that operate on the vectorization of these tensors. On the other hand, tensors can be considered as multilinear mappings. Tensor decompositions go beyond the matrix factorizations with additional subscripts and provide more detailed descriptions with different geometric and uniqueness properties [4][5]. In this chapter, we introduce the multilinear algebra concepts used throughout the dissertation.

2.1. FROM SCALARS TO TENSORS

The order is the minimum number of coordinates required to describe all the points that make up an object. Scalars are 0th-order tensors, vectors are 1st order tensors, and matrices are 2nd order tensors. We follow the notation of Cichocki, Lee, and Oseledets [6][7] to describe tensors, matrices, vectors, and scalars. Tensors are represented by underlined boldface letters such as the Nth-order $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$. We use the word mode to refer to a specific order. The tensor $\underline{\mathbf{Y}}$ has I_1 elements in the first mode, I_2 elements in the second, and so on until the Nth mode that has I_N elements. We use small letters such as $i_1 \in \{1, \ldots, I_1\}$, $i_2 \in \{1, \ldots, I_2\}$, or $i_N \in \{1, \ldots I_N\}$ to define the elements in the corresponding modes.

Several matrices are commonly used. The identity matrix is denoted with $\mathbf{I}$. Matrices that share a similar property are denoted with superscript notation. For example, in the tensor decompositions explained in Section 2.3, the N orthonormal factor matrices of multilinear singular value decomposition are explained with $\mathbf{U}^{(n)}$ for $n \in \{1, \ldots, N\}$. Such a description allows a simpler notation where superscripts are used to explain the related matrices concisely instead of a different letter.

Matrices are represented by boldface letters such as $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$. We can fix the second mode of $\mathbf{A}$ to i_2 and define a column vector $\mathbf{a}_{i_2} \in \mathbb{R}^{I_1}$. Scalars are represented by lower case letters such as a_{i_1, i_2} that represent the element at the i_1th row and i_2th column of the matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$. Along with the aforementioned

notation, we interchangeably use MATLAB notation to describe the different elements of a tensor, for example, when Greek letters or explicit elements of a tensor are described. An example is the (i_1, i_2, i_3)th element of a tensor $\underline{\mathbf{\Omega}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ that can be shown as $\underline{\mathbf{\Omega}}(i_1, i_2, i_3)$. Additionally, the frontal slice when we fix the third mode to i_3 can be shown with $\underline{\mathbf{\Omega}}(:, :, i_3)$. When all the modes but one are fixed, such as $\underline{\mathbf{\Omega}}(:, i_2, i_3)$ or $\underline{\mathbf{\Omega}}(i_1, :, i_3)$, we call this a fiber of $\underline{\mathbf{\Omega}}$.

Equipped with the scalar and tensor notation, we can define operations between tensors and matrices, namely, the tensor and matrix algebra. A simple indexing to refer to the elements of a tensor is defined. We build upon such indexing descriptions to explain various unfoldings of a tensor, such as vectorization and mode-n unfoldings. Finally, we will use such unfoldings to describe various tensor decompositions that are used throughout the dissertation.

2.2. TENSOR AND MATRIX ALGEBRA

2.2.1. INDEXING

Indexing is the process of using indices to refer to or locate the elements of a tensor. Tensors are higher-order arrays, and the elements are stored in memory using two common conventions. The first one is called the little-endian, and the second one is called the big-endian. In the little-endian format, we start counting from the first mode, whereas in the big-endian format, we start counting from the last mode. Big-endian indexing is used in Python, and in MATLAB, little-endian indexing is used. In this dissertation, we use the little-endian indexing. For example, the first and the second element of the tensor $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ are $\underline{\mathbf{A}}(1, 1, 1)$ and $\underline{\mathbf{A}}(2, 1, 1)$. On the other hand, for the big-endian indexing, the first and the second elements are $\underline{\mathbf{A}}(1, 1, 1)$ and $\underline{\mathbf{A}}(1, 1, 2)$ for the same tensor. Such a distinction is important once different coding languages are used. The basic notations, such as vectorization, unfoldings, Kronecker products, and outer products, should be consistent with the definition of the convention.

2.2.2. VECTORIZATION

The vectorization of the tensor vec($\underline{\mathbf{Y}}$) results in the vector $\mathbf{y}$ with $\mathbf{y}(\overline{i_1 i_2 i_3 \dots i_N}) = y_{i_1, i_2, i_3, \dots, i_N}$. We use the little-endian or MATLAB style ordering of the indices. This is shown with

$$\overline{i_1 i_2 i_3 \dots i_N} = i_1 + (i_2 - 1)I_1 + (i_3 - 1)I_1 I_2 + \cdots + (i_N - 1)I_1 I_2 \dots I_{N-1}\,. \tag{2.1}$$

For a matrix, this vec-operator amounts to column-wise stacking.

2.2.3. MODE-n UNFOLDING

The mode-n unfolding of a tensor is a matrix created by stacking all the fibers in the nth mode while traversing the rest in a MATLAB-style ordering. The mode-n unfolding of $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$ is depicted with the subscript in parenthesis, that is, $\mathbf{Y}_{(n)} \in \mathbb{R}^{I_n \times I_1 \dots I_{n-1} I_{n+1} \dots I_N}$. The indices have the following relation

$$\mathbf{Y}_{(n)}\left(i_n,\ \overline{i_1 \dots i_{n-1}\, i_{n+1} \dots i_N}\right) = y_{i_1, i_2, i_3, \dots, i_N}\,. \tag{2.2}$$

Two steps will result in such an unfolding: swapping the nth mode with the first mode and reshaping the tensor into a matrix of size $I_n \times I_1 \dots I_{n-1} I_{n+1} \dots I_N$. Here, the MATLAB indexing is used. Note that for the mode-1 unfolding, i.e. $n = 1$, we have $\overline{i_1 \dots i_{n-1}\, i_{n+1} \dots i_N} = \overline{i_2 \dots i_N}$.

The vectorizations of different mode-n unfoldings are related using permutation matrices. The permutation matrices are represented with $\mathbf{P}^{(\cdot)}$ that define a row-wise permutation between vectorization of the different mode-n unfoldings. An example is the permutation matrix $\mathbf{P}^{(2)}$ that permutes the vectorization of the mode-2 unfolding of $\underline{\mathbf{M}} \in \mathbb{R}^{I_1 \times \dots \times I_N}$, such that vectorization of the mode-1 unfolding of $\underline{\mathbf{M}}$ is obtained, i.e. $\text{vec}(\mathbf{M}_{(1)}) = \mathbf{P}^{(2)} \text{vec}(\mathbf{M}_{(2)})$. The matrix $\mathbf{P}^{(2)}$ can be generated by first creating a square matrix of 0s of size $\prod_{n=1}^{N} I_n \times \prod_{n=1}^{N} I_n$ and assigning $p^{(2)}_{\overline{i_1 i_2 \dots i_N}, \overline{i_2 i_1 \dots i_N}} = 1$. As an example, we illustrate the structure of the permutation matrix $\mathbf{P}^{(2)}$ in Fig. 2.1 for a tensor $\underline{\mathbf{M}} \in \mathbb{R}^{2 \times 2 \times 2}$.

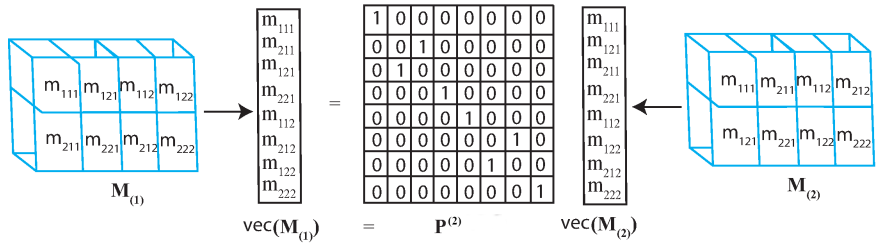


Figure 2.1: The structure of the $\mathbf{P}^{(2)}$ matrix for a tensor $\underline{\mathbf{M}} \in \mathbb{R}^{2 \times 2 \times 2}$. $\mathbf{P}$ permutes the vectorization of the second mode unfolding of $\underline{\mathbf{M}}$ to the vectorization of the first mode unfolding of $\underline{\mathbf{M}}$.

2.2.4. MODE-n FOLDING

The operation that reverts the mode-n unfolding $\mathbf{Y}_{(n)}$ of a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \dots \times I_N}$ is depicted with fold_n and has the relation

$$\text{fold}_n(\mathbf{Y}_{(n)}) = \underline{\mathbf{Y}}, \tag{2.3}$$

for $n \in \{1, \dots, N\}$.

2.2.5. MODE-n PRODUCT

The mode-n product, also called the tensor-matrix product, of a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \dots \times I_N}$ and a matrix $\mathbf{B} \in \mathbb{R}^{J \times I_n}$ yields the tensor

$$\underline{\mathbf{C}} = \underline{\mathbf{Y}} \times_n \mathbf{B} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}, \tag{2.4}$$

with entries

$$c_{i_1, i_2, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} y_{i_1, i_2, \dots, i_n, \dots, i_N}\, b_{j, i_n}. \tag{2.5}$$

The following relation holds

$$\mathbf{C}_{(n)} = \mathbf{B}\,\mathbf{Y}_{(n)}\,. \tag{2.6}$$

The mode-n product takes the linear combinations of the mode-n fibers of the tensor $\underline{\mathbf{Y}}$ with the weight specified by the rows of the matrix $\mathbf{B}$.

2

2.2.6. OUTER PRODUCT

The outer product of two tensors $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ yields a tensor $\underline{\mathbf{C}} = \underline{\mathbf{A}} \circ \underline{\mathbf{B}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N \times J_1 \times J_2 \times \cdots \times J_N}$ with entries

$$c_{i_1,\ldots,i_N,j_1,\ldots,j_N} = a_{i_1,\ldots,i_N}\, b_{j_1,\ldots,j_N}\,. \tag{2.7}$$

2.2.7. KRONECKER PRODUCT

For two tensors $\mathbf{A}^{I_1 \times \cdots \times I_N}$ and $\mathbf{B} \in \mathbb{R}^{J_1 \times \cdots \times J_N}$, the left Kronecker product $\underline{\mathbf{C}} = \underline{\mathbf{A}} \otimes \underline{\mathbf{B}}$ yields the tensor $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 J_1 \times \cdots \times I_N J_N}$ with entries

$$c_{\overline{i_1 j_1},\,\ldots,\,\overline{i_N j_N}} = a_{i_1,\ldots,i_N}\, b_{j_1,\ldots,j_N}\,. \tag{2.8}$$

Throughout the dissertation, we use the Kronecker product as the left Kronecker product (2.8), which uses the little-endian indexing. The Kronecker product and outer product are related. For two tensors $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$, we have

$$\operatorname{vec}\left(\underline{\mathbf{A}} \circ \underline{\mathbf{B}}\right) = \operatorname{vec}\left(\underline{\mathbf{A}}\right) \otimes \operatorname{vec}\left(\underline{\mathbf{B}}\right), \tag{2.9}$$

that is, the Kronecker product of the vectorizations of two tensors is equivalent to the vectorization of their outer product.

2.2.8. HADAMARD PRODUCT

For two matrices $\mathbf{A}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{I \times J}$ the Hadamard product yields the matrix

$$\mathbf{C} = \mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{1,1}b_{1,1} & \cdots & a_{1,J}b_{1,J} \\ \vdots & \ddots & \vdots \\ a_{I,1}b_{I,1} & \cdots & a_{I,J}b_{I,J} \end{bmatrix} \in \mathbf{R}^{I \times J}. \tag{2.10}$$

2.2.9. INNER PRODUCT

The inner product between two tensors is represented with

$$\langle \underline{\mathbf{A}}, \underline{\mathbf{B}} \rangle = \operatorname{vec}(\underline{\mathbf{A}})^{\top} \operatorname{vec}(\underline{\mathbf{B}}), \tag{2.11}$$

for Nth-order tensors $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ with same number of elements.

2.2.10. FROBENIUS NORM

The Frobenius norm of a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is given by

$$\|\underline{\mathbf{Y}}\|_F = \sqrt{\langle \underline{\mathbf{Y}}, \underline{\mathbf{Y}} \rangle}\,. \tag{2.12}$$

2.2.11. NUCLEAR NORM

The nuclear norm of a matrix is the sum of its singular values. For a real-valued rank R matrix $\mathbf{Y} \in \mathbb{R}^{I \times J}$, and its SVD

$$\mathbf{Y} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathsf{T}}, \tag{2.13}$$

with orthonormal $\mathbf{U} \in \mathbb{R}^{I \times R}$, $\mathbf{V} \in \mathbb{R}^{J \times R}$, and diagonal and positive $\mathbf{\Sigma} \in \mathbb{R}^{R \times R}$, the nuclear norm is defined as

$$\|\mathbf{Y}\|_* = \mathrm{Tr}(\mathbf{\Sigma}). \tag{2.14}$$

2.2.12. TENSOR-TENSOR PRODUCT

The tensor-tensor product $\underline{\mathbf{A}} \in R^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$ and $\underline{\mathbf{B}} \in R^{I_2 \times J \times I_3 \times \cdots \times I_N}$ yields the tensor

$$\mathbb{R}^{I_1 \times J \times I_3 \times \cdots \times I_N} \ni \underline{\mathbf{C}} = \underline{\mathbf{A}} \times_{\mathsf{F}} \underline{\mathbf{B}}, \tag{2.15}$$

with entries

$$\underline{\mathbf{C}}(:,:,i_3,i_4,\ldots,i_N) = \underline{\mathbf{A}}(:,:,i_3,i_4,\ldots,i_N)\ \underline{\mathbf{B}}(:,:,i_3,i_4,\ldots,i_N). \tag{2.16}$$

The tensor-tensor product is the matrix multiplication of each frontal slice of two Nth-order tensors.

2.2.13. T-PRODUCT

The t-product is proposed in [8] to define an algebraic group on 3rd-order tensors. The t-product uses the tensor-tensor product 2.2.12 and it is the building block of the t-SVD decomposition, which is introduced in Section 2.3.2. It is the circular convolution of the 3rd-mode fibers. The t-product between two 3rd-order tensors $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{I_2 \times J \times I_3}$ yields a tensor $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 \times J \times I_3}$,

$$\underline{\mathbf{C}}(i_1,j,:) = \sum_{i_2=1}^{I_2} \underline{\mathbf{A}}(i_1,i_2,:) \circledast \underline{\mathbf{B}}(i_2,j,:), \tag{2.17}$$

where $\circledast$ is the circular convolution operator. This connection provides a motivation for using the Discrete Fourier Transform (DFT), since convolution in the time domain becomes multiplication in the frequency domain. We will introduce this connection by introducing the t-product between Nth-order tensors using the DFT operations. In [9], the t-product is extended to Nth-order tensors. We express the Nth-order t-product using the general form in [10].

The t-product between two tensors $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{I_2 \times J \times I_3 \times \cdots \times I_N}$ yields a tensor $\underline{\mathbf{C}} \in \mathbb{R}^{I_1 \times J \times I_3 \times \cdots \times I_N}$,

$$\underline{\mathbf{C}} = \underline{\mathbf{A}} \star \underline{\mathbf{B}} = \left(\left(\underline{\mathbf{A}} \times_3 \mathbf{F}^{(3)} \cdots \times_N \mathbf{F}^{(N)}\right) \times_{\mathsf{F}} \left(\underline{\mathbf{B}} \times_3 \mathbf{F}^{(3)} \cdots \times_N \mathbf{F}^{(N)}\right)\right) \times_3 \mathbf{F}^{(3)^{-1}} \cdots \times_N \mathbf{F}^{(N)^{-1}}, \tag{2.18}$$

with the Discrete Fourier Transform matrix $\mathbf{F}^{(n)}$ for $n \in \{3,\ldots,N\}$ that is given by

$$\mathbf{F}^{(n)} = \frac{1}{\sqrt{I_n}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{I_n-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(I_n-1)} \\ 1 & w^3 & w^6 & \cdots & w^{3(I_n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{I_n-1} & w^{2(I_n-1)} & \cdots & w^{(I_n-1)(I_n-1)} \end{bmatrix}, \tag{2.19}$$

where $w = e^{-2\pi i/I_n}$ and $i^2 = -1$. The multiplication $\underline{\mathbf{A}} \times_n \mathbf{F}^{(n)}$ results in the one-dimensional Fourier transform of $\underline{\mathbf{A}}$ in the nth mode. For a 3rd-order tensor, we can see that (2.17) and (2.18) are equivalent. After the 1D Fourier transforms $\underline{\mathbf{A}} \times_3 \mathbf{F}^{(3)}$ and $\underline{\mathbf{B}} \times_3 \mathbf{F}^{(3)}$, we have the DC components in the first frontal slices, the lowest frequency components in the second frontal slices until the the middle frontal slices, which correspond to the highest frequency components. The frontal slices after the middle, $i_3 = \{I_3/2, \ldots, I_3\}$, exhibit Hermitian symmetry due to the DFT's properties when the input tensors are real-valued. When I_3 is odd, appropriate adaptations of the DFT can be applied to preserve the symmetry properties [11]. The circular convolution of the 3rd-mode fibers of tensors $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ is thus equivalent to matrix multiplication of the corresponding frontal slices of $\underline{\mathbf{A}} \times_3 \mathbf{F}^{(3)}$ and $\underline{\mathbf{B}} \times_3 \mathbf{F}^{(3)}$, followed by an inverse Fourier transform.

Here, we give some properties related to the t-product. More information can be obtained in [12].

- $\underline{\mathbf{A}} \star (\underline{\mathbf{B}} \star \underline{\mathbf{C}}) = (\underline{\mathbf{A}} \star \underline{\mathbf{B}}) \star \underline{\mathbf{C}}$.
- $(\underline{\mathbf{A}} + \underline{\mathbf{B}}) \star \underline{\mathbf{C}} = \underline{\mathbf{A}} \star \underline{\mathbf{C}} + \underline{\mathbf{B}} \star \underline{\mathbf{C}}$.
- The set of all invertible $I \times I \times I$ tensors forms a group under $\star$ operation.
- If $T(\underline{\mathbf{B}}) = \underline{\mathbf{A}} \star \underline{\mathbf{B}}$ where $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times J \times I_3}$ and $\underline{\mathbf{B}} \in \mathbb{R}^{J \times I_2 \times I_3}$ are real tensors, then $T : \mathbb{R}^{J \times I_2 \times I_3} \to \mathbb{R}^{I_1 \times I_2 \times I_3}$ is linear.

Equipped with the tensor and matrix algebra, we can define the tensor decompositions.

2.3. TENSOR DECOMPOSITIONS

We investigate two orthogonal tensor decompositions: MLSVD, and t-SVD decomposition. We will start by defining the tensor rank in these two frameworks.

2.3.1. MULTILINEAR SINGULAR VALUE DECOMPOSITION

The multilinear singular value decomposition (MLSVD) is proposed in [13]. The authors show that any tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ can be written as

$$\underline{\mathbf{Y}} = \underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}, \tag{2.20}$$

where $\underline{\mathbf{S}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is called the all-orthogonal core tensor and $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ for $n \in \{1, \ldots, N\}$ are called the column-wise orthonormal factor matrices. The multilinear singular value decomposition is found by applying a thin SVD on each of the N unfoldings of $\underline{\mathbf{Y}}$, i.e., $\mathbf{Y}_{(n)} = \mathbf{U}^{(n)} \mathbf{\Sigma}^{(n)} \mathbf{V}^{(n)^T}$. The thin SVD operation retains the first $\min(I_n, I_1 \ldots I_{n-1} I_n + 1 \ldots I_N)$ columns of the factor matrices and the singular value matrix of the full SVD components [14]. We assume the thin SVD operation orders the singular values $\mathbf{\Sigma}^{(n)}$ from the highest to the lowest. The columns of the factor matrices are also ordered corresponding to $\mathbf{\Sigma}^{(n)}$. The left singular vectors of each unfolding are assigned as the factor matrices $\mathbf{U}^{(n)}$ for $n \in \{1, \ldots, N\}$. Finally, the core tensor can be found by the operation

$$\underline{\mathbf{S}} = \underline{\mathbf{Y}} \times_1 \mathbf{U}^{(1)^T} \ldots \times_N \mathbf{U}^{(N)^T}. \tag{2.21}$$

Algorithm 1 MLSVD

Input : $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$
for $n = \{1, \ldots, N\}$ **do**
  $\mathbf{U}^{(n)} \mathbf{\Sigma}^{(n)} \mathbf{V}^T \leftarrow \text{svd}(\mathbf{Y}_{(n)})$ ▷ thin SVD with $I_n < I_1 \ldots I_{n-1} I_{n+1} \ldots I_N$
end for
$\underline{\mathbf{S}} \leftarrow \underline{\mathbf{Y}} \times_1 \mathbf{U}^{(1)^T} \cdots \times_N \mathbf{U}^{(N)^T}$
Return : $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}, \forall n \in \{1, \ldots, N\}, \underline{\mathbf{S}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$

The operations that decompose $\underline{\mathbf{Y}}$ into the form (2.20) are shown in Algorithm 1. The all-orthogonality of the core tensor can be expressed for $n \in \{1, ..., N\}$ as

$$\sqrt{\mathbf{S}_{(n)} \mathbf{S}_{(n)}^T} = \mathbf{\Sigma}^{(n)}, \tag{2.22}$$

where the square root operator is applied element-wise. For a real-valued tensor, each element of the diagonal singular value matrix $\mathbf{\Sigma}^{(n)}$ is greater than or equal to zero for $n \in \{1, \ldots, N\}$. The column-wise orthonormality of the factor matrices can be expressed for $n \in \{1, \ldots, N\}$ as

$$\mathbf{U}^{(n)^T} \mathbf{U}^{(n)} = \mathbf{I}. \tag{2.23}$$

For a real-valued tensor $\underline{\mathbf{Y}}$, the core tensor $\underline{\mathbf{S}}$ is also real-valued. For the MLSVD of Nth-order tensors with $N > 2$, $\underline{\mathbf{S}}$ can have both negative and positive elements. One might expect $\underline{\mathbf{S}}$ to have non-zero elements only at the super-diagonal. The super-diagonal is defined here as the elements with the same indices in all modes, i.e., $s_{i_1, \ldots, i_N}$ for $i_1 = \cdots = i_N$. The core tensor can not be super-diagonal unless the original tensor can be written exactly as a sum of rank-one tensors, i.e., Nth-order tensors that are outer products of N vectors. The latter decomposition is called the CANDECOMP/PARAFAC decomposition (CPD) [15].

One can also observe that the right orthonormal matrices $\mathbf{V}^T$ from the SVD applied to the mode-n unfoldings given in Algorithm 1 are not used. The right singular vectors $\mathbf{V}^T$ have other applications, such as measuring the similarity of two tensors [16]. On the other hand, we can observe that the MLSVD of a 2nd-order tensor is the usual SVD $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ with $\mathbf{U}^{(1)}$ as the left orthonormal matrix $\mathbf{U}$, $\mathbf{U}^{(2)}$ as the right orthonormal matrix $\mathbf{V}$, and $\mathbf{S}_{(1)} = \mathbf{S}_{(2)} = \mathbf{S} = \mathbf{\Sigma}$. In the multiplication $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, we can see that $\mathbf{U}$ makes linear combinations of the rows of $\mathbf{\Sigma}$, and $\mathbf{V}$ makes linear combinations of the columns of $\mathbf{\Sigma}$. Instead of the transpose, we can use the mode-n product and write $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ as $\mathbf{\Sigma} \times_1 \mathbf{U} \times_2 \mathbf{V}$.

Consider a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. There are N possible mode-n unfoldings as described in (2.2). The rank of each mode-n unfolding of a tensor is called the mode-n rank R_n. The multilinear ranks of a tensor are described with an N-tuple such as $(R_1, \ldots, R_N)$. The mode-n rank of a tensor is expressed with $\text{rank}_n(\underline{\mathbf{Y}})$. For $\underline{\mathbf{Y}}$ this is expressed as

$$\text{rank}_n(\underline{\mathbf{Y}}) = \text{rank}(\mathbf{Y}_{(n)}) = R_n \quad \text{for } n \in \{1, \ldots, N\}. \tag{2.24}$$

The $(R_1, ..., R_N)$ multilinear ranks of a tensor $\underline{\mathbf{Y}}$ has the following relation

$$R_i \leq \prod_{\substack{j=1 \\ j \neq i}}^{N} R_j \quad \text{for } i \in \{1, 2, \ldots, N\}. \tag{2.25}$$

We can use such ranks to truncate the core tensor and the factor matrices accordingly.

TRUNCATED MULTILINEAR SINGULAR VALUE DECOMPOSITION

The MLSVD gives rise to a well-posed low multilinear rank approximation problem, i.e., any tensor can be written using Algorithm 1. The largest contribution explained by the mode-n singular vectors is captured by the top slices of the core tensor due to the ordering defined in (2.22). Consequently, we can truncate the core tensor and take the corresponding mode-n singular vectors for a low multilinear rank approximation of a tensor. This is depicted for a 3rd-order tensor in Fig. 2.2.

Assuming that the multilinear ranks R_n for $n \in \{1, \ldots, N\}$ are known and that the signal subspace is captured by the highest singular values and the vectors, the truncated MLSVD (tr-MLSVD) is achieved by taking the first R_n columns of $\mathbf{U}^{(n)}$ for $n \in \{1, \ldots, N\}$, respectively, and truncating the core tensor by taking the first R_n elements in the respective modes. The low multilinear rank tensor $\underline{\hat{\mathbf{Y}}}$ is obtained by

$$\begin{aligned} \underline{\hat{\mathbf{Y}}} &= \underline{\hat{\mathbf{S}}} \times_1 \hat{\mathbf{U}}^{(1)} \times_2 \cdots \times_N \hat{\mathbf{U}}^{(N)}, \\ \underline{\hat{\mathbf{Y}}} &= \underline{\mathbf{S}}(1\!:\!R_1, \ldots, 1\!:\!R_N) \times_1 \mathbf{U}^{(1)}(:, 1\!:\!R_1) \cdots \times_N \mathbf{U}^{(N)}(:, 1\!:\!R_N). \end{aligned} \tag{2.26}$$

An illustration is shown in Fig. 2.2. Truncation of the core tensor and the factor

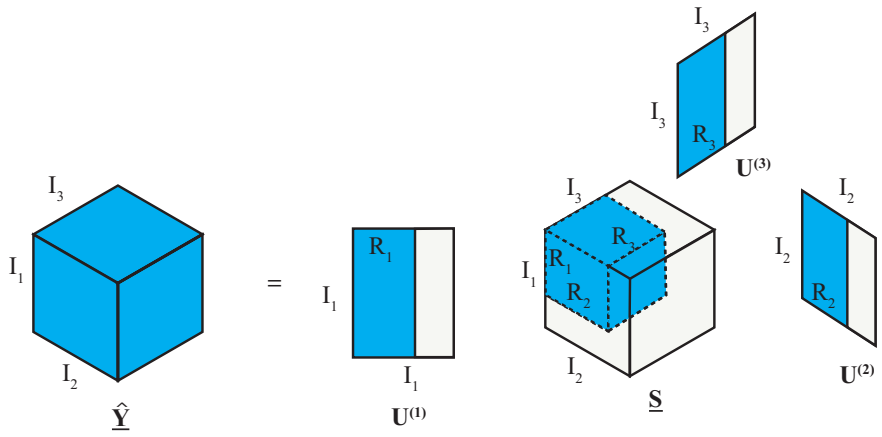


Figure 2.2: The truncated multilinear singular value decomposition of a 3rd-order tensor with ranks (R_1, R_2, R_3).

matrices according to the desired rank does not yield the best approximation. In fact, the truncated MLSVD satisfies the quasi-best approximation Theorem 10.2 in [17]. Define $\underline{\hat{\mathbf{Y}}}$ as the truncated core tensor as defined in (2.26) and $\underline{\mathbf{Y}}^{best}$ the estimated low multilinear rank tensor that is jointly optimized on the product of Stiefel Manifolds, i.e., the manifold of the product of orthonormal matrices with

sizes $I_n \times R_n$ for $n \in \{1, \ldots, N\}$. We have

$$\left\| \underline{\mathbf{Y}} - \hat{\underline{\mathbf{S}}} \times_1 \hat{\mathbf{U}}^{(1)} \times_2 \cdots \times_N \hat{\mathbf{U}}^{(N)} \right\|_F \leq \sqrt{N} \left\| \underline{\mathbf{Y}} - \underline{\mathbf{Y}}^{\text{best}} \right\|_F . \tag{2.27}$$

In an informal way, we can explain the reason why tr-MLSVD does not give the best multilinear rank approximation as follows. Consider the SVD in a mode-n unfolding. With known ranks, the best that can be done in this low-rank approximation is the projection of the matrix onto the signal subspace, denoted by the rank. We can do this process for each of the unfoldings $n \in \{1, \ldots, N\}$. However, these subspaces are not independent of each other for MLSVD. Some of the information that is truncated in a mode-n unfolding is truncated again in another mode-n unfolding, depending on the overlap. This can be visualized with the truncation of a 3rd-order core tensor, given in Fig. 2.2. For example, if we truncate in the first mode with rank R_1, and keep the other modes, we will have $\hat{\mathbf{U}}^{(1)} = \mathbf{U}^{(1)}(:, 1 : R_1)$ and $\hat{\underline{\mathbf{S}}} = \underline{\mathbf{S}}(1 : R_1, :, :)$. If we truncate in the second mode, we will have $\hat{\mathbf{U}}^{(2)} = \mathbf{U}^{(2)}(:, 1 : R_2)$ and $\hat{\underline{\mathbf{S}}} = \underline{\mathbf{S}}(:, 1 : R_2, :)$. The region in $\underline{\mathbf{S}}(R_1 : I_1, R_2 : I_2, :)$ is truncated twice. We call this the overlapping region between mode-1 and mode-2. Similarly, the mode-3 truncation will have overlapping regions with the other modes. Once part of the core is truncated in one mode, the structure of the unfolding in the other modes is altered, and thus the truncated singular vectors $\hat{\mathbf{U}}^{(n)}$ are no longer true left singular vectors of the truncated mode-n unfolding. The interdependence between the modes undermines the optimality of separate truncations. A joint optimization on a Stiefel Manifold includes the dependencies of each mode-n unfolding and provides more optimal low multilinear rank approximations [18]. This joint optimization is prohibitively complex.

The authors in [19] tackled the problem of the best rank $(R_1, \ldots, R_N)$ approximation of a tensor and introduced the higher-order orthogonal iteration (HOOI) algorithm, which is a multilinear extension of the orthogonal iterations for the matrix case, given in Section 7.3.2 of [14]. The approximation is improved compared to the tr-MLSVD. However, several local minima are found. The least squares formulation on the Frobenius norm of the residual error contains local minima. Furthermore, the authors in [20] introduced optimization algorithms that tackle the best rank $(R_1, \ldots, R_N)$ approximation of a tensor on a Stiefel manifold of orthonormal matrices of size $I_n \times R_n$. The authors found a slight improvement in the performance over HOOI and tr-MLSVD. However, the problem of global vs local minima still persists. The multilinear extension of the Eckhart-Young theorem [21], which says that the best low-rank approximation of a matrix can be obtained through the truncated SVD if we consider the Frobenius norm on the residual error, is not possible using the multilinear rank [22]. On the other hand, using the t-product defined in Section 2.2.13, an Eckhart-Young-like theorem for low-rank tensor approximation is achieved by tensor SVD (t-SVD) for 3rd-order tensors.

2.3.2. TENSOR-SVD

The tensor SVD (t-SVD) is defined for 3rd-order tensors in [8] and extended for any Nth-order tensor in [9]. The decomposition uses t-product as a building block and can be considered as modeling tensor low-rankness in the Fourier domain.

Most of the real data often has a limited bandwidth. Therefore, t-SVD decomposition is suited for applications where certain frequency bands are more important than others. The t-product explained in Section 2.2.13 always applies the 1D DFT to the modes greater than two. Hence, the t-SVD framework is an orientation-specific decomposition. Consider a 4th-order tensor with the first three indices indicating the Cartesian domain and the last axis indicating the time domain. The application of t-SVD on two cases where we switch the ordering from $I_1 \times I_2 \times I_3 \times I_4$ to $I_1 \times I_3 \times I_2 \times I_4$ gives two different decompositions. Suppose that the frequency content in the second mode and the fourth mode is expected to be significantly more bandlimited, the permutation $I_1 \times I_3 \times I_2 \times I_4$ will give a better t-SVD decomposition. In [23], a preliminary analysis through simulation or investigation on a subset of the real data is advised to assess the best possible orientation. We start with explaining the basics of the t-SVD decomposition and compare it to MLSVD in Section 2.4.

We use the breve notation for t-SVD-related components and start by defining the algebraic group related to the t-product explained in Section 2.2.13.

- The Nth-order identity tensor $\breve{\underline{\mathbf{I}}} \in \mathbb{R}^{J \times J \times I_3 \times \cdots \times I_N}$ is a tensor such that $\breve{\underline{\mathbf{I}}}(:,:,\ldots,:,1)$ is an $(N-1)$st-order identity tensor, and $\breve{\underline{\mathbf{I}}}(:,:,\ldots,:,i_n)$ is an $(N-1)$st-order zero tensor for $i_n \in \{2,\ldots,I_N\}$. For an Nth-order tensor $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times J \times I_3 \times \cdots \times I_N}$, we have

$$\breve{\underline{\mathbf{I}}} \star \underline{\mathbf{A}} = \underline{\mathbf{A}} \star \breve{\underline{\mathbf{I}}} = \underline{\mathbf{A}} \,.$$

  An illustration is provided in Fig. 2.3 for a 4th-order identity tensor.

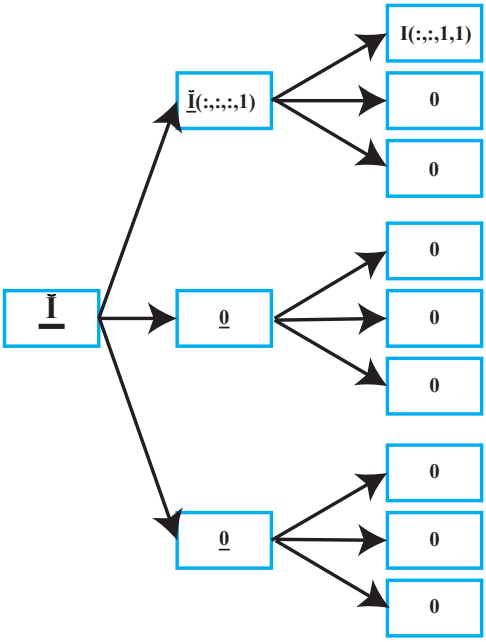


Figure 2.3: The identity tensor given in Section 2.3.2 for $\breve{\underline{\mathbf{I}}} \in \mathbb{R}^{J \times J \times 3 \times 3}$

- The transpose of $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$ is denoted by $\underline{\mathbf{A}}^\top \in \mathbb{R}^{I_2 \times I_1 \times I_3 \times \cdots \times I_N}$, and is obtained by transposing each $\underline{\mathbf{A}}(:, \ldots, :, i_n)$ and then reversing the ordering for $i_n \in \{2, \ldots, I_N\}$. An illustration is provided in Fig. 2.4 for a 4th-order tensor.

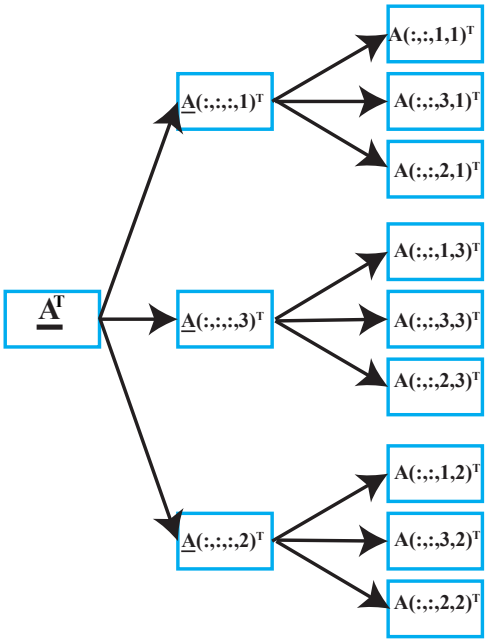


Figure 2.4: The tensor transpose described in Section 2.3.2 for $\underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times I_2 \times 3 \times 3}$.

- A tensor $\underline{\mathbf{A}} \in \mathbb{R}^{J \times J \times I_3 \times \cdots \times I_N}$ is said to be orthogonal if

$$\underline{\mathbf{A}} \star \underline{\mathbf{A}}^\top = \underline{\mathbf{A}}^\top \star \underline{\mathbf{A}} = \check{\mathbf{I}}.$$

- The inverse of a tensor $\underline{\mathbf{A}} \in \mathbb{R}^{J \times J \times I_3 \times \cdots \times I_N}$ is a tensor $\underline{\mathbf{B}} \in \mathbb{R}^{J \times J \times I_3 \times \cdots \times I_N}$ such that

$$\underline{\mathbf{A}} \star \underline{\mathbf{B}} = \underline{\mathbf{B}} \star \underline{\mathbf{A}} = \check{\mathbf{I}}. \tag{2.28}$$

With the definition of identity, orthogonality, and transpose, we can express the t-SVD decomposition. Any Nth-order tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ can be written as

$$\underline{\mathbf{Y}} = \check{\mathbf{U}} \star \check{\mathbf{S}} \star \check{\mathbf{V}}^\top, \tag{2.29}$$

where $\check{\mathbf{U}} \in \mathbb{R}^{I_1 \times I_1 \times I_3 \times \cdots \times I_N}$ and $\check{\mathbf{V}} \in \mathbb{R}^{I_2 \times I_2 \times I_3 \times \cdots \times I_N}$ are orthogonal tensors, and $\check{\mathbf{S}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$ is a forward-diagonal core tensor [12][8]. The forward-diagonal tensor has diagonal frontal slices. The t-SVD decomposition can be found using Algorithm 2. First the 1D Fourier transforms of a tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is taken for $n \in \{3, \ldots, N\}$. The SVDs of the frontal slices of the resulting tensor, that is, the slices corresponding to the first two indices, are calculated. The components of

the SVD $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$ are assigned to the corresponding locations $\breve{\underline{\mathbf{U}}}(:,:,i_3,\ldots,i_N)$, $\breve{\underline{\mathbf{V}}}(:,:,i_3,\ldots,i_N)$, and $\breve{\underline{\mathbf{S}}}(:,:,i_3,\ldots,i_N)$, respectively. Finally, the 1D inverse Fourier transforms of $\breve{\underline{\mathbf{U}}}$, $\breve{\underline{\mathbf{S}}}$, $\breve{\underline{\mathbf{V}}}$ are taken for $n \in \{3,\ldots,N\}$.

Algorithm 2 *N*th-order t-SVD

Input: $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$
$\underline{\mathbf{Y}} \leftarrow \underline{\mathbf{Y}} \times_3 \mathbf{F}^{(3)} \times_4 \mathbf{F}^{(4)} \cdots \times_N \mathbf{F}^{(N)}$ ▷ $\mathbf{F}^{(n)}$ is the DFT matrix from (2.19)
for $i_3 = \{1,\ldots,I_3\}$, $i_4 = \{1,\ldots,I_4\}$, $\ldots$, $i_N = \{1,\ldots,I_N\}$ **do**
  $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] \leftarrow \text{svd}\left(\underline{\mathbf{Y}}(:,:,i_3,i_4,\ldots,i_N)\right)$
  $\breve{\underline{\mathbf{U}}}(:,:,i_3,i_4,\ldots,i_N) \leftarrow \mathbf{U}$
  $\breve{\underline{\mathbf{S}}}(:,:,i_3,i_4,\ldots,i_N) \leftarrow \mathbf{\Sigma}$
  $\breve{\underline{\mathbf{V}}}(:,:,i_3,i_4,\ldots,i_N) \leftarrow \mathbf{V}$
end for
$\breve{\underline{\mathbf{U}}} \leftarrow \breve{\underline{\mathbf{U}}} \times_3 \mathbf{F}^{(3)^{-1}} \times_4 \mathbf{F}^{(4)^{-1}} \cdots \times_N \mathbf{F}^{(N)^{-1}}$ ▷ $\mathbf{F}^{(n)^{-1}}$ is the inverse DFT matrix
$\breve{\underline{\mathbf{S}}} \leftarrow \breve{\underline{\mathbf{S}}} \times_3 \mathbf{F}^{(3)^{-1}} \times_4 \mathbf{F}^{(4)^{-1}} \cdots \times_N \mathbf{F}^{(N)^{-1}}$
$\breve{\underline{\mathbf{V}}} \leftarrow \breve{\underline{\mathbf{V}}} \times_3 \mathbf{F}^{(3)^{-1}} \times_4 \mathbf{F}^{(4)^{-1}} \cdots \times_N \mathbf{F}^{(N)^{-1}}$
Return: $\breve{\underline{\mathbf{U}}} \in \mathbb{R}^{I_1 \times I_1 \times I_3 \times \cdots \times I_N}$, $\breve{\underline{\mathbf{S}}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$, $\breve{\underline{\mathbf{V}}} \in \mathbb{R}^{I_2 \times I_2 \times I_3 \times \cdots \times I_N}$

For t-SVD, the rank is defined using the tubal rank, which is denoted with rank_{tbl}. The tubal rank rank_{tbl} is defined as the maximum rank of any frontal slice of $\underline{\mathbf{Y}} \leftarrow \underline{\mathbf{Y}} \times_3 \mathbf{F}^{(3)} \times_4 \mathbf{F}^{(4)} \times_5 \cdots \times_N \mathbf{F}^{(N)}$, that is,

$$\text{rank}_{\text{tbl}}(\underline{\mathbf{Y}}) = \max_{i_3,\ldots,i_N} \text{rank}\left(\left(\underline{\mathbf{Y}} \times_3 \mathbf{F}^{(3)} \times_4 \mathbf{F}^{(4)} \times_5 \cdots \times_N \mathbf{F}^{(N)}\right)(:,:,i_3,\ldots,i_N)\right). \quad (2.30)$$

The tubal rank has properties [24] like the matrix rank, such as

$$\text{rank}_{\text{tbl}} \leq \min(I_1, I_2). \quad (2.31)$$

For this reason, the decomposition in (2.29) can also be expressed as

$$\underline{\mathbf{Y}} = \sum_{i=1}^{\min(I_1,I_2)} \breve{\underline{\mathbf{U}}}(:,i,:,\ldots,:) \star \breve{\underline{\mathbf{S}}}(i,i,:,\ldots,:) \star \breve{\underline{\mathbf{V}}}(:,i,:,\ldots,:)^{\mathsf{T}}. \quad (2.32)$$

The tubal rank of the t-product of two tensors $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times J \times I_3 \ldots I_N}$ and $\underline{\mathbf{X}} \in \mathbb{R}^{J \times I_2 \times I_3 \ldots I_N}$ will be less than or equal to the tubal rank of each tensor. This can be described by

$$\text{rank}_{\text{tbl}}\left(\underline{\mathbf{Y}} \star \underline{\mathbf{X}}\right) \leq \min\left(\text{rank}_{\text{tbl}}\left(\underline{\mathbf{Y}}\right), \text{rank}_{\text{tbl}}\left(\underline{\mathbf{X}}\right)\right). \quad (2.33)$$

2.3.3. TRUNCATED TENSOR-SVD

The low tubal rank approximation of $\underline{\mathbf{Y}}$, denoted by $\hat{\underline{\mathbf{Y}}}$, can be obtained by truncating the SVD given in the fourth line of Algorithm 2 with $\text{rank}_{\text{tbl}} < \min(I_1, I_2)$ components and computing

$$\hat{\underline{\mathbf{Y}}} = \sum_{i=1}^{R} \breve{\underline{\mathbf{U}}}(:,i,:,\ldots,:) \star \breve{\underline{\mathbf{S}}}(i,i,:,\ldots,:) \star \breve{\underline{\mathbf{V}}}(:,i,:,\ldots,:)^{\mathsf{T}}. \quad (2.34)$$

The truncation of the t-SVD operation is illustrated in Fig. 2.5 for a 3rd-order tensor.

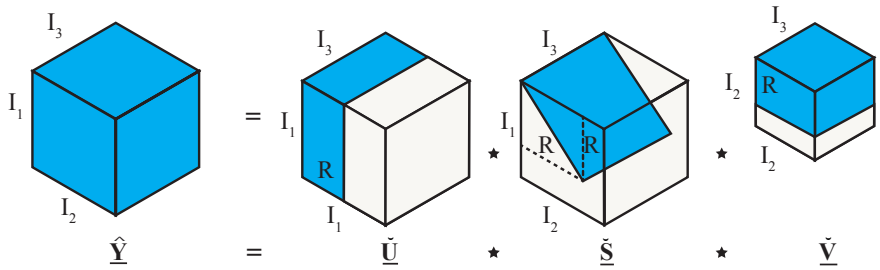


Figure 2.5: The truncated t-SVD defined in (2.34) for a 3rd-order tensor with tubal rank R.

For $R < \min(I_1, I_2)$, this truncation yields the optimal low tubal rank approximation in terms of minimizing the Frobenius norm of the residual error for 3rd-order tensors [22]. The same result is not provided for Nth-order tensors. This result is formally stated in the following theorem.

Theorem 2.3.1. Let the t-SVD of $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ be given by

$$\underline{\mathbf{Y}} = \underline{\check{\mathbf{U}}} \star \underline{\check{\mathbf{S}}} \star \underline{\check{\mathbf{V}}}^{\mathsf{T}}$$

and its truncation as in (2.34). Then, for $R < \min(I_1, I_2)$, we have

$$\underline{\hat{\mathbf{Y}}} = \arg\min_{\underline{\check{\mathbf{Y}}} \in M} \left\| \underline{\mathbf{Y}} - \underline{\check{\mathbf{Y}}} \right\|_F , \tag{2.35}$$

where the set M is defined via the t-product as

$$M = \left\{ \underline{\mathbf{C}} = \underline{\mathbf{A}} \star \underline{\mathbf{B}} \;\middle|\; \underline{\mathbf{A}} \in \mathbb{R}^{I_1 \times R \times I_3},\ \underline{\mathbf{B}} \in \mathbb{R}^{R \times I_2 \times I_3} \right\} .$$

We have provided the truncated t-SVD and the truncated MLSVD, which have a closed-form expression. If the ranks are not known, we can create a convex optimization problem such that a low tubal rank or multilinear rank approximation can be achieved. However, neither the tubal rank nor the multilinear rank expressions are convex expressions. In the following sections, we will introduce the convex relaxation of low multilinear rank and low tubal rank, such that a convex optimization algorithm can be constructed.

2.3.4. CONVEX RELAXATION OF LOW MULTILINEAR RANK

We will start by explaining the convex relaxation of the matrix rank and then generalize it to multilinear rank tensors. We can formulate an optimization problem to find a rank R approximation $\hat{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2}$ of a matrix $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2}$ as

$$\min_{\hat{\mathbf{Y}}} \left\{ \frac{1}{2} \left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\|_F^2 \right\} \quad \text{s.t.} \quad \operatorname{rank}(\hat{\mathbf{Y}}) = R . \tag{2.36}$$

The matrix rank is not a convex constraint. We can use the nuclear norm defined in Section 2.2.11 as the convex relaxation of the rank constraint and express (2.36) as

$$\min_{\hat{\mathbf{Y}}} \left\{ \frac{1}{2} ||\hat{\mathbf{Y}} - \mathbf{Y}||_F^2 + \gamma \left|\left|\hat{\mathbf{Y}}\right|\right|_* \right\}, \tag{2.37}$$

where γ is a penalty parameter related to the nuclear norm. The nuclear norm is the sum of the singular values of $\hat{\mathbf{Y}}$. By minimizing this sum, we are forcing the optimization algorithm to find a matrix that has a small sum of singular values, which indirectly forces a low rank expression. The expression in (2.37) can be solved using a proximity operator. A proximity operator is a generalization of the notion of projection in optimization. It is used to solve problems involving non-smooth convex functions efficiently [25]. Define the proximal operator of the nuclear norm with a penalty parameter γ as

$$\hat{\mathbf{Y}} = \text{prox}_{\gamma}^{\|\cdot\|_*}(\mathbf{Y}) = \arg\min_{\hat{\mathbf{Y}}} \left\{ \frac{1}{2} \|\hat{\mathbf{Y}} - \mathbf{Y}\|_F^2 + \gamma \|\hat{\mathbf{Y}}\|_* \right\}. \tag{2.38}$$

This operator (2.38) admits a closed-form solution via the soft-thresholding of the singular values of $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2}$ by γ [25]. Assuming that the SVD of $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the proximity operator returns

$$\hat{\mathbf{Y}} = \text{prox}_{\gamma}^{||\cdot||_*}(\mathbf{Y}) = \mathbf{U} \max(\mathbf{\Sigma} - \mathbf{I}\gamma, \mathbf{0}) \mathbf{V}^T, \tag{2.39}$$

where $\max(\cdot)$ is the elementwise maximum operator, $\mathbf{0}$ is a matrix consisting of 0s. Soft-thresholding is the operation where each element in the diagonal of $\mathbf{\Sigma}$ is reduced by γ, until the lowest possible value of 0 is obtained. This is done to prevent negative entries in the singular value matrix $\mathbf{\Sigma}$. The proximity operator in (2.39) is the solution of (2.37). Note that the low-rank approximation of a matrix given in (2.36) has an exact solution via truncated SVD. The convex relaxation proposed here is an algorithm that finds a low-rank approximation of $\mathbf{Y}$ using a tuning parameter γ. The only guarantee of (2.39) is that a monotonically lower rank approximation of $\mathbf{Y}$ is going to be achieved with an increasing γ.

Extending to Nth-order tensors, the multilinear low-rankness is also not a convex constraint. We can use the nuclear norm for the convex relaxation of low multilinear rank by summing the nuclear norms of each mode unfolding. For an Nth-order tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ we can express the sum of the nuclear norm (SNN) of mode-n unfoldings as

$$\left|\left|\underline{\mathbf{Y}}\right|\right|_{\text{SNN}} = \sum_{n=1}^{N} \left|\left|\mathbf{Y}_{(n)}\right|\right|_*. \tag{2.40}$$

A low multilinear rank approximation $\hat{\underline{\mathbf{Y}}}$ of tensor $\underline{\mathbf{Y}}$ can be found with the following objective function

$$\min_{\hat{\underline{\mathbf{Y}}}} \frac{1}{2} \left|\left|\hat{\underline{\mathbf{Y}}} - \underline{\mathbf{Y}}\right|\right|_F^2 + \gamma \left|\left|\hat{\underline{\mathbf{Y}}}\right|\right|_{\text{SNN}}, \tag{2.41}$$

where γ is a penalty parameter related to SNN. We can see that SNN on the right-hand side of (2.41) includes a summation of the nuclear norms of mode-n unfoldings, which are not separable. It has been proposed in [26] to separate the mode-n unfoldings in (2.41) using auxilliary variables $\hat{\underline{\mathbf{Y}}} = \hat{\underline{\mathbf{Y}}}^{(n)}$ for $n \in \{1, \ldots, N\}$,

and form the augmented Lagrangian using the Lagrangian parameters $\underline{\mathbf{\Gamma}}^{(n)}$ for $n \in \{1, \ldots, N\}$ and the Lagrangian penalty parameter ρ as

$$\min_{\underline{\hat{\mathbf{Y}}}, \underline{\hat{\mathbf{Y}}}^{(1)}, \ldots, \underline{\hat{\mathbf{Y}}}^{(N)}} \left(\frac{1}{2} \left\| \underline{\hat{\mathbf{Y}}} - \underline{\mathbf{Y}} \right\|_F^2 + \sum_{n=1}^{N} \left(\gamma \left\| \hat{\mathbf{Y}}^{(n)}_{(n)} \right\|_* + \left\langle \underline{\mathbf{\Gamma}}^{(n)}, \underline{\hat{\mathbf{Y}}}^{(n)} - \underline{\hat{\mathbf{Y}}} \right\rangle + \frac{\rho}{2} \left\| \underline{\hat{\mathbf{Y}}}^{(n)} - \underline{\hat{\mathbf{Y}}} \right\| \right) \right). \tag{2.42}$$

In this dissertation, we solve an ADMM problem in the augmented Lagrangian problem (2.42) as proposed in [27]. An alternative approach using Douglas-Rachford splitting [28] is proposed in [26]. Each subproblem related to nuclear norm in (2.42) takes the same form as the proximity operator in (2.38), applied to the mode-n unfolding. If we consider $\underline{\hat{\mathbf{Y}}}^{(1)}$, the optimization problem for this subproblem is

$$\min_{\hat{\mathbf{Y}}^{(1)}_{(1)}} \left\{ \gamma \left\| \hat{\mathbf{Y}}^{(1)}_{(1)} \right\|_* + \frac{\rho}{2} \left\| \hat{\mathbf{Y}}^{(1)}_{(1)} - \hat{\mathbf{Y}}_{(1)} \right\|_F^2 + \left\langle \mathbf{\Gamma}^{(1)}_{(1)}, \hat{\mathbf{Y}}^{(1)}_{(1)} - \hat{\mathbf{Y}}_{(1)} \right\rangle \right\}, \tag{2.43}$$

as the Frobenius norm and the inner product with the terms $\underline{\hat{\mathbf{Y}}}^{(1)}$ can be written exactly using the first mode unfolding $\hat{\mathbf{Y}}^{(1)}_{(1)}$. We can write (2.43) as

$$\min_{\hat{\mathbf{Y}}^{(1)}_{(1)}} \left\{ \frac{1}{2} \left\| \hat{\mathbf{Y}}^{(1)}_{(1)} - \left(\hat{\mathbf{Y}}_{(1)} - \frac{1}{\rho} \mathbf{\Gamma}^{(1)}_{(1)} \right) \right\|_F^2 + \frac{\gamma}{\rho} \left\| \hat{\mathbf{Y}}^{(1)}_{(1)} \right\|_* \right\}. \tag{2.44}$$

The expression in (2.44) has the same structure as (2.37). Hence, the solution can be found using the proximity operator given in (2.39) as

$$\hat{\mathbf{Y}}^{(1)} = \text{fold}_n \left(\text{prox}^{\|\cdot\|_*}_{\frac{\gamma}{\rho}} \left(\hat{\mathbf{Y}}_{(1)} - \mathbf{\Gamma}^{(1)}_{(1)} \right) \right). \tag{2.45}$$

The other unfoldings $n \in \{1, \ldots, N\}$ can be defined similarly. This is only the optimization part related to the nuclear norm. We need the primal updates related to $\underline{\hat{\mathbf{Y}}}$ and the dual updates related to the augmented Lagrangian parameters $\mathbf{\Gamma}^{(n)}$ for $n \in \{1, \ldots, N\}$. This form of alternating updates is called the alternating direction method of multipliers (ADMM) [29]. The optimization problem given in (2.42) can be solved using ADMM as proposed in [27]. We solve such a problem in Chapter 4.

2.3.5. CONVEX RELAXATION OF LOW TUBAL RANK

Similar to the low multilinear rank, we can obtain a convex relaxation for low tubal rankness using the t-SVD decomposition. For an Nth-order tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and its t-SVD decomposition

$$\underline{\mathbf{Y}} = \underline{\check{\mathbf{U}}} \star \underline{\check{\mathbf{S}}} \star \underline{\check{\mathbf{V}}}^{\mathsf{T}}, \tag{2.46}$$

with orthogonal tensors $\underline{\check{\mathbf{U}}} \in \mathbb{R}^{I_1 \times I_1 \times I_3 \cdots \times I_N}$, $\underline{\check{\mathbf{V}}} \in \mathbb{R}^{I_2 \times I_2 \times I_3 \times \cdots \times I_N}$, and the singular value tensor $\underline{\check{\mathbf{S}}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$, we have

$$\|\underline{\mathbf{Y}}\|_{\text{TNN}} = \frac{1}{I_3 \ldots I_N} \sum_{i_3=1}^{I_3} \cdots \sum_{i_N=1}^{I_N} \text{Tr} \left(\left(\underline{\check{\mathbf{S}}} \times_3 \mathbf{F}^{(3)} \cdots \times_N \mathbf{F}^{(N)} \right) (:, :, i_3, \ldots, i_N) \right). \tag{2.47}$$

The tubal nuclear norm (TNN) is the average sum of the singular values lying on the frontal slices of the one-dimensional Fourier transform of the core tensor $\underline{\mathbf{S}}$ along the modes $n \in \{3, \ldots, N\}$ [30][31]. Similar to (2.42), we can form a low tubal rank approximation function in the general form of a proximity operator as

$$\min_{\hat{\underline{\mathbf{Y}}}} \frac{1}{2} \left|\left|\hat{\underline{\mathbf{Y}}} - \underline{\mathbf{Y}}\right|\right|_F^2 + \gamma \left|\left|\hat{\underline{\mathbf{Y}}}\right|\right|_{\text{TNN}} . \tag{2.48}$$

We can solve (2.48) by assigning $\hat{\underline{\mathbf{Y}}}$ as

$$\text{Prox}_{\gamma}^{\|\cdot\|_{\text{TNN}}}(\underline{\mathbf{Y}}) = \underline{\check{\mathbf{U}}} \star \left(\max\left(\underline{\check{\mathbf{S}}} \times_3 \mathbf{F}^{(3)} \cdots \times_N \mathbf{F}^{(N)} - \gamma \underline{\mathbf{1}}, \underline{\mathbf{0}} \right) \times_3 \mathbf{F}^{(3)^{-1}} \cdots \times_N \mathbf{F}^{(N)^{-1}} \right) \star \underline{\check{\mathbf{V}}}^{\top}, \tag{2.49}$$

where $\gamma \underline{\mathbf{1}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is a tensor with all entries equal to γ, $\max(\cdot)$ is the elementwise maximum operator, and $\underline{\mathbf{0}}$ is a tensor whose elements are 0.

It is important to note that the core tensor in (2.49) is soft-thresholded after applying the one-dimensional Fourier transforms to the modes $n \in \{3, \ldots, N\}$, as described in Algorithm 2. The singular values of a complex matrix are nonnegative real numbers, so the soft-thresholding with zero as the smallest value is valid.

2.4. COMPARISON OF MLSVD AND T-SVD FRAMEWORKS

We summarize the key properties comparing MLSVD and t-SVD [10]:

- The core tensor $\underline{\check{\mathbf{S}}}$ in t-SVD is unique, while the core tensor $\underline{\mathbf{S}}$ in MLSVD is generally not unique. It is unique up to sign ambiguity if the mode-n singular values are distinct for $n \in \{1, \ldots, N\}$.
- The core tensor $\underline{\check{\mathbf{S}}}$ of t-SVD decomposition after the 1D Fourier transform in modes $n = \{3, \ldots, N\}$, has non-negative entries, while the core tensor $\underline{\mathbf{S}}$ of MLSVD can contain negative entries.
- The truncated MLSVD does not yield the optimal low multilinear rank approximation, whereas the truncated t-SVD provides the optimal low tubal rank approximation for 3rd-order tensors, as described in Theorem 2.3.1.
- Permuting $\underline{\mathbf{Y}}$ along the first two modes results in the same core tensor $\underline{\check{\mathbf{S}}}$ for t-SVD but with permuted factor tensors $\underline{\check{\mathbf{U}}}$ and $\underline{\check{\mathbf{V}}}$. However, permuting the first or second mode with any other mode greater than or equal to 3 results in a different t-SVD decomposition.
- The tubal rank R satisfies $R \leq \min(R_1, R_2)$. This implies that any tensor with low multilinear rank in the first two modes will also have low tubal rank. Extending this further, a tensor that has low multilinear rank in all modes will also exhibit low tubal rank under all possible permutations.

  For instance, consider the permuted tensors $\underline{\mathbf{Y}}^{\text{perm3}} \in \mathbb{R}^{I_3 \times I_1 \times I_2 \times I_4 \cdots I_N}$ and $\underline{\mathbf{Y}}^{\text{perm2}} \in \mathbb{R}^{I_2 \times I_3 \times I_1 \times I_4 \cdots \times I_N}$. The corresponding tubal ranks satisfy

  $$R^{\text{perm3}} \leq \min(R_3, R_1), \quad R^{\text{perm2}} \leq \min(R_2, R_3),$$

  where R^{perm2} and R^{perm3} are the tubal ranks of the respective permuted tensors. Therefore, any low multilinear rank tensor will also have a low tubal rank decomposition.

The intuitive reasoning for low tubal rank approximation is that DFT sparsifies the core tensor of the t-SVD decomposition. Most of the real data often has a limited bandwidth. Therefore, some frequency bands are expected to have less information, which can be truncated. We will introduce the application domain in the next chapter and motivate the usage of low multilinear rank and low tubal rank tensor approximations for DCEUS denoising.

3

ULTRASOUND FOR PROSTATE CANCER

3.1. ULTRASOUND IMAGING

Ultrasound imaging is a medical imaging technique that uses sound waves to unravel structures within the body. Since its initial development in the mid-20th century, ultrasound technology has revolutionized diagnostic medicine by providing a non-invasive, relatively affordable, and widely accessible means of visualizing soft tissues and organs. This section traces the historical evolution of ultrasound imaging, covering key technical advancements, including the development of B-mode imaging, Doppler imaging, and dynamic contrast-enhanced ultrasound imaging (DCEUS), which is the modality that is used in this dissertation. The DCEUS acquisitions are collected from patients enrolled in an ethically approved clinical trial for prostate cancer diagnosis and treatment monitoring. All imaging procedures and patient participation follow the protocols approved by the Institutional Review Board (IRB) under the International Recognition Procedure (IRP). Informed consent is obtained from each participant prior to data collection. We introduce the ultrasound machines and the settings that are used in such clinical trials. We analyze the DCEUS sequences for the diagnosis of prostate cancer by modeling the perfusion and the dispersion dynamics of the bubble movement. The Section 3.2 introduces this model and describes the cancer markers. We compare the extracted features with the ground truth obtained after the biopsy-proven client's prostate is annotated. Furthermore, we explain the procedures, such as the delineation, registration, and annotation of the prostate. In Sections 3.3 and 3.4 we bridge the gap between tensor denoising and DCEUS acquisitions.

3.1.1. EARLY DEVELOPMENTS IN ULTRASOUND TECHNOLOGY

The origins of ultrasound can be traced back to the late 19th and early 20th centuries, when scientists such as Pierre Curie and Jacques Curie discovered the piezoelectric effect in 1880 [32]. This discovery laid the groundwork for ultrasound transducers, as the piezoelectric effect converts electrical signals into sound waves, a principle essential for ultrasound technology. Later, in 1877, John

William Strutt (also known as Lord Rayleigh) published "The Theory of Sound," which showed that the waves generated via the piezoelectric crystals could be recorded such that the once invisible structures to the human eye could become visible [33]. During World War I and II, technology advancements were driven by the military need for submarine detection, further refining ultrasound's capabilities in detecting and interpreting sound waves [34].

In the 1940s, doctors and physicists began to apply ultrasound in the medical field. Dr. Karl Dussik, an Austrian neurologist, and his brother Friederich are credited as one of the first to use ultrasound for medical diagnosis, attempting to detect brain tumors using ultrasound transmission. However, the skull significantly attenuates ultrasound signals. The brothers' attempts were later realized in 1952 as reflections regarding bone thickness rather than brain tumors. This undermined confidence in the use of ultrasound for medical diagnosis [34]. The 1950s saw further advancements, particularly with Dr. John Wild, who pioneered the use of ultrasound for detecting tissue abnormalities in the human body. Wild's work demonstrated that ultrasound could effectively differentiate between healthy and cancerous tissues [35].

One of the most significant advancements in ultrasound technology was the development of B-mode, or brightness-mode, imaging in the 1950s. Early ultrasound systems, such as A-mode (amplitude mode), only displayed a 1st-order graph of amplitude versus time, which was useful for identifying the location of structures but not their spatial relationships. B-mode imaging, developed by Dr. Ian Donald and colleagues, introduced a 2nd-order grayscale display, where variations in the amplitude of reflected sound waves translated into different brightness levels on the screen.

B-mode imaging quickly became the cornerstone of diagnostic ultrasound, providing real-time, cross-sectional images of anatomical structures. It allowed for greater precision in detecting abnormalities within soft tissues. B-mode's visual capabilities allowed clinicians to observe fetal development, identify organ pathologies, and perform other diagnostic assessments with unprecedented clarity.

In the 1970s, Doppler imaging expanded the applications of ultrasound further. By using the Doppler effect, which is the change in frequency of sound waves as they reflect off moving objects, clinicians could measure and visualize blood flow within the body. Doppler ultrasound, particularly color Doppler, is now widely used for assessing vascular health, detecting blood clots, and monitoring cardiac function.

The advancements of 3rd-order and 4th-order ultrasound in the late 1990s and early 2000s marked a new era in ultrasound imaging. Unlike traditional 2nd-order ultrasound, which provides a single cross-sectional image, 3rd-order ultrasound compiles multiple 2nd-order images to create a volumetric representation of the anatomy.

3rd-order ultrasound found value in detailed imaging of the fetus, as well as in breast and cardiac imaging, providing enhanced visualization. With the addition of time, real-time 3rd-order imaging became possible, allowing practitioners to observe dynamic processes, such as fetal movements and cardiac cycles by analyzing a 4th-order tensor.

In 1988, the first contrast agents were introduced that were used in magnetic resonance imaging [36]. Subsequently, medical ultrasound research picked up

on the value of using ultrasound contrast agents (UCAs) and improved the visibility of tissue vascularity. UCAs are small bubble-encapsulated shells that are several micrometers in size. However, the stability of the contrast agents was a problem. The UCAs could not pass through the right atrium, the right ventricle, and the transpulmonary circulation. In the decade between 1990-2000, the so-called second generation of contrast agents was introduced, including the one used in this dissertation, SonoVue®. The phospholipid shell that encapsulates sulfur fluoride prevents the SonoVue bubbles from bursting in the commonly resonant frequencies around 3 MHz. We focus on dynamic contrast-enhanced ultrasound, which uses SonoVue UCAs to image the tumor vasculature due to its promising results in the diagnosis of prostate cancer.

3.1.2. DYNAMIC CONTRAST-ENHANCED ULTRASOUND

4th-order dynamic contrast-enhanced ultrasound is a non-linear imaging technique that enhances the non-linear echoes from the bubbles while suppressing the echoes from the tissue. Such a non-linear oscillation is achieved by the administration of UCAs. When the ultrasound wave resonates with the UCAs, the unequal extension and the retraction of the shell result in a non-linear echo. On the other hand, the surrounding areas, such as the tissue, respond linearly to the ultrasound wave. This is especially true at the low ultrasound pressures employed in DCEUS to minimize bubble destruction. Several techniques can enhance these non-linear echoes while suppressing the linear ones.

When a low-pressure sound wave is transmitted, the echoes from tissue do not generate any harmonics [37]. Therefore, we can utilize the harmonics of the fundamental frequency of the ultrasound wave to capture bubble-related information. The fundamental frequency is used in the receiver interfaces in the B-mode imaging. This frequency is denoted with f_0. In harmonic imaging, we still transmit the ultrasound waves with the fundamental frequency at f_0 but use frequencies such as $2f_0$ or $3f_0$ in the receiver end. Such filters are used in harmonic imaging [38]. An analogous extension is the subharmonic imaging [39] and ultraharmonic imaging [40]. In the former, half of the fundamental frequency is used, while in the latter, the frequency between the fundamental and the first harmonic is used at the receiver end. Another interesting approach to non-linear imaging is based on the transmission of coded pulses.

In coded pulse imaging, several ultrasound pulses (short waves) are transmitted consecutively. During beamforming, the received echoes are summed and scaled in such a way that the linear echoes are suppressed while the non-linear echoes are enhanced [41]. In pulse inversion (PI) [42], the second ultrasound pulse is the negative of the first one. Consider the first ultrasound wave as p_1. The second pulse that is transmitted is $p_2 = -p_1$. In amplitude modulation (AM) [43], the second pulse is half of the first pulse, that is, $p_2 = 1/2p_1$. By combining both techniques, the authors in [44] suggested the second pulse to be $p_2 = -1/2p_1$. This is called the pulse inversion amplitude modulation (PIAM) imaging. In the receiver end, we can merge the two pulses by

$$\begin{aligned} p_{PI} &= 1/2p_1 + 1/2p_2\,, \\ p_{AM} &= 1/3p_1 - 2/3p_2\,, \\ p_{PIAM} &= 1/3p_1 + 2/3p_2\,. \end{aligned} \tag{3.1}$$

Assuming that the scatterer responds linearly to the ultrasound wave, a change in the phase or the amplitude of the transmitted wave will result in a linear change in the received wave. If the scatterer responds non-linearly, such as in the case of UCAs, the received signal will have non-linear responses to the phase or amplitude changes. Therefore, the equations in (3.1) can be used to suppress the linear echoes while keeping the non-linear responses.

3.1.3. COMMERCIAL SCANNERS

There are other operations that the signal goes through before being recorded as DICOM data [45]. This is shown in Fig. 3.1. The TX-RX controller is responsible

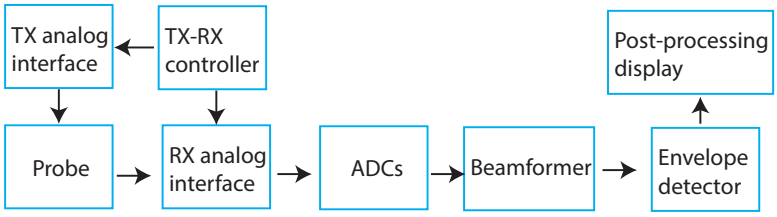


Figure 3.1: The operations the data goes through in a conventional scanner.

for generating the signal that will propagate inside the medium and which arrays in the transducer will be used for receiving the propagated signal. In the RX analog interface, one can use the PI, AM, or PIAM filters that enhance the non-linear echoes. The ADC is an analog-to-digital converter. The beamformer is responsible for delaying and summing the received echoes accordingly so that a signal in the focused direction is enhanced. The envelope detector returns the absolute value of the Hilbert transform of the RF signal. In the post-processing display, the log transform is taken using a predefined dynamic range.

Different companies use different hardware components, which causes a problem in the reproducibility of experiments [46]. In this dissertation, we use two versions of an ultrasound scanner from GE Healthcare. In both cases, the probe is the same. In Section 5, the recordings are captured on a LOGIQ E9 device, whereas in Section 6 the recordings are captured on a LOGIQ E10 device. The settings that were used by the machine are given in Table 3.1.

We fixed the dynamic range to 42 dB, the volume scan to the highest possible area to cover all the possible prostate sizes, and the image quality is 'low' for the E9 and 'Mid1' for the E10. The low image quality results in a 0.25-Hz frame rate for the E9 machine and a 1-Hz frame rate for the E10 machine. The mechanical index is a unitless metric representing the mechanical interaction between the ultrasonic waves and the bubbles. It is defined as the peak negative pressure in MPa divided by the square root of the central frequency given in MHz. The mechanical index (MI), also related to the power level, is set to 0.1 to minimize the bubble destruction. The endocavity transducer is driven at 4 MHz for the E9, whereas it is driven at 3.5 MHz for the E10 machine. The gain represents the DC component that is added to each voxel during the post-processing display. The gain is set to 55 dB for LOGIQ E10, while it is set to 30 dB for LOGIQ

Table 3.1: The settings that were used during the recordings.

	LOGIQ E10	LOGIQ E9
field of view:	120 deg	120 deg
depth:	50 mm	60-80 mm
image quality	Mid1 1 frame/s	Low 0.25 frame/s
sound wave frequency	3.6 MHz	4 MHz
gain ($a^{(2)}$)	55 dB	30 dB
dynamic range ($a^{(3)}$)	42 dB	42 dB
mechanical index	0.1	0.1
power level	10%	10%

E9. These settings were set by a common consensus among the urologists and Angiogenesis Analytics, who are the data providers.

A 2.4-mL bolus of SonoVue contrast agent is intravenously administered into the patient's arm, and a two-minute DCEUS sequence is recorded using a GE RIC5-9-D 3D/4D endocavitary probe. Within this time window, the UCAs are assumed to reach the prostate at least once. This contrast-enhanced acquisition provides insight into the vasculature of the prostate, which is of particular relevance in the context of prostate cancer.

3.2. PROSTATE CANCER

Prostate cancer is the most prevalent cancer among men after skin cancer. In 2024, approximately 300 000 new diagnoses are anticipated among American men, of which 35 000 are expected to be lethal [47]. A key indicator of cancer is angiogenesis [48–51]. Angiogenesis is the formation of new capillaries from pre-existing ones due to several pathological and physiological reasons [52]. It is not only related to prostate cancer, but it is a common process associated with wound healing, chronic inflammation, and maturation of the corpus luteum (a temporary endocrine gland related to ovulation). In physiological situations such as the maturation of the corpus luteum, angiogenesis is required for the formation of the capillaries that provide the nutrients and hormones to maintain the uterine environment. It is self-limited, i.e., after the pregnancy is finished, angiogenesis subsides or ends. In certain pathologic states that are not malignant, angiogenesis is prolonged but, again, self-limited. Examples are pyogenic granuloma, keloid formation, or retrolental fibroplasia [52]. On the other hand, in malignant cases such as cancer, the formation of angiogenesis continues until the host dies or the tumor is eradicated. The stages of angiogenesis until the metastasis are shown in Fig. 3.2.

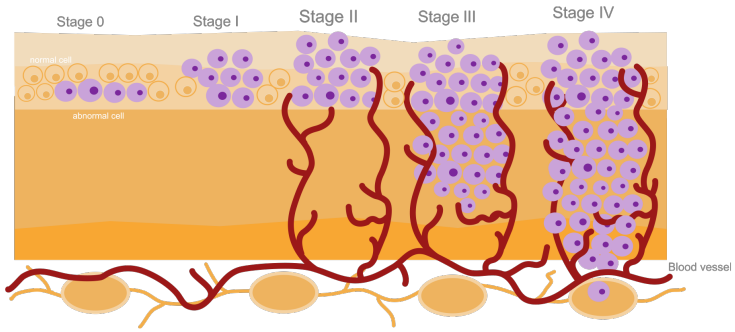


Figure 3.2: The stages of angiogenesis. [1]

Once a mutation occurs, like other angiogenetic stimuli such as wounds, the vessels around the tumor thicken, and the resting endolithic cells start containing an increased amount of ribosomes, endoplasmic reticulum, and Golgi apparatus. In fact, only the side of the vessel that faces the tumor thickens, while the other side remains thin. The nutrients start diffusing to the tumor. The membrane fragments and the neo-vessels start developing. Once they are elongated enough, microvascularization starts to feed the tumor further. The more aggressive tumor spreads through metastasis to find a new spot in the vascular system to restart the process.

The newly formed vessels have different characteristics from the healthy ones. For example, they are more tortuous and highly irregular [53][54]. Therefore, by imaging the UCAs trickling through the microvasculature, it is possible to differentiate the angiogenic and healthy vessels. It is possible to model the movement of UCAs using fluid dynamics. Once inserted into the system, the UCAs disperse and perfuse in the microvasculature. We can plot the time evolution of the UCAs through a voxel (a 3rd-order micro region) and create the time-intensity curves (TICs). We can consider the relation of the voxel with its surroundings through time and analyze the dispersion and perfusion. These features that describe the malignancy are the biomarkers of prostate cancer.

The biomarkers of prostate cancer used in this dissertation are extracted using the contrast-ultrasound dispersion imaging (CUDI) model. CUDI enables the extraction of several quantitative imaging features that facilitate the diagnosis of prostate cancer [55–69]. In essence, all the models approach the bubble dynamics as an anisotropically dispersive, multi-trajectory fluid flow in a porous medium [70] [71]. The system behavior is governed by both fluid dynamics and the Brownian motion of gas-filled microbubbles. Blood serves as the convective medium transporting the UCAs, a phenomenon modeled under the framework of convective-dispersion theory.

[1] Credit: Adobe Stock | #589239991

3.2.1. CONVECTIVE-DISPERSION MODELING

Consider a 4th-order tensor $\underline{\tilde{\mathbf{L}}}(i_1, i_2, i_3, i_4)$ that represents the intensity data over time from which the perfusion and dispersion characteristics of UCAs can be deduced. The first three modes represent the canonical 3rd-order Cartesian axes, and the last mode represents the time. There is a linear relation between the UCA concentration and the intensity $\underline{\tilde{\mathbf{L}}}(i_1, i_2, i_3, i_4)$ [55]. Hence, we can use $\underline{\tilde{\mathbf{L}}}(i_1, i_2, i_3, i_4)$ to capture the UCA concentration at a specific time and space. The UCA bubbles flow according to a convective-dispersion model [67][69]

$$\frac{\partial \underline{\tilde{\mathbf{L}}}}{\partial i_4} = \nabla \cdot (\mathbf{D} \nabla \underline{\tilde{\mathbf{L}}}) - \mathbf{v} \nabla \underline{\tilde{\mathbf{L}}}, \tag{3.2}$$

where $\mathbf{D} \in \mathrm{R}^{3x3}$ is the 2nd-order convective dispersion parameter and $\mathbf{v} \in \mathrm{R}^{3x1}$ is the velocity vector, ∇ is the differential in Cartesian domain, and $\nabla \cdot$ is the divergence. The continuous partial differential equation is provided in (3.2), which is commonly discretized using the frame-rate and the voxel size of the DCEUS acquisitions and solved for $\underline{\tilde{\mathbf{L}}}(i_1, i_2, i_3, i_4)$.

In 3rd-order space, the convective-dispersion spreading can occur at different rates along each mode, for example, faster along the i_1 direction than i_2, and can have cross-coupling effects, for example, the spreading along i_1 might depend on the gradients along i_2. For this reason, we use a matrix to represent $\mathbf{D}$. On the other hand, $\mathbf{v}$ is the bulk movement of the UCA, for example, due to the velocity of the blood, and can be represented with a vector. We can write the anisotropic locally constant convective dispersion parameter as

$$\mathbf{D} = \begin{bmatrix} d_{i_1 i_1} & d_{i_1 i_2} & d_{i_1 i_3} \\ d_{i_2 i_1} & d_{i_2 i_2} & d_{i_2 i_3} \\ d_{i_3 i_1} & d_{i_3 i_2} & d_{i_3 i_3} \end{bmatrix}. \tag{3.3}$$

On the other hand, the locally constant velocity vector can be written as

$$\mathbf{v} = [\, v_{i_1} \; v_{i_2} \; v_{i_3} \,]^T. \tag{3.4}$$

The dispersion parameter $\mathbf{D}$ is symmetric since $d_{i_1 i_2} = d_{i_2 i_1}$, $d_{i_1 i_3} = d_{i_3 i_1}$, and $d_{i_2 i_3} = d_{i_3 i_2}$. The local constancy of the $\mathbf{v}$ and $\mathbf{D}$ allows the estimation of these parameters through the analysis of $\underline{\tilde{\mathbf{L}}}(i_1, i_2, i_3, i_4)$ in a region, for example, a spherical shell with a radius of 3 mm, covering the early stages of an angiogenic tumor. Instead of tracking individual UCAs, the macroscopic movement of UCA concentration through time is analyzed using the model given in (3.2). For example, a spherical region around a voxel can be selected, and $\mathbf{D}$ and $\mathbf{v}$ can be estimated by analyzing the time-evolution of the UCA concentrations locally in that region. A spherical region is needed to cover any direction in which the UCAs can move in the Cartesian domain.

Several methods have been proposed that aim directly at solving (3.2). These are the 3D convective dispersion paper [69] and the system-identification paper [72]. The former assumes locally constant dispersion and velocity inside a spherical shell and solves (3.2) by approaching it as a least squares problem with L2-regularization. The latter approaches the problem by modeling the TICs $\underline{\tilde{\mathbf{L}}}(i_1, i_2, i_3, :)$ in a spherical shell as a linear dynamic system and finding the causal relation between the center voxel and the voxels in the spherical shell surrounding it. The dispersion and velocity are extracted from this relation. We use the

convective dispersion (CD) prefix to describe the features **D** and **v** as shown in Table 3.3.

The model fitting-based approaches solve the convective dispersion model described in (3.2) in 1*D*, assuming an infinitely long tube. The physiological parameters, such as the mean transit time and the skewness [73], are extracted from the fitted models. The solution to the 1D convective dispersion equation is the local density random walk (LDRW) model [74]. There are other models, such as the log-normal model [75] and the gamma-variate model [76], that model the perfusion characteristic of the bubble dynamics. These models are not based on a solution of the convective dispersion model [77], which is the focus of this dissertation.

The similarity-based metrics [63][57] compare the TICs from voxels that are neighbors in a spherical shell. The underlying hypothesis is that there is a positive correlation between the degree of similarity and the dispersion [61]. In cancerous regions, a high locally constant **D** and a high similarity of TICs are observed.

We will describe the time-intensity curve fitting in detail since it is used intensively as an assessment of how well the denoising algorithms perform.

3.2.2. LOCAL DENSITY RANDOM WALK MODEL

The modeling of the bubble dynamics is examined in [74], where the authors propose a spatially 1st-order characterization of the convective-dispersion process. Consider an infinitely long tube with a constant cross-section that imitates the circulatory system as described in Fig. 3.3.

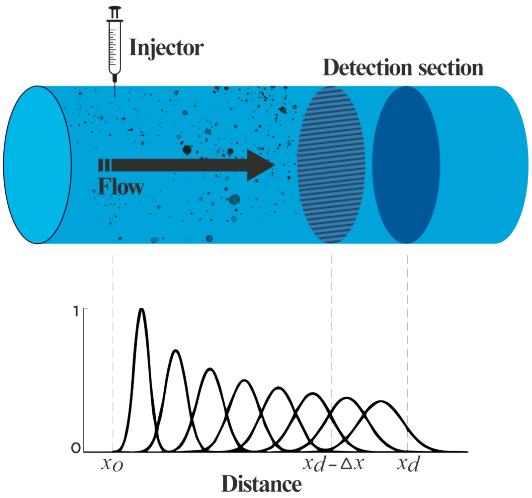


Figure 3.3: Modelling of the bubble dynamics in an infinitely long tube.

The UCAs are assumed to be introduced to the system fast and are modeled by

a Dirac function. The carrier in this tube is assumed to have a constant velocity v.

From 3rd-order space, we can have $\tilde{\mathbf{L}}(x, i_4)$, a spatially 1st-order characterization of (3.2). We use x as an arbitrary variable that represents any of the canonical Cartesian axes $x \in (i_1, i_2, i_3)$. The concentration function $\tilde{\mathbf{L}}(x, i_4)$ is a normal distribution that moves along the tube with the same velocity as the carrier v and spreads with a variance that is a linear function of time. This linear increase in the variance can be seen in Fig. 3.3. This is due to the Brownian motion of the UCAs (comparable to molecular diffusion) and the velocity gradients across the tube (flow profile). When imaging UCAs passing through a (micro)vascular network, like in the prostate, the UCA spatial spread is mostly due to the multi-path trajectories through the microvasculature. This process is typically referred to as apparent diffusion or dispersion [78], and it is denoted by D. If we assume that there is no dispersion and the bubbles move with constant velocity, we would see the same time evolution in all the detection sites as the one given just after the injection point x_0. Although denoted by $\tilde{\mathbf{L}}$, this function represents a scalar-valued concentration field evaluated over space and time, and is discretized into a matrix for computational purposes.

The spatially 1st-order convective dispersion is given by

$$\frac{\partial \tilde{\mathbf{L}}(x, i_4)}{\partial i_4} = D \frac{\partial^2 \tilde{\mathbf{L}}(x, i_4)}{\partial x^2} - v \frac{\partial \tilde{\mathbf{L}}(x, i_4)}{\partial x}. \tag{3.5}$$

The solution of the differential equation (3.5) gives a Gaussian distribution profile of UCAs at a location x that is sufficiently far away from the injection site [78]. However, the TIC $\underline{\tilde{\mathbf{L}}}(x, :)$ is a skewed distribution resulting from the temporal sampling of UCAs with a Gaussian concentration profile at the site x.

The authors in [55] solved (3.5) locally using the relation of the UCA concentration between $x_d - \Delta x$ and the detection site x_d. The result is independent of the distance; it only depends on time. The dispersion and velocity are assumed to be constant for the distance Δx. The model is named the local density random walk (LDRW) model, and can be expressed for any voxel $x \in (i_1, i_2, i_3)$ as

$$\tilde{\mathbf{I}}_x(i_4) = \alpha \sqrt{\frac{\kappa}{2\pi i_4}} e^{-(\kappa/2i_4)(i_4-\mu)^2}, \tag{3.6}$$

with the skewness related parameter $\kappa = \frac{v^2}{D}$, μ as the mean transit time, α as the area under the TIC. Note that $\tilde{\mathbf{I}}_x(i_4)$ is a scalar-valued function that represents the time evolution of each voxel in $\underline{\tilde{\mathbf{L}}}$. It is shown with the lowercase of $\underline{\tilde{\mathbf{L}}}$ as described in Section 2.1. Commonly, in ultrasound scanners, logarithmic compression is applied for visualization purposes. The logarithmic compression is expressed with

$$\underline{\mathbf{L}} = a^{(1)} \ln\left(\underline{\tilde{\mathbf{L}}}\right) + a^{(2)}, \tag{3.7}$$

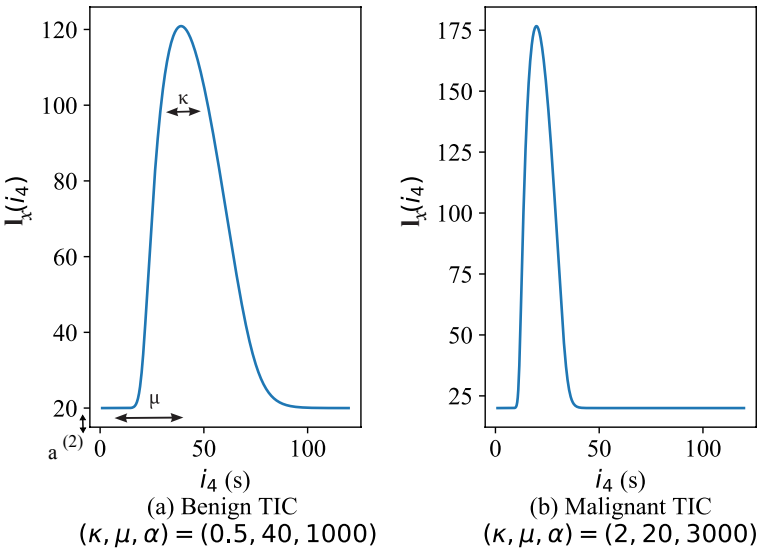
with a DC off-set $a^{(2)}$, log-compression related $a^{(1)} = 255 \log_{10}(e) 10/a^{(3)}$, the dynamic range $a^{(3)}$ in dB as described in Table 3.1, element-wise natural logarithm operator $\ln(\cdot)$ and the Euler's number e. The general tensor form of the log-compression is given in (3.7). Since all the operations are element-wise, it can be readily applied to the spatially 1st-order characterization $\tilde{\mathbf{L}}$.

The resulting model is called the modified local density random walk model (mLDRW). We can use mLDRW to fit the time evolution and have

$$\mathbf{I}_x(i_4) = a^{(1)} \ln\left(\alpha\sqrt{\frac{\kappa}{2\pi i_4}}\left(\exp-\left(\frac{\kappa(i_4-\mu)^2}{2i_4}\right)\right)\right) + a^{(2)}. \tag{3.8}$$

The TICs of the logarithmically compressed DCEUS acquisitions can be fitted with (3.8) to each voxel $x \in (i_1, i_2, i_3)$, and the parameters related to the UCA kinetics can be extracted.

We have plotted a malignant and benign TIC in Fig. 3.4. The benign TICs have a low κ, high μ, and low α. This can be observed in the subplot (a). The UCAs reach and leave a malignant tumor with a high angiogenic structure faster than benign regions. Due to the increased number of vessels, a higher concentration of UCAs reaches the malignant region as well. This property is reflected in subplot (b) with a high κ, low μ, and high α. We summarized the terminology used to define the TICs and the convective dispersion equations in Table 3.2.



(a) Benign TIC $(\kappa, \mu, \alpha) = (0.5, 40, 1000)$

(b) Malignant TIC $(\kappa, \mu, \alpha) = (2, 20, 3000)$

Figure 3.4: An example of a malignant and benign TIC plotted using (3.8).

Table 3.2: Abbreviations related to the time-intensity curves and the convective dispersion equation.

D	convective-dispersion
v	velocity
μ	mean transit time
κ	skewness
$\alpha^{(1)}$	scaling related to the dynamic range
$\alpha^{(2)}$	gain of the system
α	area under the TIC

Commonly, the first pass of the UCAs is fitted with (3.8). In addition to κ, μ, and α, several other features can be extracted from the TICs. Peak time (PT) is the time until a TIC reaches its peak, peak intensity (PI) is the maximum intensity of a TIC, appearance time (AT) is the time when TIC first reaches to 5% of its peak intensity, wash-in-time (WIT) is the difference between peak time and and the appearance time, wash-in-rate (WIR) is the ratio of peak intensity and the wash-in-time, root mean squared error (RMSE) is the square root of the norm of the TIC fitting error averaged over the total time. In this dissertation, we use the root mean squared error of such a fitting to assess how well the denoising algorithms perform. The model fitting is abbreviated from the other features with the prefix model-fitter (MFR) and incorporated into the summary of the CUDI features Table 3.3.

3.2.3. SPATIOTEMPORAL SIMILARITY ANALYSIS

The spatiotemporal similarity features are another cancer marker that is used in this dissertation. The similarity metrics are shown to correlate with the dispersion of the UCAs through the multipath trajectories of the vasculature. The time-evolution of a voxel $\tilde{\mathbf{I}}_{i_1 i_2 i_3}(i_4)$ is selected, and temporal correlation and spectral coherence are calculated with the neighboring voxels in a spherical shell in [61][56][57]. This is shown in Fig. 3.5. The inner diameter is selected as 1 mm, and the outer diameter is defined as 2.5 mm. The spherical shell kernel is chosen such that an omnidirectional convective-dispersion quantity is investigated. The inner radius is selected as a value greater than the system's resolution 0.8 mm. The clinically significant tumors are at least 0.5 cm^3. These radii are selected to cover the early stages of clinically significant prostate cancer and avoid the correlated signals within the system's resolution.

The average temporal correlation of the middle TIC to the neighboring TICs is calculated using the Pearson correlation coefficient [56]. Similarly, the average correlation in the Fourier Domain is calculated in spectral coherence, discarding the phase information. Finally, the mutual information is calculated in [65]. The temporal correlation encapsulates the whole time information, including the time shifts. The spectral coherence discards phase information; hence, it is invariant to time shifts. Mutual information extends the scope to non-linear relations. All three similarity metrics are used to distinguish malignant and benign voxels. The prefix similarity analysis (SA) is used to distinguish the metrics related to this

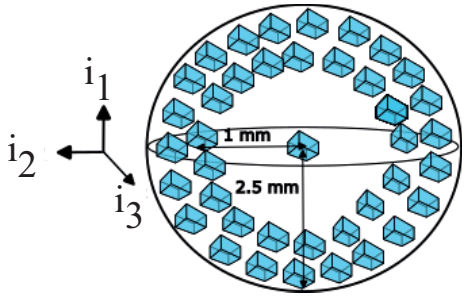


Figure 3.5: The spherical shell. The similarity analysis compares the TIC in the center with the TICs that reside between 1 mm and 2.5 mm radii.

section, and we incorporate it into the CUDI features Table 3.3.

3.2.4. PROSTATE ANNOTATIONS

The prostate cancer diagnosis is made using an annotated ground truth obtained through several steps: prostatectomy, delineation, and registration, as shown in Fig. 1.1. The prostate cancer patients sign an IRP. The patients with prostate cancer undergo radical prostatectomy, where the prostate is removed from the body. The pathologist fixes the prostate in formalin, cuts it into 4 mm slices, and delineates the cancerous regions. An image representing the slices can be seen in Fig. 3.6.

There is a mismatch between the digital reconstruction and the shape of the prostate inside the DCEUS acquisitions due to the shrinking of the prostate after its removal and fixation in formalin. For this reason, a registration process stretches and deforms the digital reconstruction given in Fig. 3.7 to the shape of the prostate inside the DCEUS acquisition. The digital reconstruction from Fig. 3.6 to Fig. 3.7 is done using radial basis functions [83] [84]. After this registration, the locations close to the prostate's voxels have been given a high probability of belonging to the annotations made by the pathologist. We define the ground truth by using the voxels with a high probability (>0.95) of belonging to a cancerous region.

We used two datasets from two clinical trials. The first trial was run in the Second Affiliated Hospital of Zhejiang University, located in Hangzhou, Zhejiang, China [2]. After obtaining the written consent, six patients underwent prostatectomy. The resulting delineation by the urologists resulted in $18{\cdot}10^3$ benign voxels and $25 \cdot 10^3$ malignant voxels. A LOGIQ E9 machine, with the settings provided in Table 3.1 was used. The second trial was done in two hospitals, the Amsterdam University Medical Center and the Netherlands Cancer Institute. After the International Recognition Procedure (IRP) approval, 94 patients underwent prostatectomy. The corresponding clinical trial is conducted by Auke Jager and

[2] Pintong Huang, Department of Ultrasound, Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China

Table 3.3: The abbreviations of the CUDI features, the chapters in which they are used, and the corresponding references in the literature.

	Abbreviation	References	Chapters
MFR-κ	Model-fitter, κ	[55] [74] [79] [54] [80]	5,6
MFR-μ	Model-fitter, μ		5,6
MFR-α	Model-fitter, α		5,6
MFR-WIT	Model-fitter, wash-in-time		5
MFR-WIR	Model-fitter, wash-in-rate		5
MFR-PI	Model-fitter, peak intensity		5
MFR-PT	Model-fitter, peak time		5
MFR-AT	Model-fitter, appearance time		5
CD-D	Convective dispersion, dispersion	[69] [72] [81]	6
CD-v	Convective dispersion, velocity		6
VE-Ev	Velocity entropy, entropy	[81]	6
VE-CEv	Velocity entropy, conditional entropy		6
SA-ρ	Similarity analysis - spectral coherence	[61] [56] [56]	6
SA-r	Similarity analysis - pearson correlation coefficent		6
SA-MI	Similarity analysis, mutual information	[65]	6

presented in [85]. The urologists delineated the locations of the malignant and benign voxels. There were $70 \cdot 10^7$ benign voxels and $20 \cdot 10^4$ malignant voxels. The same LOGIQ E10 machine was used in this trial with the settings provided in Table 3.1. The summary is given in Table 3.4.

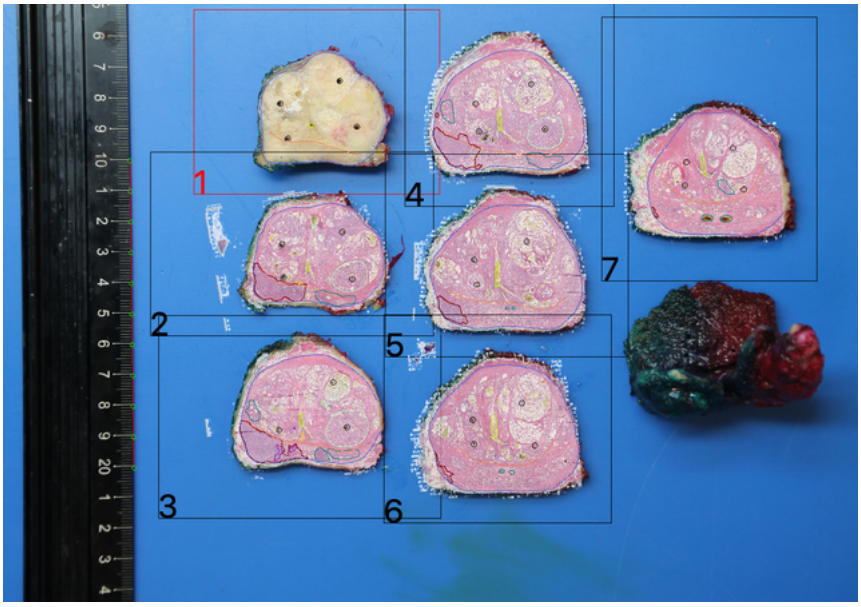


Figure 3.6: Slicing made by the pathologist [82] and the overlayed annotations. The slice numbers 1 to 7 represent each cut made to the prostate from the apex to the base. The first slice is commonly discarded.

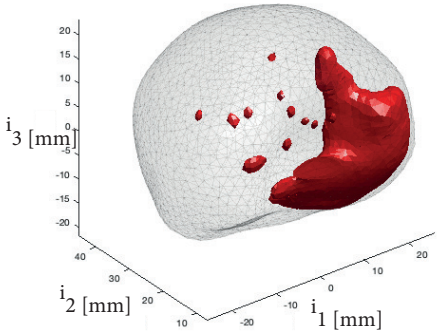


Figure 3.7: The surface mesh with the annotation of clinically significant prostate cancer is shown in bold red [83].

Table 3.4: The summary of the clinical trials.

Hospital	Number of Patients	Number of Malignant Voxels	Number of Benign Voxels	US Scanner
Second Affiliated Hospital of Zheijang University	6	$25 \cdot 10^3$	$18 \cdot 10^3$	LOGIQ E9
Amsterdam University Medical Center and Netherlands Cancer Institute	94	$20 \cdot 10^4$	$70 \cdot 10^7$	LOGIQ E10

3.3. SPECKLE IN DYNAMIC CONTRAST-ENHANCED ULTRASOUND

The dynamic contrast-enhanced ultrasound data is a 4th-order tensor, where the first three modes are the spatial domain, and the last mode is the time domain. After the operations described in the commercial scanners section in Section 3.1.3 and the corresponding Fig. 3.1, the Digital Imaging and Communications in Medicine (DICOM) data is received. This data is the log-compressed intensities (3.7) that are linearly correlated with the UCA concentration [55]. Hence, all the prostate cancer biomarkers described in Section 3.2 are extracted using the intensities from the DICOM data as a linearly correlated measure of the UCA concentration.

The DCEUS data pipeline, starting from the DICOM extract, is depicted in Fig. 1.1. The spatial information of DICOM data has a spherical format, which includes azimuth, elevation, and axial information. We transform the spherical domain to a Cartesian domain using $0.25 \times 0.25 \times 0.25$ mm voxel sizes following the procedures defined in [64], where a linear interpolation is applied to reach the given resolution. We downsampled by three voxels such that 0.75 mm voxel size is obtained. The spatial resolution of the DCEUS acquisitions is fixed when the axial distance from the probe is considered. However, with an increased elevation and azimuth, the spatial resolution decreases. The procedure to estimate the resolution for 3rd-order and 4th-order DCEUS is explained in [61] and [64], respectively. The resolution is equalized through space using a varying Wiener deconvolution filter, such that a resolution of approximately 1.6 mm is obtained [62]. The selected voxel size is approximately half of the equalized resolution. After the aforementioned steps, the speckle noise still persists due to the coherent imaging of UCA in one resolution cell.

The speckle noise in DCEUS sequences can be recognized with a granular appearance, such as the example given in Fig. 3.8. We have plotted the $i_1 i_3$ slices of a DCEUS acquisition at three separate UCA states. The early appearance time is defined as the average time when the microbubbles reach 5% of their peak intensity. The peak intensity is the average time that the microbubbles reach their highest intensity. Finally, the wash-out time is the time when the average intensity drops to 50% of the peak intensity. The speckle noise is additive in the log-compressed DCEUS acquisitions plotted in Fig. 3.8. The speckle noise is best seen in the early appearance time, on the edge of the slices. The observed intensities are speckles added on top of the gain $a^{(2)}$.

A comprehensive review of the statistics of speckle noise under various scenarios, such as speckle noise with strong reflectors and uniformly distributed

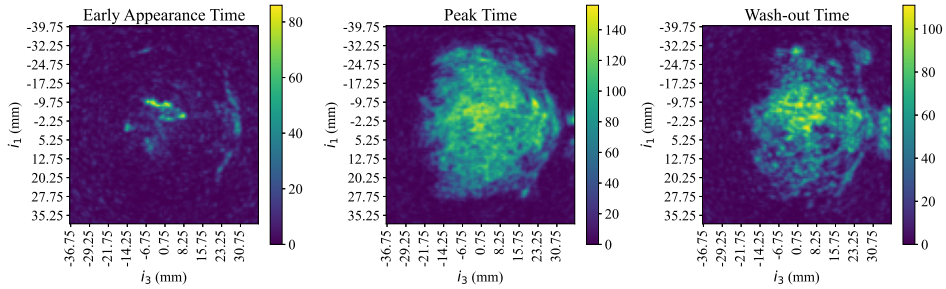


Figure 3.8: The $i_1 i_3$ slices at $i_2 = 20$ mm of a DCEUS acquisition of a cancerous prostate. The early appearance time, the peak time, and the wash-out time of UCAs are shown from left to right.

reflectors, is presented in Chapter 7 of the PhD dissertation [86]. There are several probability distributions, such as the K-distribution [87], Nakagami distribution [88], Generalized Gamma distribution [89], or Rician distribution [90]. In this dissertation, we assume the fully-developed speckle noise scenario and model the speckle noise as Rayleigh distributed with a scaling parameter of 1 that is multiplied before the log transform by the true intensity. Let $\underline{\tilde{\mathbf{Y}}}$ be the noisy DCEUS tensor and $\underline{\tilde{\mathbf{M}}}$ be the speckle noise tensor, where each element $\tilde{m}_{i_1,i_2,i_3,i_4} \sim$ Rayleigh(1) is Rayleigh distributed with a scaling parameter 1. We can express the noisy DCEUS tensor as

$$\underline{\tilde{\mathbf{Y}}} = \underline{\tilde{\mathbf{L}}} \odot \underline{\tilde{\mathbf{M}}}, \tag{3.9}$$

where each TIC $\underline{\tilde{\mathbf{L}}}$ follows the LDRW model given in (3.6). The signal model (3.9) where each element of $\underline{\tilde{\mathbf{L}}}$ is multiplied with a Rayleigh distributed speckle noise with a scaling parameter 1 comes from the relation between the mean-scattering amplitude of the particles in the resolution cell and the scaling parameter of the Rayleigh function [91]. The greater the number of scatterers, the greater the scaling parameter of the Rayleigh distribution. We estimate UCA concentrations related to the scaling parameter. Hence, we fix the scaling parameter to 1, and estimate the parameter that is multiplied by the random variable.

The log-compression is applied in commercial scanners before visualization, as shown in Fig. 3.1. After the log-compression (3.7) the noisy DCEUS tensor $\underline{\mathbf{Y}}$ can be expressed as

$$\begin{aligned} \underline{\mathbf{Y}} &= a^{(1)} \ln(\underline{\tilde{\mathbf{L}}}) + a^{(1)} \ln(\underline{\tilde{\mathbf{M}}}) + a^{(2)} \\ &= \underline{\mathbf{L}} + \underline{\mathbf{M}} + a^{(2)}. \end{aligned} \tag{3.10}$$

The log compression $m_{i_1,i_2,i_3,i_4} = a^{(1)} \ln(\tilde{m}_{i_1,i_2,i_3,i_4}) + a^{(2)}$ changes the probability density function of the noise [11]. The summation with the constant $a^{(2)}$ only shifts the probability distribution without changing its shape. So we ignore $a^{(2)}$, and focus on the transformation $m_{i_1,i_2,i_3,i_4} = a^{(1)} \ln(\tilde{m}_{i_1,i_2,i_3,i_4})$

The Rayleigh probability distribution with a scaling parameter 1 is denoted by

$p(\tilde{m}_{i_1,i_2,i_3,i_4};1)$ and can be expressed as

$$p(\tilde{m}_{i_1,i_2,i_3,i_4};1) = \tilde{m}_{i_1,i_2,i_3,i_4} \exp(-\tilde{m}^2_{i_1,i_2,i_3,i_4}), \quad \tilde{m}_{i_1,i_2,i_3,i_4} > 0, \tag{3.11}$$

where exp is the element-wise exponentiation operator. Let the Nth-order tensor $\underline{\tilde{\mathbf{M}}}$ represent the noise tensor, where each element follows a Rayleigh distribution with a scaling parameter of 1. The probability distribution $p(\underline{\tilde{\mathbf{M}}};1)$ becomes

$$p(\underline{\tilde{\mathbf{M}}};1) = \prod_{i_1=1}^{I_1} \cdots \prod_{i_4=1}^{I_4} \tilde{m}_{i_1,\ldots,i_4} \exp(\frac{-\tilde{m}^2_{i_1,\ldots,i_4}}{2}), \tag{3.12}$$

due to the statistical independence between voxels. For each element of the tensor $\underline{\mathbf{M}}$, we have

$$p(m_{i_1,i_2,i_3,i_4};1) = p(\tilde{m}_{i_1,\ldots,i_4};1)(\frac{\partial \tilde{m}_{i_1,\ldots,i_4}}{\partial m_{i_1,i_2,i_3,i_4}}). \tag{3.13}$$

The derivative of $\tilde{m}_{i_1,i_2,i_3,i_4}$ with respect to m_{i_1,i_2,i_3,i_4} is

$$\frac{\partial \tilde{m}_{i_1,i_2,i_3,i_4}}{\partial m_{i_1,i_2,i_3,i_4}} = \frac{\tilde{m}_{i_1,i_2,i_3,i_4}}{a^{(1)}}. \tag{3.14}$$

We can plug (3.14) into (3.13), resulting in

$$\begin{aligned} p(m_{i_1,i_2,i_3,i_4};1) &= \frac{\tilde{m}_{i_1,i_2,i_3,i_4}}{a^{(1)}} \tilde{m}_{i_1,i_2,i_3,i_4} \exp(\frac{-\tilde{m}^2_{i_1,i_2,i_3,i_4}}{2}), \\ &= \frac{\tilde{m}^2_{i_1,i_2,i_3,i_4}}{a^{(1)}} \exp(\frac{-\tilde{m}^2_{i_1,i_2,i_3,i_4}}{2}). \end{aligned} \tag{3.15}$$

If we plug the relation $\tilde{m}_{i_1,i_2,i_3,i_4} = \exp(m_{i_1,i_2,i_3,i_4}/a^{(1)})$ to (3.15), we will have the Fisher-Tippett distribution with the scaling parameter $a^{(1)}$

$$p(m_{i_1,i_2,i_3,i_4};1) = \frac{\exp\left(2m_{i_1,i_2,i_3,i_4}/a^{(1)} - \frac{\exp(2m_{i_1,i_2,i_3,i_4}/a^{(1)})}{2}\right)}{a^{(1)}}. \tag{3.16}$$

For a 4th-order tensor $\underline{\mathbf{M}}$ we have

$$p(\underline{\mathbf{M}};1) = \prod_{i_1=1}^{I_1} \cdots \prod_{i_4=1}^{I_4} \frac{\exp\left(2m_{i_1,i_2,i_3,i_4}/a^{(1)} - \frac{\exp(2m_{i_1,i_2,i_3,i_4}/a^{(1)})}{2}\right)}{a^{(1)}}, \tag{3.17}$$

due to the statistical independence between the voxels. A histogram that illustrates the probability distribution of Fisher-Tippet is given in Fig. 3.9. The Fisher-Tippett distribution has an asymmetrical left tail and a symmetric Gaussian-like shape around 0. Commonly, the speckle noise is modeled as WGN with sparse outliers [92]. The sparse outliers here would be the left tail, which has a low probability of occurring and has values greater than 3 standard deviations of the distribution. Out of 10^5 realizations shown in Fig. 3.9, approximately 5 percent of realizations are found to be 3 standard deviations beyond the mean.

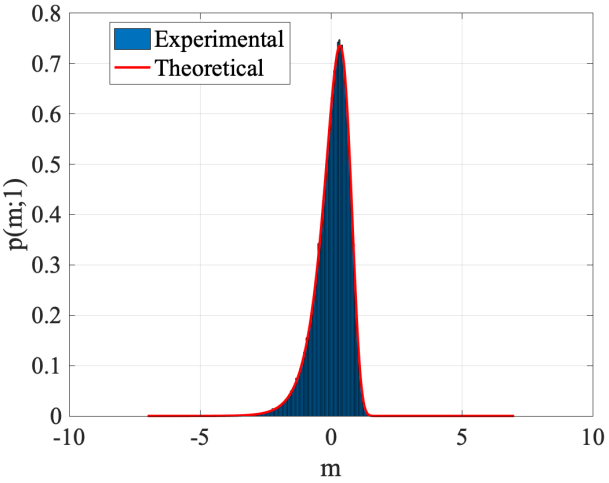


Figure 3.9: The histogram of the Fisher-Tippet (Log Rayleigh) distribution $p(m;1)$ given in (3.16) with m on the x-axis and the probability density on the y-axis. The experimental probability distribution is the histogram of 10^5 realizations of the logarithm of the Rayleigh distribution with scaling parameter 1. The theoretical probability distribution is plotted using (3.16) and 10^5 m values equally spaced between -7 and 7. Here $\alpha^{(1)}$ is selected as 1.

Up until now, we have explained tensor algebra, tensor decompositions, prostate cancer, and the DCEUS processing pipeline. We aim to despeckle the spatially equalized DCEUS acquisitions in the Cartesian format using orthogonal tensor decomposition methods based on t-SVD and MLSVD framework while incorporating the noise statistics. In the next section, we will motivate the usage of low tubal rank and multilinear rank tensor decomposition methods for modeling DCEUS.

3.4. LOW TUBAL RANK AND MULTILINEAR RANK MODELING OF DCEUS

This dissertation is motivated by the application of SVD to the Casorati matrix of DCEUS, which is in fact the fourth mode unfolding of the tensor. Define the 4th-order tensor $\underline{\mathbf{Y}}$ as the DCEUS sequences. The authors in [93] applied SVD to $\mathbf{Y}_{(4)}$ and kept the highest few singular values corresponding to the UCA kinetics. The flattening of $\underline{\mathbf{Y}}$ into $\mathbf{Y}_{(4)}$ removes the spatial structure provided in the first three modes, the spatial domain. We postulated that the spatial domain has a correlation that can be incorporated into the low-rank approximation. We will investigate the MLSVD and the t-SVD framework to achieve this.

The insight behind the structured spatial domain comes from several features

that are used as markers of prostate cancer. The similarity analysis described in Section 3.2.3 shows that the clinically significant prostate cancer tumors that are grown at least 0.5 cm^3 in diameter show high levels of similarity in a spherical shell. The angiogenesis creates a chaotic network of microvasculature around the tumor, wherein UCAs perfuse and disperse similarly. Instead of imaging individual bubbles, which is called super resolution imaging in the ultrasound community [94], in this dissertation, we focus on a macro scale. We aim at diagnosing clinically significant prostate cancer, which has grown in size and malignancy. These regions create structured regions that can be modeled as low-rank.

In MLSVD, we look at four flattenings of $\underline{\mathbf{Y}}$ and recover a spatially and temporally low multilinear rank tensor. The spatial flattenings are related to the views with respect to the canonical basis in the Cartesian system. The first mode i_1 represents the x-axis, the second mode i_2 represents the y-axis, the third mode i_3 represents the z-axis, and the fourth mode i_4 represents the time. The flattening with respect to i_1 stacks the information in the x-axis and flattens the information in the i_2, i_3, and i_4 modes. The other mode-n unfoldings for $n \in \{1, \ldots, 4\}$ can be defined similarly. The following SVD operation approaches to capture the structures in the x axis. We can either iteratively approach this low multilinear rank approximation through SNN as described in Section 2.3.4, or estimate the ranks and apply tr-MLSVD, which is described in Section 2.3.1. We are going to investigate such approaches in Chapters 4 and 5. The authors in [13] show that the MLSVD investigates the highest contributions in each mode-n unfolding and is a good approach to the multilinear rank modeling. However, we can see a similar problem in terms of flattening and the accompanying loss of structure. The mode-1 unfolding looks at an x-axis view but flattens the y-axis, z-axis, and time modes. The mode-2 unfolding looks at a y-axis view, while simultaneously flattening the other modes. The question regarding keeping the tensor structure intact and applying tensor operations results in the t-SVD framework.

In low tubal rank modeling of the DCEUS, we use the Fourier Transform as a building block. The motivation is to capture the low-frequency contents that allow us to look at major scale structures. In [74], the TICs are shown to have the majority of the information below 0.5 Hz. This motivates us to model temporally the DCEUS sequences as band-limited information with the majority of the information in the low-frequency regime. Spatially, we can propose a similar motivation where the clinically significant prostate cancer tumors tend to grow in space and show similar characteristics. We can model this as a low spatial frequency. The t-SVD framework explained in Section 2.3.2 can be applied to the natural tensor ordering of the 4th-order DCEUS acquisitions. We can apply a 1D Fourier transform to the 3rd and the 4th modes. Then we are limited to the investigation of the frequency contents only in two modes. Since the t-SVD framework is orientation-dependent, in the sense that different permutations of the $\underline{\mathbf{Y}}$ result in different t-SVD decompositions, we can investigate the different permutations of $\underline{\mathbf{Y}}$ and apply the low tubal rank approximation. We investigate such an approach in Chapter 6.

4

COMPARISONS OF CONVEX OPTIMIZATION ALGORITHMS FOR SPECKLE DENOISING: A SIMULATION STUDY

In this chapter, we run simulations of denoising a 3rd-order low multilinear rank tensor with speckle noise expressed in (3.13) by modeling it as WGN with sparse outliers, using the MLSVD and t-SVD framework explained in Section 2.3. This chapter answers the first research question.

4.1. INTRODUCTION

Speckle noise is an inherent characteristic of coherent imaging modalities occurring in various systems, including synthetic aperture radar [95], sonar [96], and ultrasound imaging [92]. This phenomenon arises due to the superposition of waves reflected from a resolution cell, which is defined by the bandwidth of the transmitted signal. When enough scatterers have sizes comparable to the transmitted signal's wavelength and the system's point spread function is broader, the received signal exhibits a granular appearance. Speckle noise is commonly reduced to enhance the visibility of structures and improve the quantification of the underlying system's dynamics.

In many applications, speckle noise is assumed to be multiplicative [95][96][92]. Several techniques were proposed in the literature to remove multiplicative noise. State-of-the-art methods can be categorized as the total variation (TV) regularization [97] [98], and nonlocal low-rank-based methods [96] [99]. Total regularization methods optimize a loss function that includes a data fidelity term and a total variational term. The data fidelity term provides a least squares approximation,

This chapter is based on the article published as "Denoising of the Speckle Noise by Robust Low-rank Tensor Decomposition" by M. Calis, B. Hunyadi, European Signal Processing Conference (EUSIPCO), 2024, 32

while the total variation regularization term smoothes the data while preserving the edges. The nonlocal low-rank-based methods first cluster the data into similar patches and apply denoising by returning a low-rank approximation of each patch. This low-rank approximation is made through low multilinear rank-based algorithms such as higher-order orthogonal iteration (HOOI) [96] or a convex relaxation of the low multilinear rankness [100]. However, these algorithms are better suited for white Gaussian noise (WGN). The aforementioned techniques can be improved by changing the WGN assumption about the speckle noise.

In [92], the log-transformed Rayleigh distribution is approximated as WGN with sparse outliers. The sparse outliers have a detrimental effect on the low multilinear rank approximation of the data [26]. The robust tensor decomposition methods recover the low multilinear rank tensor while capturing noise as a sparse and additive term. In most algorithms, the sparsity constraint is relaxed using the L1 norm. The convex relaxation of the tensor low multilinear and tubal rankness differs and can be categorized into the sum of the nuclear norm-based (SNN) and the sum of tubal nuclear norm-based (TNN) methods as defined in Chapter 2. In [101], an SNN method is used as a robust tensor decomposition method. Assuming that the tensor is not simultaneously low multilinear rank in all modes, the authors in [102] introduced a new low multilinear rank relaxation that is better suited.

The authors in [103] introduced TNN as a tighter convex surrogate for tubal rank. Using TNN, the authors in [104] solved the robust tensor decomposition problem. In [30], the authors extended the TNN to consider low tubal rankness in all modes simultaneously and introduced an orientation invariant tubal nuclear norm (OITNN-O) method. If the tensor is not low tubal rank in all modes, a method named OITNN-L is introduced. TNN-based algorithms are shown to perform better than their SNN counterparts. This chapter reviews SNN and TNN-based robust tensor decomposition methods and applies them in the context of speckle denoising. In addition, we propose a weighting scheme for SNN when the underlying ranks are known. In parallel with the truncated multilinear singular value decomposition (tr-MLSVD), the proposed algorithm preserves the predominant singular values while subjecting the remaining ones to soft thresholding. By incorporating a sparsity constraint, our objective is to examine the effect of sparsity in denoising speckle noise.

In Section 4.2, we explain the notation. In Section 4.3, we formulate the problem. In Section 4.5, we introduce the proposed algorithm. In Section 4.7 and Section 4.8, we compare the performance of the proposed algorithm to other methods in the literature and report the results. Finally, in Section 4.9, we discuss the results and propose possible future work.

4.2. NOTATION AND TENSOR PRELIMINARIES

Tensors are represented by underlined boldface letters such as $\underline{\mathbf{Y}}$. Matrices are represented by boldface letters such as $\mathbf{I}$. The numbers given as superscripts in parentheses refer to the different matrices or tensors that share a similar property. An example could be the relation $\underline{\mathbf{K}} = \sum_{n=1}^{N} \underline{\mathbf{K}}^{(n)}$, where $\underline{\mathbf{K}}$ is written as sum of N tensors. Vectors are represented with boldface lowercase letters such as $\mathbf{t}_{ij}$ that represent the (i, j)th vector of $\underline{\mathbf{T}} \in \mathbb{R}^{I \times J \times K}$. Scalars are represented by lowercase letters such as a_{ij} that represent the element at the ith row and jth column

of $\mathbf{A} \in \mathbb{R}^{I \times J}$. The mode-$n$ unfolding of $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is $\mathbf{Y}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N}$. Mode-$(n, n+1)$ unfolding of a tensor is described further in the chapter and shown with $\underline{\mathbf{Y}}_{[n]} \in \mathbb{R}^{I_n \times D/(I_n I_{n+1}) \times I_{n+1}}$, where $D = \prod_n I_n$. The Hadamard product is shown with $\circledast$. The Frobenius norm is the square root of the sum of each element and is shown by $||\cdot||_F$. The cardinality, defined as the number of non-zero elements, is shown with $||\cdot||_0$. The nuclear norm is shown with $||\cdot||_*$. The penalty parameter of the augmented Lagrangian is defined as ρ. The superscript t without the parentheses is used to define the state of a tensor in an iterative optimization algorithm. For example, $\underline{\mathbf{K}}^t$ is the state of $\underline{\mathbf{K}}$ at tth iteration in a maximum number of T iterations. More information regarding tensor algebra is provided in Chapter 2.

4.3. PROBLEM FORMULATION

Let $\tilde{\underline{\mathbf{Y}}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ denote the Nth-order tensor to be denoised. We model the tensor $\tilde{\underline{\mathbf{Y}}}$ as the element-wise multiplication of the low tubal rank or low multilinear rank tensor $\tilde{\underline{\mathbf{L}}}$ and the noise tensor $\tilde{\underline{\mathbf{M}}}$, that is,

$$\tilde{\underline{\mathbf{Y}}} = \tilde{\underline{\mathbf{L}}} \circledast \tilde{\underline{\mathbf{M}}}, \tag{4.1}$$

where each element of $\tilde{\underline{\mathbf{M}}}$ is modeled by the Rayleigh noise with a scaling parameter of 1. Taking the element-wise logarithm of (4.1) transforms the multiplicative noise into additive noise. Let the natural logarithms $\ln \tilde{\underline{\mathbf{Y}}}$, $\ln \tilde{\underline{\mathbf{L}}}$ and $\ln \tilde{\underline{\mathbf{M}}}$ be denoted with $\underline{\mathbf{Y}}$, $\underline{\mathbf{L}}$ and $\underline{\mathbf{M}}$, respectively. Each element of $\underline{\mathbf{M}}$ follows the Fisher-Tippett distribution.

In [92], using Taylor expansion, the Fisher-Tippett distribution is approximated as WGN with sparse outliers. With these assumptions, we can model the received tensor $\underline{\mathbf{Y}}$ as the sum of a low multilinear rank tensor $\underline{\mathbf{L}}$, sparse outlier $\underline{\mathbf{O}}$ and WGN $\underline{\mathbf{W}}$, that is,

$$\begin{aligned} \ln(\tilde{\underline{\mathbf{Y}}}) &= \ln(\tilde{\underline{\mathbf{L}}}) + \ln(\tilde{\underline{\mathbf{M}}}), \\ \underline{\mathbf{Y}} &= \underline{\mathbf{L}} + \underline{\mathbf{M}}, \\ \underline{\mathbf{Y}} &\approx \underline{\mathbf{L}} + \underline{\mathbf{O}} + \underline{\mathbf{W}}. \end{aligned} \tag{4.2}$$

This chapter uses tensor approaches to compare convex methods to recover $\underline{\mathbf{L}}$ from the speckled $\underline{\mathbf{Y}}$. For $N > 2$, and ranks $R_n < I_n$ for $n \in \{1, \ldots, N\}$, this problem can be formulated as

$$\begin{aligned} &\min_{\underline{\mathbf{L}}, \underline{\mathbf{O}}} \left\{ \frac{1}{2} \| \underline{\mathbf{L}} + \underline{\mathbf{O}} - \underline{\mathbf{Y}} \|_F^2 + \lambda \| \underline{\mathbf{O}} \|_0 \right\} \\ &\text{s.t.} \quad \operatorname{rank}\left(\mathbf{L}_{(n)}\right) = R_n, \quad \text{for } n \in \{1, \ldots, N\}. \end{aligned} \tag{4.3}$$

The objective function given in (4.3) is not convex. A common convex relaxation of the cardinality is the L1 norm. For the convex relaxation of the tensor ranks as defined in Chapter 2, we investigate two approaches, namely SNN and TNN. In the following section, we will introduce SNN and TNN. In addition, we will introduce a weighted SNN proximity operator, which can be used if the multilinear ranks of the tensor are known. We aim to apply various convex relaxations to solve (4.3) and compare the results in the context of speckle denoising for various SNRs, which has not been done before. Since the convex optimization

problems for low multilinear rankness and low tubal rankness are compared to tr-MLSVD and HOOI, we will start with their optimization problems. More information regarding the tensor algebra and the tensor decompositions is provided in Chapter 2.

4.4. OPTIMIZATION PROBLEMS OF HOOI AND TR-MLSVD

Assuming that the Nth-order low multilinear rank tensor $\underline{\mathbf{L}}$ has the structure $\underline{\mathbf{L}} = \underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \cdots \times_N \mathbf{U}^{(N)}$ as defined in Chapter 2 with ranks R_n for $n \in \{1, \ldots, N\}$, both HOOI and tr-MLSVD solve the following optimization function

$$\min_{(\mathbf{U}^{(n)})^T \mathbf{U}^{(n)} = \mathbf{I}} \| \underline{\mathbf{Y}} - \underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \cdots \times_N \mathbf{U}^{(N)} \|^2 , \tag{4.4}$$

where $\underline{\mathbf{S}} = \mathbf{Y} \times_1 \mathbf{U}^{(1)^T} \cdots \times_N \mathbf{U}^{(N)^T}$. The sparse noise tensor $\underline{\mathbf{O}}$ as defined in (4.3) is not incorporated into the optimization problem (4.4). Since a least squares function is incorporated, the errors due to the outliers are expected to disrupt the estimation. We can incorporate a new variable into the optimization problem and mitigate the error due to the presence of outliers. Additionally, we can apply a convex relaxation to the low multilinear rankness and define the sum of nuclear norms.

4.5. SUM OF NUCLEAR NORM

The sum of the nuclear norm is a direct extension of the matrix rank to the tensor rank using the multilinear rank, i.e., the rank of the different unfoldings of the tensor. In [101], the optimization problem given in (4.3) is solved through the relaxation of the low multilinear rankness using the sum of the nuclear norm. We have the following optimization problem defined as

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{O}}} \left\{ \frac{1}{2} \| \underline{\mathbf{L}} + \underline{\mathbf{O}} - \underline{\mathbf{Y}} \|_F^2 + \gamma_L \sum_{n=1}^{N} \| \mathbf{L}_{(n)} \|_* + \gamma_O \| \underline{\mathbf{O}} \|_1 \right\} . \tag{4.5}$$

This can be solved by following [101], where N auxiliary variables are defined to separate $\mathbf{L}_{(n)}$ for $n \in \{1, \ldots, N\}$ and a proximity operator is employed alternately. Add auxiliary variables regarding $\underline{\mathbf{L}} = \underline{\mathbf{K}}$ and $\underline{\mathbf{O}} = \underline{\mathbf{R}}$. We have

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{O}}, \underline{\mathbf{K}}, \underline{\mathbf{R}}} \left\{ \frac{1}{2} \| \underline{\mathbf{L}} + \underline{\mathbf{O}} - \underline{\mathbf{Y}} \|_F^2 + \gamma_L \sum_{n=1}^{N} \| \mathbf{K}_{(n)} \|_* + \gamma_O \| \underline{\mathbf{R}} \|_1 \;\middle|\; \underline{\mathbf{K}} = \underline{\mathbf{L}},\ \underline{\mathbf{R}} = \underline{\mathbf{O}} \right\} . \tag{4.6}$$

We add N auxiliary variables $\underline{\mathbf{K}}^{(1)}, \ldots, \underline{\mathbf{K}}^{(N)}$ in an attempt to separate the mode-n unfoldings as done in [26].

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{O}}, \underline{\mathbf{K}}^{(1)}, \ldots, \underline{\mathbf{K}}^{(N)}, \underline{\mathbf{R}}} \left\{ \frac{1}{2} \| \underline{\mathbf{L}} + \underline{\mathbf{O}} - \underline{\mathbf{Y}} \|_F^2 + \gamma_L \sum_{n=1}^{N} \| \mathbf{K}_{(n)}^{(n)} \|_* + \gamma_O \| \underline{\mathbf{R}} \|_1 \;\middle|\; \underline{\mathbf{K}}^{(n)} = \underline{\mathbf{L}} \text{ for } n \in \{1, \ldots, N\}, \quad \underline{\mathbf{R}} = \underline{\mathbf{O}} \right\} . \tag{4.7}$$

Then we have the following augmented Lagrangian $\mathcal{L}_\rho$ of the optimization problem (4.7),

$$\mathcal{L}_\rho(\underline{\mathbf{L}}, \underline{\mathbf{O}}, \underline{\mathbf{K}}^{(1)}, \ldots, \underline{\mathbf{K}}^{(N)}, \underline{\mathbf{R}}, \underline{\mathbf{\Gamma}}^{(1)}, \ldots, \underline{\mathbf{\Gamma}}^{(N)}, \underline{\mathbf{\Lambda}}) = \frac{1}{2}\|\underline{\mathbf{L}} + \underline{\mathbf{O}} - \underline{\mathbf{Y}}\|_F^2 + \gamma_L \sum_{n=1}^{N} \|\mathbf{K}_{(n)}^{(n)}\|_* + \gamma_O \|\underline{\mathbf{R}}\|_1 + \sum_{n=1}^{N} \left(\langle \underline{\mathbf{\Gamma}}^{(n)}, \underline{\mathbf{K}}^{(n)} - \underline{\mathbf{L}} \rangle + \frac{\rho}{2} \|\underline{\mathbf{K}}^{(n)} - \underline{\mathbf{L}}\|_F^2 \right) + \langle \underline{\mathbf{\Lambda}}, \underline{\mathbf{R}} - \underline{\mathbf{O}} \rangle + \frac{\rho}{2} \|\underline{\mathbf{R}} - \underline{\mathbf{O}}\|_F^2 . \tag{4.8}$$

The augmented Lagrangian (4.8) can be solved using ADMM [105]. We introduce the iteration parameter t and update the primal and dual augmented Lagrangian variables given in (4.8). Define the tth iteration of augmented Lagrangian variables with $\underline{\mathbf{\Gamma}}^{(n)^t}, \underline{\mathbf{K}}^{(n)^t}$ for $n \in \{1, \ldots, N\}$ and $\underline{\mathbf{R}}^t, \underline{\mathbf{\Lambda}}^t, \underline{\mathbf{L}}^t, \underline{\mathbf{O}}^t$ with $0 < t < T$. The initializations of the augmented Lagrangian variables are shown with $t = 0$. The ADMM iterations stop when the differences between the $t+1$ and tth iteration of all augmented Lagrangian parameters are less than $\epsilon = 0.1$ or when the maximum number of iterations T is reached.

We will calculate the derivatives of $\mathcal{L}_\rho$ with respect to the augmented Lagrangian variables in Gauss-Seidel nature [105] and assign

$$\begin{aligned}
\underline{\mathbf{L}}^{t+1} &= \frac{\partial \mathcal{L}_\rho(\underline{\mathbf{L}}^t, \underline{\mathbf{O}}^t, \underline{\mathbf{K}}^{(1)^t}, \ldots, \underline{\mathbf{K}}^{(N)^t}, \underline{\mathbf{R}}^t, \underline{\mathbf{\Gamma}}^{(1)^t}, \ldots, \underline{\mathbf{\Gamma}}^{(N)^t}, \underline{\mathbf{\Lambda}}^t)}{\partial \underline{\mathbf{L}}^t}, \\
\underline{\mathbf{O}}^{t+1} &= \frac{\partial \mathcal{L}_\rho(\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{O}}^t, \underline{\mathbf{K}}^{(1)^t}, \ldots, \underline{\mathbf{K}}^{(N)^t}, \underline{\mathbf{R}}^t, \underline{\mathbf{\Gamma}}^{(1)^t}, \ldots, \underline{\mathbf{\Gamma}}^{(N)^t}, \underline{\mathbf{\Lambda}}^t)}{\partial \underline{\mathbf{O}}^t}, \\
\underline{\mathbf{K}}^{(n)^{t+1}} &= \frac{\partial \mathcal{L}_\rho(\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{O}}^{t+1}, \underline{\mathbf{K}}^{(1)^t}, \ldots, \underline{\mathbf{K}}^{(N)^t}, \underline{\mathbf{R}}^t, \underline{\mathbf{\Gamma}}^{(1)^t}, \ldots, \underline{\mathbf{\Gamma}}^{(N)^t}, \underline{\mathbf{\Lambda}}^t)}{\partial \underline{\mathbf{K}}^{(n)^t}}, \\
\underline{\mathbf{R}}^{t+1} &= \frac{\partial \mathcal{L}_\rho(\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{O}}^{t+1}, \underline{\mathbf{K}}^{(1)^{t+1}}, \ldots, \underline{\mathbf{K}}^{(N)^{t+1}}, \underline{\mathbf{R}}^t, \underline{\mathbf{\Gamma}}^{(1)^t}, \ldots, \underline{\mathbf{\Gamma}}^{(N)^t}, \underline{\mathbf{\Lambda}}^t)}{\partial \underline{\mathbf{R}}^t}, \\
\underline{\mathbf{\Gamma}}^{(n)^{t+1}} &= \frac{\partial \mathcal{L}_\rho(\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{O}}^{t+1}, \underline{\mathbf{K}}^{(1)^{t+1}}, \ldots, \underline{\mathbf{K}}^{(N)^{t+1}}, \underline{\mathbf{R}}^{t+1}, \underline{\mathbf{\Gamma}}^{(1)^t}, \ldots, \underline{\mathbf{\Gamma}}^{(N)^t}, \underline{\mathbf{\Lambda}}^t)}{\partial \underline{\mathbf{\Gamma}}^{(n)^t}}, \\
\underline{\mathbf{\Lambda}}^{t+1} &= \frac{\partial \mathcal{L}_\rho(\underline{\mathbf{L}}^{t+1}, \underline{\mathbf{O}}^{t+1}, \underline{\mathbf{K}}^{(1)^{t+1}}, \ldots, \underline{\mathbf{K}}^{(N)^{t+1}}, \underline{\mathbf{R}}^{t+1}, \underline{\mathbf{\Gamma}}^{(1)^{t+1}}, \ldots, \underline{\mathbf{\Gamma}}^{(N)^{t+1}}, \underline{\mathbf{\Lambda}}^t)}{\partial \underline{\mathbf{\Lambda}}^t},
\end{aligned} \tag{4.9}$$

where $\underline{\mathbf{K}}^{(n)^{t+1}}$ and $\underline{\mathbf{\Gamma}}^{(n)^{t+1}}$ are updated for $n \in \{1, \ldots, N\}$.

We first calculate $\underline{\mathbf{L}}^{t+1}$ and $\underline{\mathbf{O}}^{t+1}$. We solve jointly

$$\min_{\underline{\mathbf{L}}^t, \underline{\mathbf{O}}^t} \left\{ \frac{1}{2}\|\underline{\mathbf{L}}^t + \underline{\mathbf{O}}^t - \underline{\mathbf{Y}}\|_F^2 + \sum_{n=1}^{N} \left(\langle \underline{\mathbf{\Gamma}}^{(n)^t}, \underline{\mathbf{K}}^{(n)^t} - \underline{\mathbf{L}}^t \rangle + \frac{\rho}{2} \|\underline{\mathbf{K}}^{(n)^t} - \underline{\mathbf{L}}^t\|_F^2 \right) + \langle \underline{\mathbf{\Lambda}}^t, \underline{\mathbf{R}}^t - \underline{\mathbf{O}}^t \rangle + \frac{\rho}{2} \|\underline{\mathbf{R}}^t - \underline{\mathbf{O}}^t\|_F^2 \right\}. \tag{4.10}$$

Setting the derivatives with respect to $\underline{\mathbf{L}}^t$ and $\underline{\mathbf{O}}^t$ to zero leads to

$$\begin{aligned} \underline{\mathbf{L}}^t + \underline{\mathbf{O}}^t - \underline{\mathbf{Y}} - \sum_{n=1}^{N} \left(\underline{\mathbf{\Gamma}}^{(n)^t} + \rho \left(\underline{\mathbf{K}}^{(n)^t} - \underline{\mathbf{L}}^t \right) \right) = \underline{\mathbf{0}}, \\ \underline{\mathbf{L}}^t + \underline{\mathbf{O}}^t - \underline{\mathbf{Y}} - \underline{\mathbf{\Lambda}}^t - \rho \left(\underline{\mathbf{R}}^t - \underline{\mathbf{O}}^t \right) = \underline{\mathbf{0}}. \end{aligned} \tag{4.11}$$

We can solve the above equations to obtain

$$\begin{aligned} \underline{\mathbf{L}}^{t+1} &= \frac{\rho(\underline{\mathbf{Y}} - \underline{\mathbf{R}}^t)}{C} + \frac{1+\rho}{C} \sum_{n=1}^{N} \underline{\mathbf{\Gamma}}^{(n)^t} + \frac{\rho(1+\rho)}{C} \sum_{n=1}^{N} \underline{\mathbf{K}}^{(n)^t} - \frac{1}{C} \underline{\mathbf{\Lambda}}^t, \\ \underline{\mathbf{O}}^{t+1} &= \underline{\mathbf{Y}} + \sum_{n=1}^{N} \underline{\mathbf{\Gamma}}^{(n)^t} + \rho \sum_{n=1}^{N} \underline{\mathbf{K}}^{(n)^t} - (1+\rho N) \underline{\mathbf{L}}^{t+1}, \end{aligned} \tag{4.12}$$

where C is $\rho(N+1+\rho N)$. Then we calculate $\underline{\mathbf{K}}^{(1)^{t+1}}, \ldots, \underline{\mathbf{K}}^{(N)^{t+1}}$ and $\underline{\mathbf{R}}^{t+1}$. For $\underline{\mathbf{K}}^{(1)^{t+1}}, \ldots, \underline{\mathbf{K}}^{(N)^{t+1}}$ we have

$$\min_{\underline{\mathbf{K}}^{(1)^t}, \ldots, \underline{\mathbf{K}}^{(N)^t}} \left\{ \gamma_L \sum_{n=1}^{N} \| \mathbf{K}_{(n)}^{(n)^t} \|_* + \sum_{n=1}^{N} \left(\langle \underline{\mathbf{\Gamma}}^{(n)^t}, \underline{\mathbf{K}}^{(n)^t} - \underline{\mathbf{L}}^{t+1} \rangle + \frac{\rho}{2} \| \underline{\mathbf{K}}^{(n)^t} - \underline{\mathbf{L}}^{t+1} \|_F^2 \right) \right\}. \tag{4.13}$$

If we write the augmented Lagrangian for $\underline{\mathbf{K}}^{(n)^{t+1}}$, we have

$$\min_{\underline{\mathbf{K}}^{(n)^t}} \left\{ \gamma_L \| \mathbf{K}_{(n)}^{(n)^t} \|_* + \langle \underline{\mathbf{\Gamma}}^{(n)^t}, \underline{\mathbf{K}}^{(n)^t} - \underline{\mathbf{L}}^{t+1} \rangle + \frac{\rho}{2} \| \underline{\mathbf{K}}^{(n)^t} - \underline{\mathbf{L}}^{t+1} \|_F^2 \right\}. \tag{4.14}$$

We can complete the right-hand side to a complete square,

$$\min_{\underline{\mathbf{K}}^{(n)^t}} \left\{ \frac{\gamma_L}{\rho} \| \mathbf{K}_{(n)}^{(n)^t} \|_* + \frac{1}{2} \left\| \underline{\mathbf{K}}^{(n)^t} - \left(\underline{\mathbf{L}}^{t+1} - \frac{1}{\rho} \underline{\mathbf{\Gamma}}^{(n)^t} \right) \right\|_F^2 \right\}. \tag{4.15}$$

This can be solved using the proximity operator [104] for each $n \in \{1, \ldots, N\}$ as

$$\underline{\mathbf{K}}^{(n)^{t+1}} = \text{fold}_n \left(\text{prox}_{\frac{\gamma_L}{\rho}}^{\|\cdot\|_*} \left(\left(\underline{\mathbf{L}}^{t+1} - \frac{1}{\rho} \underline{\mathbf{\Gamma}}^{(n)^t} \right)_{(n)} \right) \right), \tag{4.16}$$

where fold_n is the operator given in (2.3) that reverses the mode-n unfolding back to the tensor, and the proximity operator of the nuclear norm is defined in Section 2.3.4.

For $\underline{\mathbf{R}}^{t+1}$ we have

$$\min_{\underline{\mathbf{R}}^t} \left\{ \gamma_O \| \underline{\mathbf{R}}^t \|_1 + \langle \underline{\mathbf{\Lambda}}^t, \underline{\mathbf{R}}^t - \underline{\mathbf{O}}^{t+1} \rangle + \frac{\rho}{2} \| \underline{\mathbf{R}}^t - \underline{\mathbf{O}}^{t+1} \|_F^2 \right\}. \tag{4.17}$$

This can be solved using the proximity operator [104]

$$\underline{\mathbf{R}}^{t+1} = \text{prox}_{\frac{\gamma_O}{\rho}}^{\|\cdot\|_1} \left(\underline{\mathbf{O}}^{t+1} - \frac{1}{\rho} \underline{\mathbf{\Lambda}}^t \right). \tag{4.18}$$

where the L1 norm proximity operator is

$$\begin{aligned}\text{prox}^{\|\cdot\|_1}_{\frac{\gamma_O}{\rho}}\left(\underline{\mathbf{O}}^{t+1}-\frac{1}{\rho}\underline{\mathbf{\Lambda}}^{t}\right) &= \text{sign}\left(\underline{\mathbf{O}}^{t+1}-\frac{1}{\rho}\underline{\mathbf{\Lambda}}^{t}\right)\\ &\odot \max\left\{\left|\underline{\mathbf{O}}^{t+1}-\frac{1}{\rho}\underline{\mathbf{\Lambda}}^{t}\right|-\frac{\gamma_O}{\rho},\ \underline{\mathbf{0}}\right\}.\end{aligned} \tag{4.19}$$

where γ_O/ρ is subtracted element-wise, and the absolute value $|\cdot|$ is applied element-wise. We can update the Lagrangian multipliers with the following equations

$$\begin{aligned}\underline{\mathbf{\Gamma}}^{(n)^{t+1}} &= \underline{\mathbf{\Gamma}}^{(n)^{t}}+\rho\left(\underline{\mathbf{K}}^{(n)^{t+1}}-\underline{\mathbf{L}}^{t+1}\right),\\ \underline{\mathbf{\Lambda}}^{t+1} &= \underline{\mathbf{\Lambda}}^{t}+\rho\left(\underline{\mathbf{R}}^{t+1}-\underline{\mathbf{O}}^{t+1}\right).\end{aligned} \tag{4.20}$$

The algorithm starts with initializations $\{\underline{\mathbf{\Gamma}}^{(n)^0},\underline{\mathbf{K}}^{(n)^0}\}=\underline{\mathbf{0}}\in\mathbb{R}^{I_1,\ldots,I_N}$ for $n\in\{1,\ldots,N\}$, $\{\underline{\mathbf{R}}^0,\underline{\mathbf{\Lambda}}^0,\underline{\mathbf{L}}^0,\underline{\mathbf{O}}^0\}=\underline{\mathbf{0}}\in\mathbb{R}^{I_1,\ldots,I_N}$ and runs until the maximum iteration T is reached or until the difference between $t+1$st and tth version of each initialized variable is less than or equal to ϵ. The tensor $\underline{\mathbf{L}}^t$ is the low multilinear rank estimate of TSPCP. The algorithmic steps are provided in Algorithm 3.

Algorithm 3 Tensor Stable Principal Component Pursuit

Given $\underline{\mathbf{Y}}\in\mathbb{R}^{I_1,\ldots,I_N}$, $\gamma_L,\gamma_O,\rho,\epsilon$ initialize $\{\underline{\mathbf{\Gamma}}^{(n)^0},\underline{\mathbf{K}}^{(n)^0}\}=\underline{\mathbf{0}}\in\mathbb{R}^{I_1,\ldots,I_N}$ for $n\in\{1,\ldots,N\}$, $\{\underline{\mathbf{R}}^0,\underline{\mathbf{\Lambda}}^0,\underline{\mathbf{L}}^0,\underline{\mathbf{O}}^0\}=\underline{\mathbf{0}}\in\mathbb{R}^{I_1,\ldots,I_N}$, $t=0$,

while $t<T$ or $\Big(\left\|\underline{\mathbf{L}}^{t+1}-\underline{\mathbf{L}}^{t}\right\|_F>\epsilon$ and $\left\|\underline{\mathbf{O}}^{t+1}-\underline{\mathbf{O}}^{t}\right\|_F>\epsilon$ and $\left\|\underline{\mathbf{K}}^{(n)^{t+1}}-\underline{\mathbf{K}}^{(n)^{t}}\right\|_F>\epsilon$ for $n\in\{1,\ldots,N\}$ and $\left\|\underline{\mathbf{\Gamma}}^{(n)^{t+1}}-\underline{\mathbf{\Gamma}}^{(n)^{t}}\right\|_F>\epsilon$ for $n\in\{1,\ldots,N\}$ and $\left\|\underline{\mathbf{R}}^{t+1}-\underline{\mathbf{R}}^{t}\right\|_F>\epsilon$ and $\left\|\underline{\mathbf{\Lambda}}^{t+1}-\underline{\mathbf{\Lambda}}^{t}\right\|_F>\epsilon\Big)$ **do**

  $\underline{\mathbf{L}}^{t+1},\underline{\mathbf{O}}^{t+1}\leftarrow$ see (4.12)
  for $n=\{1,2,3,4\}$ **do**
    $\underline{\mathbf{K}}^{(n)^{t+1}}\leftarrow$ see (4.16)
  end for
  $\underline{\mathbf{R}}^{t+1}\leftarrow$ see (4.18)
  for $n=\{1,2,3,4\}$ **do**
    $\underline{\mathbf{\Gamma}}^{(n)^{t+1}}\leftarrow$ see (4.20)
  end for
  $\underline{\mathbf{\Lambda}}^{t+1}\leftarrow$ see (4.20)
  $t\leftarrow t+1$
end while
Return: $\underline{\mathbf{L}}^t$

It is suggested in [106] that the proximity operator given in (4.16), which is part of Algorithm 3, is problematic for denoising. This is due to the global soft thresholding operator applied to all the singular values. A better algorithm for denoising tasks can be achieved by the weighted sum of the singular values.

4.5.1. WEIGHTED SUM OF NUCLEAR NORM

We propose a weighting scheme for SNN methods that utilizes the known multilinear ranks of the true tensor. The problem is still convex with this additional update. The largest singular values corresponding to the signal subspace would be kept the same at each optimization iteration, while the lowest singular values corresponding to the noise subspace would be reduced by γ_L/ρ. For the nth mode unfolding with the corresponding rank R_n, the weighted thresholding matrix $\tilde{\mathbf{\Sigma}}_{\gamma_L/\rho}$ becomes

$$\tilde{\mathbf{\Sigma}}^{(n)}_{\frac{\gamma_L}{\rho}} = \max\left\{\mathbf{\Sigma}^{(n)} - \mathbf{C}^{(n)}\frac{\gamma_L}{\rho},\ \mathbf{0}\right\}, \tag{4.21}$$

where

$$\mathbf{C}^{(n)} = \begin{bmatrix} \mathbf{0} \in \mathbb{R}^{R_n \times R_n} & \mathbf{0} \in \mathbb{R}^{R_n \times (I_n - R_n)} \\ \mathbf{0} \in \mathbb{R}^{(I_n - R_n)\times R_n} & \mathbf{I} \in \mathbb{R}^{(I_n - R_n)\times (I_n - R_n)} \end{bmatrix}, \tag{4.22}$$

for $n \in \{1, \ldots, N\}$. Hence, the proximity operator defined in (4.16) becomes

$$\text{prox}^{\|\cdot\|_*}_{\frac{\gamma_L}{\rho}}\left(\left(\underline{\mathbf{L}}^{t+1} - \frac{1}{\rho}\underline{\mathbf{\Gamma}}^{(n)^t}\right)_{(n)}\right) = \mathbf{U}^{(n)}\,\tilde{\mathbf{\Sigma}}^{(n)}_{\frac{\gamma_L}{\rho}}\left(\mathbf{V}^{(n)}\right)^{\mathsf{T}}, \tag{4.23}$$

for $n \in \{1, \ldots, N\}$. The problem in (4.5) can be solved in the same way as [101], only by changing the proximity operator from (4.16) to (4.23). We name this method weighted SNN (wSNN). This step only changes the updates of $\underline{\mathbf{K}}^{(n)}$ in $n \in \{1, \ldots, N\}$ given in (4.16), while the rest of Algorithm 3 is the same.

4.6. TUBAL NUCLEAR NORM

The tubal nuclear norm is introduced using the framework of t-SVD [8]. Detailed information is provided in Chapter 2. We will add the definition of orientation invariant tubal nuclear norm (OITNN) as described in [30].

Definition 4.6.1. Mode-$(n, n+1)$ unfolding: The mode-$(n, n+1)$ unfolding of $\underline{\mathbf{L}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ creates a 3D tensor $\underline{\mathbf{L}}_{(n,n+1)} \in \mathbb{R}^{I_n \times D/(I_n I_{n+1}) \times I_{n+1}}$ by permuting the nth mode of $\underline{\mathbf{L}}$ to the first dimension, the $(n+1)$th mode to the last dimension, and grouping the remaining modes. Here, D is defined as $\prod_{n=1}^{N} I_n$. This unfolding is also denoted as $\underline{\mathbf{L}}_{[n]}$. For a 3rd-order tensor, this operation gives three permutations of the tensor, where the orders are permuted circularly.

Given the t-SVD of $\underline{\mathbf{L}}$ as $\underline{\mathbf{L}} = \underline{\breve{\mathbf{U}}} \star \underline{\breve{\mathbf{S}}} \star \underline{\breve{\mathbf{V}}}^{\mathsf{T}}$, the tubal nuclear norm is given by the average of the sum of the frontal slices of the 1D Fourier Transform of the core tensor, that is,

$$\|\underline{\mathbf{L}}\|_{\text{TNN}} = \frac{1}{I_3}\sum_{i_3=1}^{I_3} \text{Tr}\left(\left(\underline{\breve{\mathbf{S}}} \times_3 \mathbf{F}^{(3)}\right)(:, :, i_3)\right). \tag{4.24}$$

The OITNN is defined as

$$\|\underline{\mathbf{L}}\|_{\text{OITNN}} = \frac{1}{N}\sum_{n=1}^{N} \|\underline{\mathbf{L}}_{[n]}\|_{\text{TNN}}. \tag{4.25}$$

In [104], the optimization problem given in (4.3) is solved similarly to SNN while only changing the rank relaxation into the tubal nuclear norm. This can be formulated as

$$\min_{\underline{\mathbf{L}},\underline{\mathbf{O}}} \left\{ \frac{1}{2}\|\underline{\mathbf{L}}+\underline{\mathbf{O}}-\underline{\mathbf{Y}}\|_F^2 + \gamma_L\|\underline{\mathbf{L}}\|_{\text{TNN}} + \gamma_O\|\underline{\mathbf{O}}\|_1 \right\}. \tag{4.26}$$

Under the tensor incoherence condition (see [104]), the exact recovery of the true tensor $\underline{\mathbf{L}}$ is guaranteed for the noiseless case. This algorithm is named the tensor stable principal component pursuit (TSPCP). In [31], an additional constraint regarding the infinity norm of the tensor $\underline{\mathbf{L}}$ is added to the objective function such that the absolute values of $\underline{\mathbf{L}}$ are below a certain threshold, and the outliers are captured at $\underline{\mathbf{O}}$ instead. This optimization problem can be shown as

$$\begin{aligned} \min_{\underline{\mathbf{L}},\underline{\mathbf{O}}} \left\{ \frac{1}{2}\|\underline{\mathbf{L}}+\underline{\mathbf{O}}-\underline{\mathbf{Y}}\|_F^2 + \gamma_L\|\underline{\mathbf{L}}\|_{\text{TNN}} + \gamma_O\|\underline{\mathbf{O}}\|_1 \right\} \\ \text{s.t.} \quad \|\underline{\mathbf{L}}\|_\infty \leq \alpha. \end{aligned} \tag{4.27}$$

The constraint regarding the infinity norm is handled by clipping the maximum value of the absolute value of $\underline{\mathbf{L}}$ to α at each iteration of the optimization. Similar algorithmic steps described in Algorithm 3 can be followed with a change of updates of $\underline{\mathbf{K}}^{(n)}$ in $n \in \{1,\ldots,N\}$ given in (4.16). Instead of the convex relaxation of SNN defined in Section 2.3.4, the convex relaxation of TNN defined in Section 2.3.5 can be used.

Only the low-rankness in the third mode is incorporated in (4.27) and (4.26). The authors in [30] extended the analysis to all the modes for general Nth-order tensors $\underline{\mathbf{L}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$, $\underline{\mathbf{O}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$, and $\underline{\mathbf{W}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ using the mode-$(n, n+1)$ unfolding. According to Lemma 1 in [30], any tensor that has a low Tucker rank has a low orientation invariant average rank. This enables the direct application of two convex proxy methods, OITNN-O and OITNN-L, onto low Tucker rank applications using the orientation invariant tubal nuclear norm (OITNN).

The OITNN-O considers the low tubal rankness in all orientations and solves the optimization problem that is defined as

$$\begin{aligned} \min_{\underline{\mathbf{L}},\underline{\mathbf{O}}} \left\{ \frac{1}{2}\|\underline{\mathbf{L}}+\underline{\mathbf{O}}-\underline{\mathbf{Y}}\|_F^2 + \gamma_L\|\underline{\mathbf{L}}\|_{\text{OITNN}} + \gamma_O\|\underline{\mathbf{O}}\|_1 \right\} \\ \text{s.t.} \quad \|\underline{\mathbf{L}}\|_\infty \leq \alpha. \end{aligned} \tag{4.28}$$

The OITNN-L considers $\underline{\mathbf{L}}$ as the sum of N auxiliary tensors, which can have a low tubal rank in their respective mode. Such modeling is claimed to be better if the tensor might have significantly different tubal ranks in different mode-$(n, n+1)$ unfoldings [30]. Assuming that an incoherence condition holds, this optimization problem is defined as

$$\begin{aligned} \min_{\underline{\mathbf{L}}^{(1)},\ldots,\underline{\mathbf{L}}^{(N)},\underline{\mathbf{O}}} \left\{ \frac{1}{2}\left\|\sum_{n=1}^{N}\underline{\mathbf{L}}^{(n)}+\underline{\mathbf{O}}-\underline{\mathbf{Y}}\right\|_F^2 + \frac{\gamma_L}{N}\sum_{n=1}^{N}\left\|\underline{\mathbf{L}}^{(n)}_{[n]}\right\|_{\text{TNN}} + \gamma_O\|\underline{\mathbf{O}}\|_1 \right\} \\ \text{s.t.} \quad \left\|\sum_{n}\underline{\mathbf{L}}^{(n)}\right\|_\infty \leq \alpha. \end{aligned} \tag{4.29}$$

The exact recovery of true $\underline{\mathbf{L}}$ is not guaranteed for the noiseless case when the infinity norm constraint is incorporated.

Table 4.1: Summary of the methods that are used in this chapter.

	Constraints	Noise	Optimization problem
HOOI [19]	Ranks	WGN	(4.4)
tr-MLSVD [13]	Ranks	WGN	(4.4)
wSNN	Ranks	WGN + Sparse	(4.5) and (4.23)
SNN [101]	-	WGN + Sparse	(4.5)
TNN [107]	$\|\|\underline{\mathbf{L}}\|\|_\infty$	WGN + Sparse	(4.27)
OITNN-O [30]	$\|\|\underline{\mathbf{L}}\|\|_\infty$	WGN + Sparse	(4.28)
OITNN-L [30]	$\|\|\underline{\mathbf{L}}\|\|_\infty$	WGN + Sparse	(4.29)
TSPCP [104]	-	WGN + Sparse	(4.26)

4.7. SIMULATION

This section compares the denoising performance of the methods given in Table 1 through a Monte Carlo simulation. Here, we have generated two 3D tensors of size $\underline{\mathbf{L}} \in \mathbb{R}^{20\times 20\times 20}$ with unequal ranks and equal ranks. This is done by generating a core tensor of size $(10, 10, 10)$ and $(5, 10, 15)$ from a normal distribution, multiplying it in each mode with unitary factor matrices, rescaling the resulting tensor between 1 and 2, and finally taking the logarithm. The noise tensor is generated by taking the logarithm of tensor $\underline{\mathbf{M}} \in \mathbb{R}^{20\times 20\times 20}$ where each entry $m_{i_1 i_2 i_3}$ follows the Rayleigh distribution with a scaling parameter of 1. The SNR of the problem is defined by

$$10\log_{10}\left(\frac{\mathbb{E}\left[\left(\underline{\mathbf{L}}-\mathbb{E}[\underline{\mathbf{L}}]\right)^2\right]}{\mathbb{E}\left[\left(\underline{\mathbf{M}}-\mathbb{E}[\underline{\mathbf{M}}]\right)^2\right]}\right). \tag{4.30}$$

We have scaled the variance of $\underline{\mathbf{L}}$ while fixing the scaling parameter of the noise such that SNRs of $\{-5, 0, 5, 10, 15, 20, 25, 30\}$ are obtained. The noisy tensor $\underline{\mathbf{Y}}$ is obtained by the summation of $\underline{\mathbf{L}}$ and $\underline{\mathbf{M}}$. The normalized error is used as a performance metric, which is defined as

$$20\log_{10}\left(\frac{\|\hat{\underline{\mathbf{L}}}-\underline{\mathbf{L}}\|_F}{\|\underline{\mathbf{L}}\|_F}\right). \tag{4.31}$$

Ten numbers in the range $(1, 100)$ are traversed for optimization of the parameters γ_O and γ_L for OITNN-L, while the range $(1, 30)$ is traversed for the rest of the robust tensor decomposition methods. We fixed the value of ρ to 1 and set the maximum iteration number to 500. The true ranks are used for tr-MLSVD and HOOI, and the true $||\underline{\mathbf{L}}||_\infty$ is assigned to α for TNN, OITNN-O, and OITNN-L.

4.8. RESULTS

The simulation defined in Section 4.7 is repeated for 20 random initialization of the noise tensor and true tensor. The normalized error is calculated for various SNRs and plotted in Fig. 4.1. For each run of the algorithm, the best-performing

tuning parameters γ_O and γ_L are selected to calculate the normalized error. For all SNRs, the wSNN performs better than the truncated multilinear singular value decomposition and HOOI. More improvement is found for low SNR scenarios. The best-performing method is OTINN_L for both ranks when the SNR is less than 20 dB. For SNRs greater than 20 dB, the wSNN performs better with a small improvement compared to HOOI.

We ran pairwise t-tests to assess the statistical significance. For the scenario with multilinear ranks $(5, 10, 15)$, all the algorithms had a statistically significant difference with $p < 0.05$, except between TNN and tr-MLSVD at 20 SNR, and between OITNN-O and OITNN-O at 30 dB. For the scenario with ranks $(10, 10, 10)$, all the algorithms had a statistically significant difference with $p < 0.05$, except between TNN and SNN at 0 dB, and between OITNN-O and OITNN-O at 25 dB.

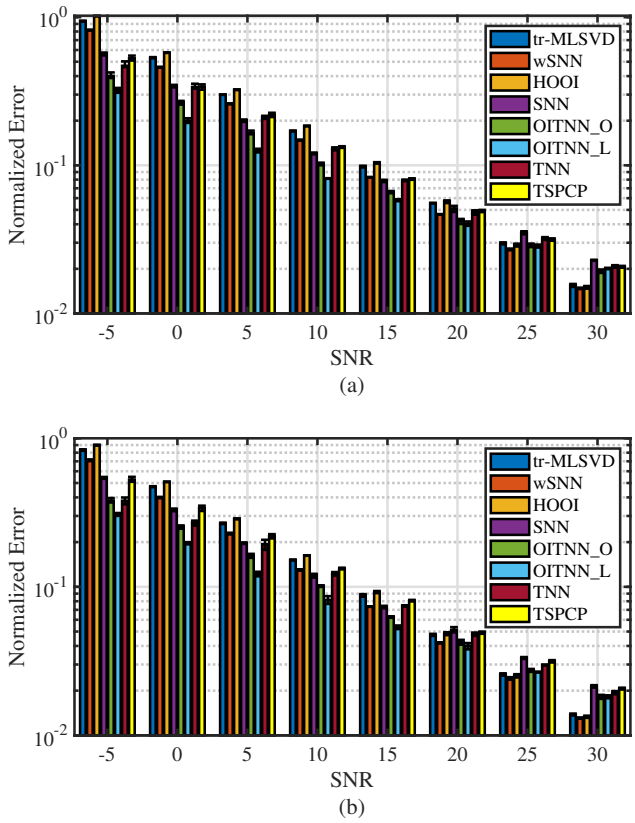


Figure 4.1: The plot of the normalized error versus SNR for low multilinear rank $\underline{\mathbf{L}}$ denoised with the aforementioned methods over 20 Monte Carlo simulations. The plots with subscripts (a) and (b) represent the ranks $(10, 10, 10)$ and $(5, 10, 15)$, respectively.

4.9. DISCUSSION AND CONCLUSION

This chapter compares low multilinear or tubal rank tensor decomposition methods for despeckling applications. We have proposed a weighting scheme for SNN-based algorithms when the ranks are known and compared the results to robust tensor decomposition methods. All robust tensor decomposition algorithms outperform HOOI and tr-MLSVD for low SNRs because of the better noise model. We have observed that wSNN performs better than HOOI and tr-MLSVD for high SNR scenarios. The weights of wSNN are arranged such that the highest singular values that correspond with the true rank are not updated during the proximity operator update. This resembles the tr-MLSVD, where the singular values with an index greater than the rank are truncated for denoising. Alongside this property, the sparsity constraint that is added to the formulation of wSNN results in a better recovery. For low SNR scenarios, wSNN performed worse than SNN. This is due to the corruption of the singular values with the noise subspace. The utilization of the proximity operator shown in (4.23) restricts the updating of singular values associated with the rank. Consequently, the denoising capabilities of wSNN are constrained by a diminished search space, thus contributing to its reduced performance under low SNR conditions.

The selection of the weight $\mathbf{C}^{(n)}$ (4.22) is specifically designed to keep the highest mode-n singular values associated with the rank in each mode. This can be thought of as a regularization parameter that is incorporated into the tr-MLSVD. In tr-MLSVD, the highest R_n for $n \in \{1, \ldots, N\}$ section of the core tensor is kept along with the corresponding factor matrices as explained in Chapter 2. The convex relaxation using SNN explained in Section 2.3.4 applies a soft-thresholding to the mode-n singular values iteratively. The application of the weight proposed in (4.23) allows for the iterative soft-thresholding of the weights regularized with a sparsity-related parameter. We investigated this approach to model the speckle noise, which is approximated as WGN with sparse outliers. The highest R_n mode-n singular values for $n \in \{1, \ldots, N\}$ are kept at each iteration like tr-MLSVD, the noise subspace is soft-thresholded. This operation is regularized with a sparsity-related parameter that affects the calculation of SVD in each step, hence indirectly also affects all the mode-n singular values for $n \in \{1, \ldots, N\}$.

Among the convex relaxations, TNN performs better than SNN for low SNRs. The algorithm proposed in [107] considers the low tubal rankness in only one of the modes using the t-SVD framework. In our simulations, we generated a tensor with a low multilinear rank in all modes. Using the relation between the multilinear rank and the tubal rank given in Section 2.4, the resulting tensor is also of low tubal rank. Algorithms such as OITNN-L and OITNN-O that incorporate the low tubal rankness in all modes have performed better. Although OITNN-O is expected to perform worse when the tensor has a low tubal rank in some modes than others, we have not observed this in our simulation. OITNN-L is found to be the algorithm that performed the best for low SNR scenarios, regardless of the ranks. The sparse outliers have minimal effect on the low multilinear and tubal rank approximation for SNRs greater than 20 dB. This might be because, as the noise intensity decreases, the noise after the logarithmic transformation becomes closer to a Gaussian distribution, diminishing the impact of sparsity.

One of the limitations of this chapter is the grid search on the low tubal rank and sparsity-related tuning parameters γ_L and γ_O. We have defined the range such that a convex range of normalized mean squared error has been observed.

By extending the grid search, a better comparison can be made among the convex low tubal rank approximation methods. This chapter aims to understand the importance of sparsity constraint in denoising speckle noise. The HOOI and tr-MLSVD are better suited for denoising WGN. The fact that the robust low multilinear rank approximation methods perform better than HOOI and tr-MLSVD proves the added benefit of the sparsity constraint. We compare the results of the SNN and TNN-based methods in a heuristic manner. A better approach for this comparison would be optimizing the tuning parameters using methods such as [108].

Another limitation of the study is the need for a quantization analysis for the tensor $\underline{\mathbf{L}}$. During the simulation, we scaled the variance of $\underline{\mathbf{L}}$ to reach the predefined SNRs. In many despeckling applications, such as ultrasound, the input is 8-bit unsigned integers. An analysis that covers the effect of the quantization on the despeckling applications is left for future work, along with real-world applications. In addition, we have assumed that the rank of the true tensor $\underline{\mathbf{L}}$ is known and defined the weighting parameter given in (4.22) according to this rank. The ranks of the system can be estimated using the outlier-resistant rank estimation methods such as the one given in [109]. The algorithm could be modified such that the ranks are not known. In such a case, the higher singular values can be thresholded less using techniques such as [100]. Finally, non-linear transformations such as the log transformation might increase the rank of the tensor. In real-world applications with limited quantization, further analysis is required.

TNN and TSPCP are defined for 3D tensors that can be applied to speckle denoising applications where 2D nonlocal correlated patches are stacked to create the tensor. For applications like 4D ultrasound or MRI, orientation-insensitive TNN methods such as OITNN-O and OITNN-L could be used. SNN methods are directly applicable to any order tensor. Hence, they can be immediately applied to a 4D speckle denoising application. In conclusion, we have reviewed various low tubal rank and low multilinear rank approximation methods for denoising speckle noise and found that incorporating the sparsity constraint is useful, especially for low SNR scenarios.

4.10. CHAPTER SUMMARY

In this chapter, we introduced the optimization problems that tackle the low multilinear rankness and low tubal rankness introduced in Section 2.3 with the signal model introduced in Section 3.3. We investigated the convex relaxation of the Fisher-Tippett distribution, that is, WGN with sparse outliers. We summarized the optimization problems in Table 4.1.

We only conducted simulations in this chapter and compared the denoising performances of the tensor decomposition techniques in various SNRs. In low SNR scenarios, the incorporation of sparsity had a significant effect, and OITNN-L gave the best denoising performance. In high SNR scenarios, the wSNN performed the best. Although statistically significant, the normalized error difference between wSNN and the tr-MLSVD was relatively close. Since the tr-MLSVD can be calculated faster due to its closed-form solution, we further investigate its denoising performance through subsequent prostate cancer classification accuracy using in-vivo data in Chapter 5. The computation time of OITNN-L was significantly higher than OITNN-O. Therefore, we investigate the prostate cancer

classification performance of OITNN-O in Chapter 6.

5

DENOISING OF DYNAMIC CONTRAST-ENHANCED ULTRASOUND USING MULTILINEAR SINGULAR VALUE DECOMPOSITION

In this chapter, we evaluate the prostate cancer classification performance after tr-MLSVD-based denoising using six in-vivo recordings obtained in Zhejiang University, provided to us by Prof. Pintong Huang. The tr-MLSVD works by truncating MLSVD to pre-defined ranks, prompting an investigation into suitable rank estimation methods. In addition to the noise model provided in Section 3.3, in this chapter we examine the effect of various other noise types such as the motion artifacts, breathing, and additive WGN. The motion artifacts and breathing types of noise are added before the log-compression. The additive WGN is incorporated after the log-compression. This chapter partially answers the second research question.

5.1. INTRODUCTION

Prostate cancer (PCa) is found to be the most common malignancy among American men, except for skin cancer, and according to the estimates for 2021, 248 530 new cases are expected, of which 34 130 people will die from the disease [110]. The recommended guidelines for detection and grading of PCa are $\geq$ 10-core systematic biopsy [111], which consists of an invasive procedure that can still miss significant PCa lesions [112]. In addition, infectious complications

This chapter is based on the article published as "Denoising of Dynamic Contrast-enhanced Ultrasound Sequences: A Multilinear Approach." by M. Calis, M. Mischi, A. van der Veen, B. Hunyadi, International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSIGNALS), 2022, 15

might arise due to the insertion of needles through the rectal wall [113]. Also, because of the poor patient selection by prostate-specific antigen blood testing, about 3 of 4 biopsies are, in retrospect, unnecessary, as no cancer is found. Recent advances in imaging techniques can potentially reduce the needle cores in systematic biopsies and overcome unnecessary biopsies [114].

One of the featured techniques to classify PCa is the detection of angiogenesis, which is the rapid growth of new blood vessels around the tumourous region. The newly formed capillaries create a tortuous and highly irregular network that can be analyzed by modeling the flow dynamics of the ultrasound contrast agents (UCAs) [115] [116]. Although several modalities have been proposed that can monitor the UCAs, we focus on dynamic contrast-enhanced ultrasound (DCEUS) because of the promising results for PCa localization [117] [67]. In DCEUS, contrast agents that are smaller than 10 μm are imaged using a low-energy ultrasound pulse. The frequency of the pulse is around the resonance frequency of the UCAs, which enhances the contrast between the microbubbles and the surrounding tissue. The time-intensity curves (TICs) of these contrast agents correlate with the underlying vasculature [118], which can be exploited to classify PCa.

There has been a substantial amount of work in the analysis of the perfusion characteristics of TICs [53] [54]. Although increased perfusion is expected due to the increased microvascular density, contradictory effects of angiogenesis and perfusion have been observed [119]. The flow resistance might be reduced due to the introduction of arteriovenous shunts and lack of vasomotor control. This effect might not be observed due to the small diameter of the newly formed microvessels and the increased interstitial pressure because of the extravascular leakage [120]. A different path has been taken by [121] [74] where the authors have modeled the multi-trajectory bubble transport inside the prostate as a convective-dispersion process and introduced a new model based on dispersion characteristics, namely the local density random walk model (LDRW) [55]. The authors have reported a good correlation between dispersion and angiogenesis, even though the TICs suffer from a low signal-to-noise (SNR) ratio.

Several techniques were proposed to remove the noise of various origins from the DCEUS acquisitions. Spatial filters were used to denoise the speckle noise [87] [122], and temporal filters were proposed to denoise the clutter [123]. These methods assumed that the temporal and spatial frequencies of the desired signal and the noise were different. Using a different approach, the authors in [93] proposed the blind source separation (BSS) of the DCEUS acquisition into sources, where the desired signal was recovered by discarding the noise sources. The best-performing BSS was found to be singular value decomposition (SVD) for modeling the time-intensity curves. The dynamics of the UCAs have been captured in the first few singular values of the decomposition. Although several cancer markers, such as the similarity between TICs [63] and solutions to convective dispersion models [69] [68] have been extended to 3rd-order, the denoising algorithms do not take the tensor structure of the recording into consideration.

The SVD denoising analyzed by [93] creates the Casorati matrix, which vectorizes the spatial modes into rows and the temporal mode into columns. The vectorization of the spatial modes into rows results in a loss of spatial information regarding the location of the voxels. The multilinear singular value decomposition (MLSVD) proposed by [13] keeps the tensor format of the data and generalizes the concept of SVD to higher orders. The information that is retained

by keeping the tensor format of the data is hypothesized to enable a better representation of the bubble dynamics and, hence, a better denoising capability. We introduced truncated multilinear singular value decomposition (tr-MLSVD) in Chapter 2. In the literature, the tr-MLSVD has been applied for clutter filter denoising in power Doppler images [124] [125]. Improved sensitivity and specificity have been observed. As far as the authors know, no work has been done to analyze the performance of tr-MLSVD in TIC dispersion modeling. This chapter aims to answer the following research question: Can the retained 3rd-order structural information provided by tr-MLSVD improve the classification performance of TIC dispersion modeling compared to SVD?

5.2. SIGNAL MODEL

In nearly all commercial scanners, the recordings are logarithmically compressed to deal with the large dynamic ranges. Here, we model the logarithmically compressed DCEUS acquisitions as the addition of the original signal and the noise,

$$\underline{\mathbf{Y}} = \underline{\mathbf{L}} + \underline{\mathbf{E}}, \tag{5.1}$$

where $\underline{\mathbf{L}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ stands for the original signal, $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ stands for the received signal and $\underline{\mathbf{E}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ stands for the noise. When only multiplicative noise is considered, and the logarithmic compression is applied, $\underline{\mathbf{E}}$ stands for the speckle noise [126]. After the logarithmic compression, speckle noise is shown to obey the Fisher-Tippet distribution, which can be approximated as white Gaussian noise with outliers that have a fixed variance [92]. There can be several other noise sources in practical scenarios, such as motion artifacts due to the urologist's probe handling and the patient's breathing.

In (5.1), the signal model after log-compression is given. Before log-compression, we have the $\tilde{\underline{\mathbf{L}}}$ as explained in Section 3.2.2. The time-intensity curves $\tilde{\underline{\mathbf{L}}}$ are assumed to follow the LDRW model [55] described in Section 3.2.2, which is given as

$$\tilde{\mathbf{l}}_{i_1 i_2 i_3}(i_4) = \alpha_{i_1 i_2 i_3} \sqrt{\frac{\kappa_{i_1 i_2 i_3}}{2\pi(i_4 - t_0)}} e^{-\frac{\kappa_{i_1 i_2 i_3}(i_4 - t_0 - \mu_{i_1 i_2 i_3})^2}{2(i_4 - t_0)}}, \tag{5.2}$$

where $\tilde{\mathbf{l}}_{i_1 i_2 i_3}(i_4)$ represents the i_4th time sample of (i_1, i_2, i_3) voxel before the log-compression operation of the DCEUS data.

Here, $\kappa_{i_1 i_2 i_3}$ is the local dispersion-related parameter independent of the injection site's distance. For low values of dispersion, a symmetric curve and a high $\kappa_{i_1 i_2 i_3}$ value are observed. This is expected to represent malignant regions [121] [55]. On the other hand, a low $\kappa_{i_1 i_2 i_3}$ is expected to represent the benign regions.

5.3. PROPOSED ALGORITHM

The data tensor $\underline{\mathbf{Y}}$ is a 4th-order DCEUS acquisition where the first three are the spatial mode in the Cartesian domain, and the fourth is the temporal mode. We recover $\underline{\mathbf{L}}$ by truncating $\underline{\mathbf{Y}}$ with ranks (R_1, R_2, R_3, R_4). This assumption is expected to hold since the movement of the microbubbles is bounded by the spatial locations of the vascular architecture inside the prostate, and their temporal characteristics are a latent variable of indicator dilution models. In addition, we

assume that the noise $\underline{\mathbf{E}}$ is independent of the signal itself. With these assumptions, the problem at hand transforms into a tensor rank estimation problem, where the rank that defines the signal subspace will be estimated, and the original signal will be recovered.

The multilinear truncation is done on each mode-n unfolding separately. This can be described as

$$\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{U}^{(1)}\mathbf{\Sigma}^{(1)}\mathbf{V}^{(1)^T} = \mathbf{\Sigma}^{(1)} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{V}^{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{U}^{(2)}\mathbf{\Sigma}^{(2)}\mathbf{V}^{(2)^T} = \mathbf{\Sigma}^{(2)} \times_1 \mathbf{U}^{(2)} \times_2 \mathbf{V}^{(2)} \\
\mathbf{Y}_{(3)} &= \mathbf{U}^{(3)}\mathbf{\Sigma}^{(3)}\mathbf{V}^{(3)^T} = \mathbf{\Sigma}^{(3)} \times_1 \mathbf{U}^{(3)} \times_2 \mathbf{V}^{(3)} \\
\mathbf{Y}_{(4)} &= \mathbf{U}^{(4)}\mathbf{\Sigma}^{(4)}\mathbf{V}^{(4)^T} = \mathbf{\Sigma}^{(4)} \times_1 \mathbf{U}^{(4)} \times_2 \mathbf{V}^{(4)}
\end{aligned}$$

Truncate each singular vector by taking first R_n elements for $n \in \{1, \ldots, 4\}$, in their respective modes. This can be described by the operation

$$\mathbf{Y}_{(n)} = \begin{bmatrix} \bar{\mathbf{U}}^{(n)} & \tilde{\mathbf{U}}^{(n)} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}^{(n)}_{R_n} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}^{(n)}_{I_n - R_n} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}^{(n)^T} \\ \tilde{\mathbf{V}}^{(n)^T} \end{bmatrix} \tag{5.3}$$

where $\bar{\mathbf{U}}^{(n)}$ is the column-wise stacking of first R_n vectors representing the singular vectors of the highest singular values, $\tilde{\mathbf{U}}^{(n)}$ is the column-wise stacking of the $I_n - R_n$ columns for $n \in \{1, \ldots, N\}$ that represent the singular vectors of the rest of the singular values, $\bar{\mathbf{V}}^{(n)^T}$ and $\tilde{\mathbf{V}}^{(n)^T}$ are defined in a similar way but represent the right singular vectors of the nth unfolding of $\underline{\mathbf{Y}}$.

The multilinear ranks of $\underline{\mathbf{L}}$ are estimated using the SCORE algorithm proposed by [109]. Let $\boldsymbol{\phi}^{(n)}$ be the squared sum of the columns of the mode-n unfolding of $\underline{\mathbf{S}}$, that is,

$$\boldsymbol{\phi}^{(n)} = \frac{1}{(D/I_n)} \text{diag}(\mathbf{S}^T_{(n)} \mathbf{S}_{(n)}), \tag{5.4}$$

where D/I_n represents the multiplication of the sizes of all orders except the nth order. Note that $\boldsymbol{\phi}^{(n)}$ is not the singular values, which is expressed as $\text{diag}(\mathbf{S}_{(n)}\, \mathbf{S}_{(n)})^T$. It is rather the normalized powers of the columns of $\mathbf{S}_{(n)}$. The most contributing columns are selected for the minimum description length estimation [127]. The intuition behind this is the following. There are I_n rows and D/I_n columns in $\mathbf{S}_{(n)}$. The columns of $\mathbf{S}_{(n)}$ represent the interaction of the combinations of the columns of the factor matrices from the other modes than n with the structure defined by the mode-n factor matrices. If a column has a high norm, that suggests the interaction of the mode-n factor matrix with the combination of the columns of the factor matrices other than mode-n is strong. If the coupling is not strong, we can say that the combination of the factor matrices does not contribute much to describing the original tensor. For example, the factor matrices that explain a region with outliers are not expected to be coupled with the other factor matrices of the low multilinear rank tensor. Hence, they will have a low column norm in $\boldsymbol{\phi}^{(n)}$. We can select and remove such columns to have a rank estimation technique that is robust to outliers.

In each mode, the removal of columns is executed by the parameter ρ, which is suggested to be between 0.0001 and 0.01. Let $\mathbf{B}^{(n)} \in \mathbb{R}^{D/I_n \times D/I_n}$ denote the

matrix that selects the $\rho D/I_n$ columns of $\mathbf{S}_{(n)}$ with the highest norm, while discarding the rest. We have the robust singular values $\mathbf{s}^{(n)}$ at the diagonal of $\mathbf{S}_{(n)}\,\mathbf{B}^{(n)}\,\mathbf{B}^{(n)^T}\mathbf{S}_{(n)}^T$, that is,

$$\mathbf{s}^{(n)} = \frac{1}{(\rho D/I_n)} diag(\mathbf{S}_{(n)}\,\mathbf{B}^{(n)}\,\mathbf{B}^{(n)^T}\mathbf{S}_{(n)}^T).$$

The rank $\hat{R}_n$ for each mode $n \in \{1,\ldots,4\}$ are estimated using the MDL criterion and $\mathbf{s}^{(n)}$,

$$\hat{R}_n = \underset{r}{\operatorname{argmin}} -2\log\left(\frac{\prod_{i=r+1}^{I_n}(s_i^{(n)})^{1/(I_n-r)}}{\frac{1}{I_n-r}\sum_{i=r+1}^{I_n} s_i^{(n)}}\right)^{\rho(D/I_n)(I_n-r)} + r(2I_n-r)\log(\rho(D/I_n)). \tag{5.5}$$

The truncation is done by projecting each mode to the column subspace represented by the estimated rank

$$\hat{\underline{\mathbf{L}}} = \underline{\mathbf{Y}} \times_1 \bar{\mathbf{U}}^{(1)}\bar{\mathbf{U}}^{(1)\mathsf{T}} \times_2 \bar{\mathbf{U}}^{(2)}\bar{\mathbf{U}}^{(2)\mathsf{T}} \times_3 \bar{\mathbf{U}}^{(3)}\bar{\mathbf{U}}^{(3)\mathsf{T}} \times_4 \bar{\mathbf{U}}^{(4)}\bar{\mathbf{U}}^{(4)\mathsf{T}},$$

where $\hat{\underline{\mathbf{L}}}$ represent the estimate of the TICs. We are going to compare SVD and tr-MLSVD through simulation and in-vivo analysis by analyzing the model fitting performance.

5.4. SIMULATION RESULTS

In this section, a theoretical analysis is done to compare the performances of SVD and tr-MLSVD for two noise scenarios. The TICs are generated according to the model described in (5.2). The speckle noise is added for the best-case scenario. Additionally, motion artifacts and white Gaussian noise are added for the worst-case scenario. The resulting signals are logarithmically compressed and then truncated using SVD and MLSVD. The ranks are estimated using two methods that are abbreviated as *mlsvd_min* and *mlsvd_score*. In the former, the rank that gives the least MSE is chosen, and in the latter, the SCORE algorithm proposed in [109] is used. The estimated signals are fit using the LDRW model given in (5.2) using the algorithm described in [55] after reverting the logarithmic compression. For each algorithm, the mean squared error is calculated by

$$\text{MSE} = \frac{1}{I_1 I_2 I_3 I_4}\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\sum_{i_3=1}^{I_3}\sum_{i_4=1}^{I_4}\left(\hat{l}_{i_1 i_2 i_3 i_4} - l_{i_1 i_2 i_3 i_4}\right)^2. \tag{5.6}$$

We consider a signal $\tilde{\underline{\mathbf{L}}} \in \mathbb{R}^{10\times 10\times 10\times 30}$ which holds TICs that obey the LDRW model as described in (5.2). The voxel size is 0.75 mm, and the time step is 4 seconds. The simulated region holds three different TICs, which have the parameters that are commonly observed in the literature [93]. Let the first malignant region be defined as a block located at the indices $i_1 \in \{2,3,4,5\}$, $i_2 \in \{2,3,4,5\}$, and $i_3 \in \{2,3,4,5\}$. For this region defined by the aforementioned indices $\kappa_{i_1 i_2 i_3} = 0.5 \pm 0.1$, $\mu_{i_1 i_2 i_3} = 30 \pm 1$ and $\alpha_{i_1 i_2 i_3} = 1000 \pm 10$. The second malignant region is located at the indices $i_1 \in \{2,3,4,5\}$, $i_2 \in \{6,7,8,9\}$, and $i_3 \in \{6,7,8,9\}$. For this region defined by the aforementioned indices

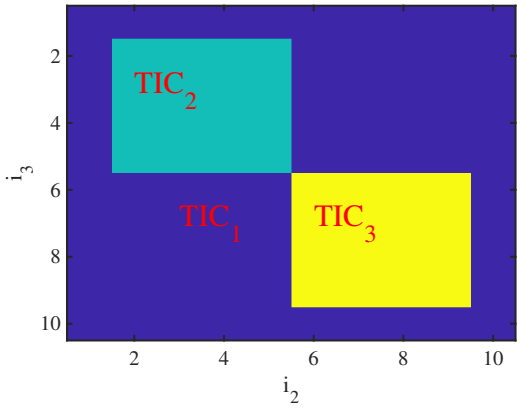


Figure 5.1: The simulation setup of a 3rd-order rectangular region with three different TICs. The slice at $i_1 = 2$ is shown. Three different simulation areas where the dark blue (majority of the slice) represents the first region, the light blue (left top) rectangle represents the second region, and the yellow part (bottom right) represents the third region.

$\kappa_{i_1 i_2 i_3} = 1 \pm 0.1$, $\mu_{i_1 i_2 i_3} = 25 \pm 1$ and $\alpha_{i_1 i_2 i_3} = 1600 \pm 10$. The benign region is the area that is not covered by the first two blocks. This region has $\kappa_{i_1 i_2 i_3} = 2 \pm 0.1$, $\mu_{i_1 i_2 i_3} = 20 \pm 1$, and $\alpha_{i_1 i_2 i_3} = 1200 \pm 10$. A slice at $i_1 = 2$ that represents these three different regions is provided in Fig. 5.1. An example of TICS denoted in regions TIC_1, TIC_2, and TIC_3 is plotted in Fig. 5.2. The y-axis is the TICs before log compression, denoted with $\tilde{\mathbf{I}}_{i_1 i_2 i_3}(i_4)$. The x-axis is the time stamps i_4 in seconds. The mean values of the TIC parameters $\kappa_{i_1 i_2 i_3}$, $\mu_{i_1 i_2 i_3}$, and $\alpha_{i_1 i_2 i_3}$ described in Table 5.1 are used to generate TIC_1, TIC_2, and TIC_3. During the simulation, the $\kappa_{i_1 i_2 i_3}$ is varied with ± 0.1, $\mu_{i_1 i_2 i_3}$ is varied with ± 1, and $\alpha_{i_1 i_2 i_3}$ with ± 10 to introduce heterogeneity inside the regions. The benign region (majority of the block) shown as dark blue is assigned as region 1, the light blue (left top) rectangle represents region 2, and the yellow part (bottom right) represents region 3. These numbers are used to refer to the malignant and benign regions. For example, TIC_1 will refer to the $\kappa_{i_1 i_2 i_3}, \mu_{i_1 i_2 i_3}, \alpha_{i_1 i_2 i_3}$ values inside the region 1, TIC_2 for region 2 and TIC_3 for region 3. The other parameters are represented in the same fashion. The true rank of the original signal $\underline{\mathbf{L}}$ (log-compressed $\underline{\tilde{\mathbf{L}}}$) in this setup is $(2, 3, 3, 3)$. There are three different TIC regions that create three distinctive components in the 2nd, 3rd, and the 4th modes. In the first mode, the rank is 2. This is because TIC_1 and TIC_2 share the same locations in the 1st mode, that is, the region defined by $i_1 \in \{2, 3, 4, 5\}$. Therefore, the information in mode-1 can be defined by two components, the TIC_3 and the region $i_1 \in \{2, 3, 4, 5\}$. In the 2nd and the 3rd modes, TIC_1 exists in the region $\{2, 3, 4, 5\}$ and TIC_2 exists in the region $\{6, 7, 8, 9\}$. Therefore, an extra component is required for differentiating the locations of three TICs, resulting in a rank of 3. A plot of the multilinear singular values of $\underline{\mathbf{Y}}$ is shown in Fig. 5.3. We plotted the modal singular values

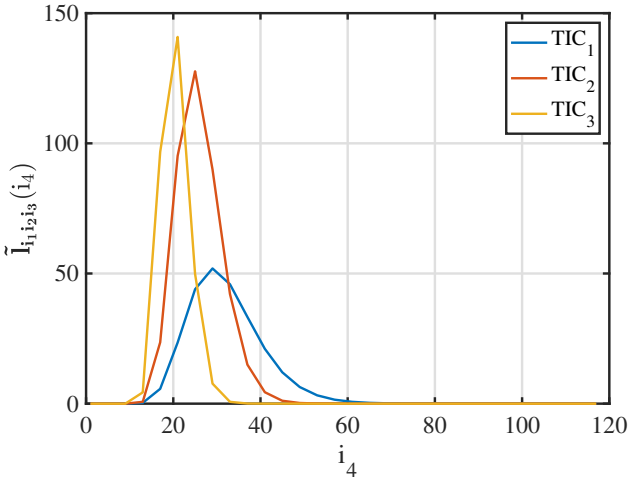


Figure 5.2: The three TICs generated in Section 5.4 with the LDRW model given in (5.2).

of $\underline{\mathbf{L}}$ with additive Fisher-Tippett noise. The singular values are shown on a logarithmic scale. We can see a linear decrease of modal singular values after $i_1 = 2$, $i_2 = 3$, $i_3 = 3$, and $i_4 = 3$, confirming the multilinear ranks.

Two noise scenarios are simulated. For Scenario 1, the TICs are noised with Rayleigh-shaped multiplicative noise. After the log-compression, the multiplicative noise is Fisher-Tippett distributed and becomes additive as described in Section 3.3. For Scenario 2, a variety of noise sources have been added. In addition to the multiplicative noise, we have added a sinusoidal breathing artifact with an amplitude of 0.5 mm and a frequency of 0.2 Hz, a random-walk displacement that has the maximum translation of 0.0125 mm at each step to simulate the probe-handling of the urologist before the log-compression. Additionally, a WGN with 4 dB SNR, where the SNR is calculated with respect to the averaged TICs is added after the log-compression. The error measures given in (5.6) are calculated and averaged across 100 Monte Carlo simulations for each case. A plot of MSE can be observed in Fig. 5.4 and Fig. 5.6 for Scenario 1 and Scenario 2, respectively. Subplots (a) show the MSE over 100 Monte Carlo simulations when the highest 1 to 7 principal components are used for truncation. Since tr-MLSVD does not share the same i_1 axis, the MSE is drawn as a straight line on the same plot. The line with the circle markers is the average MSE over 100 Monte Carlo simulations when the best-performing truncation is applied, and the line with cross markers is the MSE when the ranks are estimated using the SCORE algorithm. Subplots (b) and (c) represent the histogram of the ranks that give the least MSE and the rank estimated by the SCORE algorithm, respectively. In 87 percent of the cases, the correct rank is estimated by the SCORE algorithm, which can be seen in plot (c) of Fig. 5.4. For this case, the performance of *mlsvd_min* and *mlsvd_score* overlap, suggesting that the SCORE algorithm gives the least MSE over 100 Monte Carlo simulations. Although a performance improvement

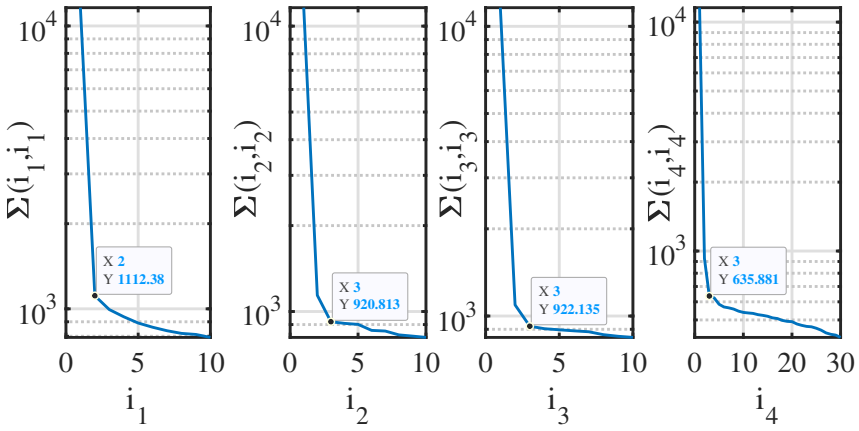


Figure 5.3: The singular values of $\underline{\mathbf{Y}}$ in each mode-n unfolding for $n \in \{1,\ldots,N\}$ defined in Section 5.4 Scenario 1. There is only additive Fisher-Tippett noise.

5

over SVD is observed in Fig. 5.6, the ranks are not estimated correctly when a variety of noise sources are added. The reasons are discussed in Section 5.5.

The estimated $\hat{\kappa}_{i_1 i_2 i_3}$ $\hat{\mu}_{i_1 i_2 i_3}$, and $\hat{\alpha}_{i_1 i_2 i_3}$ parameters are shown in Fig. 5.5 and Fig. 5.7, respectively. On the x-axis, the true values are shown in parentheses. On the y-axis, the mean and the standard deviations of the estimated TIC parameters are shown. A lower standard deviation is observed in the estimates in Scenario 1 compared to Scenario 2. The $\hat{\kappa}_{i_1 i_2 i_3}$ and $\hat{\mu}_{i_1 i_2 i_3}$ are more accurate compared to estimates of $\hat{\alpha}_{i_1 i_2 i_3}$. In both scenarios, we can observe that the tr-MLSVD and SVD are able to capture the three distinctive malignant and benign TIC features.

Table 5.1: Summary of simulation scenarios

	$TIC_1(\kappa_{i_1i_2i_3}, \mu_{i_1i_2i_3}, \alpha_{i_1i_2i_3})$	$TIC_2(\kappa_{i_1i_2i_3}, \mu_{i_1i_2i_3}, \alpha_{i_1i_2i_3})$	$TIC_3(\kappa_{i_1i_2i_3}, \mu_{i_1i_2i_3}, \alpha_{i_1i_2i_3})$	Number of Voxels TIC_1	Number of Voxels TIC_2	Number of Voxels TIC_3	Noise Types	Number of Simulations
Scenario 1	$(0.5 \pm 0.1, 30 \pm 1, 1000 \pm 10)$	$(1 \pm 0.1, 25 \pm 1, 1600 \pm 10)$	$(2 \pm 0.1, 20 \pm 1, 1200 \pm 10)$	872	64	64	Multiplicative	100
Scenario 2	$(0.5 \pm 0.1, 30 \pm 1, 1000 \pm 10)$	$(1 \pm 0.1, 25 \pm 1, 1600 \pm 10)$	$(2 \pm 0.1, 20 \pm 1, 1200 \pm 10)$	872	64	64	Multiplicative, Breathing, Motion, WGN	100

5.4.1. EXPERIMENTAL RESULTS

A recording of 6 patients was acquired from the Second Affiliated Hospital of Zhejiang University (Hangzhou, Zhejiang, China). Written consent was obtained. A 2.4-mL bolus of SonoVue® was intravenously injected, and a 4th-order recording in contrast mode was obtained with a LOGIQ E9 scanner equipped with a RIC5-9-D endocavitary transducer driven at 4 MHz. The volume rate was fixed to 0.25 Hz by setting the image quality to low, and the disruption of microbubbles was avoided by fixing the mechanical index to 0.1. The patients underwent prostatectomy after the recording. The prostate was sliced with 4-mm thickness, and for each slice, an annotation was made by the pathologist. The annotations were

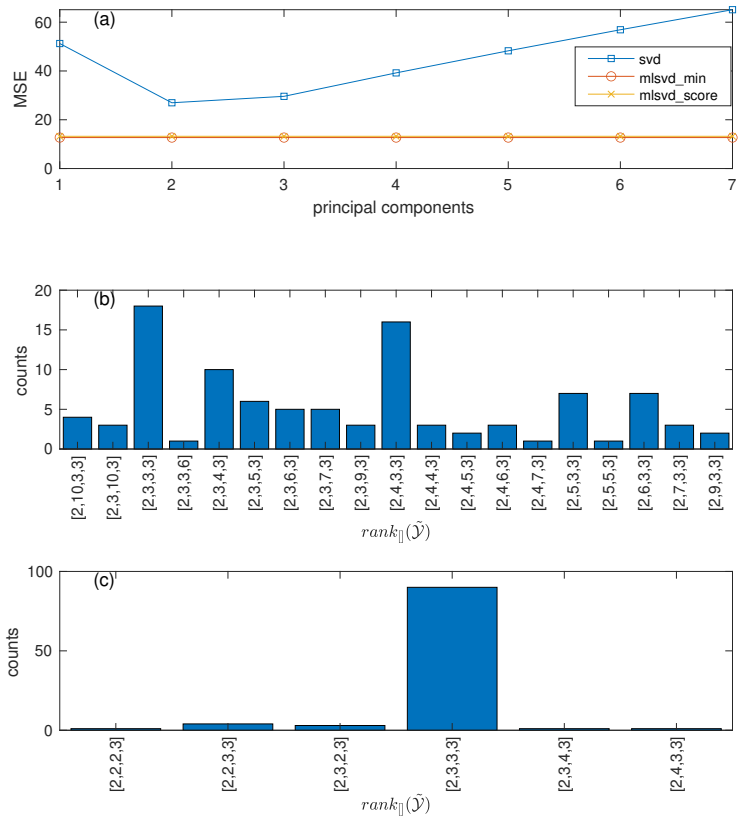


Figure 5.4: The performance comparison for Scenario 1. In (a), the MSE of SVD when several principal components are used to truncate is shown along with two lines that show the best-performing tr-MLSVD error over 100 Monte Carlo simulations and the performance of tr-MLSVD after the ranks are estimated with the SCORE algorithm. In (b), the histogram of the multilinear ranks that give the least MSE after tr-MLSVD at each simulation is shown. In (c), the histogram of the ranks that are estimated with the SCORE algorithm is shown.

registered back to the domain of the recording, and the ground truth was obtained. Only two out of the six patients had significant lesions with Grade Group > 3. A region of interest is selected in the benign and malignant regions that are reasonably close to the true annotations. There are around 25 000 voxels, of which 18 000 are malignant.

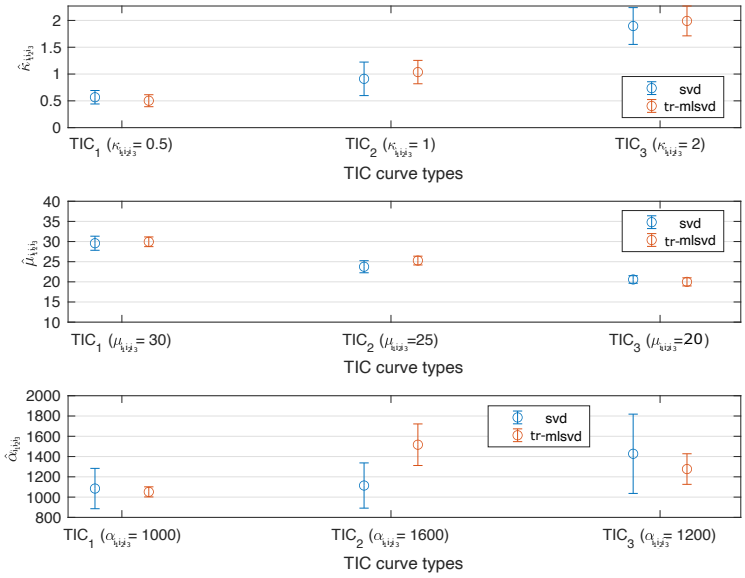


Figure 5.5: Plot that shows the mean and the standard deviation of the $\hat{\kappa}_{i_1 i_2 i_3}$, $\hat{\mu}_{i_1 i_2 i_3}$, and $\hat{\alpha}_{i_1 i_2 i_3}$ for three different TICs in Scenario 1. The circles represent the mean values, whereas the error bars represent the standard deviation. The mean and the standard deviation are calculated over 100 Monte Carlo simulations.

Table 5.2: The estimated model fit parameters across all patients. The mean and the standard deviations are shown.

	SVD		MLSVD-SCORE $\rho = 10^{-4}$	
	Benign	Malignant	Benign	Malignant
MFR-κ	0.51($\sigma = 0.35$)	0.56($\sigma = 0.33$)	0.39($\sigma = 0.18$)	0.48($\sigma = 0.22$)
MFR-PV	94.73($\sigma = 29.57$)	104.05($\sigma = 27.53$)	87.87($\sigma = 26.33$)	98.49($\sigma = 24.03$)
MFR-PT	33.76($\sigma = 9.57$)	30.00($\sigma = 8.09$)	32.56($\sigma = 6.95$)	29.24($\sigma = 5.62$)
MFR-AT	15.12($\sigma = 4.96$)	13.72($\sigma = 4.23$)	14.32($\sigma = 3.44$)	13.41($\sigma = 3.11$)
MFR-WIT	18.64($\sigma = 9.41$)	16.28($\sigma = 7.44$)	18.24($\sigma = 5.91$)	15.84($\sigma = 4.08$)
MFR-μ	35.28($\sigma = 11.48$)	31.31($\sigma = 9.57$)	34.29($\sigma = 8.31$)	30.58($\sigma = 6.24$)
MFR-α	172.99($\sigma = 126.38$)	184.39($\sigma = 123.05$)	143.78, ($\sigma = 77.96$)	154.09, ($\sigma = 68.92$)

The spatial resolution of the recording is regularized in space, and the data is downsampled by a factor of 3 as described by [64]. The MLSVD is applied, and the signal is truncated using the ranks estimated by the SCORE [109] algorithm, where $\rho = 10^{-4}$ is selected. The LDRW model is fit as described in [55], and the perfusion and dispersion parameters are extracted. The results can be seen in Table 5.2 and Table 5.3. In the former, the mean and the standard deviation of the features are shown. In the latter, the sensitivity, specificity, and area under the receiver operating characteristic curves are shown. The classification

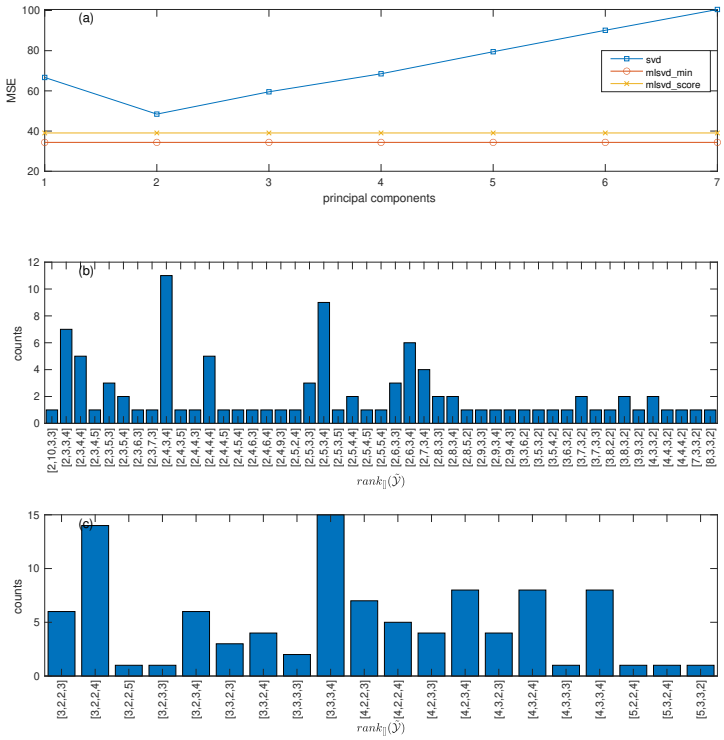


Figure 5.6: The performance comparison for Scenario 2. In (a), the MSE of SVD when several principal components are used to truncate is shown along with two lines that show the best-performing tr-MLSVD error over 100 Monte Carlo simulations and the performance of tr-MLSVD after the ranks are estimated with the SCORE algorithm. In (b), the histogram of the ranks that give the least MSE at each simulation is shown. In (c), the histogram of the ranks estimated with the SCORE algorithm is shown.

is done by determining the point in the receiver operating characteristic curve that is the closest to the upper left corner (sensitivity and specificity equal to 1) in Euclidean distance. In Table 5.2, we can observe that the difference between the means of the malignant and benign voxels increased after denoising with tr-MLSVD. Furthermore, the standard deviation of the features is less for tr-MLSVD, suggesting a more precise estimation of benign and malignant features. In Table 5.3, six out of seven model fit features have a higher ROC-AUC metric when tr-MLSVD is used. The model fit features that have a high ROC-AUC metric are

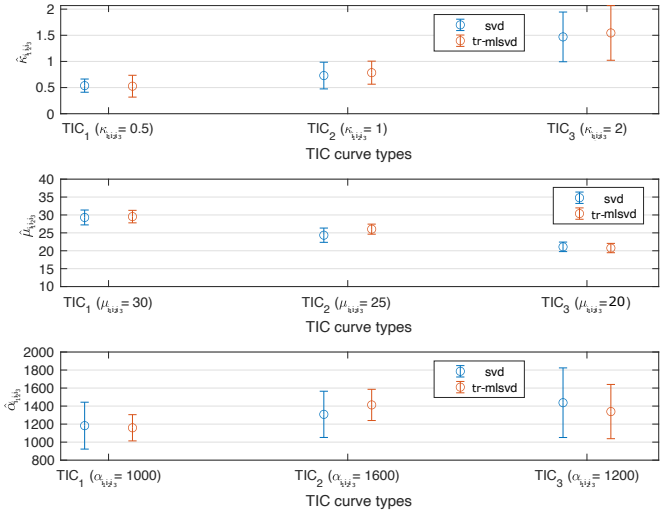


Figure 5.7: Plot that shows the mean and the standard deviation of the $\hat{\mu}_{i_1 i_2 i_3}$, $\hat{\kappa}_{i_1 i_2 i_3}$, and $\hat{\alpha}_{i_1 i_2 i_3}$ for three different TICs at Scenario 2. The circles represent the mean values, whereas the error bars represent the standard deviation. The mean and the standard deviation are calculated over 100 Monte Carlo simulations.

Table 5.3: The classification performance across all the patients.

	SVD	MLSVD-SCORE ($\rho = 10^{-4}$)
[MFR-κ_{sens}, MFR-κ_{spec}, MFR-κ_{AUC}, *threshold*]	[0.54,0.53,0.57,0.51]	[0.48,0.65,**0.63**,0.45]
[MFR-PI_{sens}, MFR-PI_{spec}, MFR-PI_{AUC}, *threshold*]	[0.60,0.54,0.61,99]	[0.63,0.57,**0.64**,93]
[MFR-PT_{sens}, MFR-PT_{spec}, MFR-PT_{AUC}, *threshold*]	[0.57,0.62,0.61,31]	[0.59,0.65,**0.63**,30]
[MFR-α_{sens}, MFR-α_{spec}, MFR-α_{AUC}, *threshold*]	[0.54,0.51,0.55,150]	[0.46,0.64,**0.58**,160]
[MFR-WIR_{sens}, MFR-WIR_{spec}, MFR-WIR_{AUC}, *threshold*]	[0.58,0.59,0.63,6.6]	[0.62,0.59,**0.66**,5.7]
[MFR-AT_{sens}, MFR-AT_{spec}, MFR-AT_{AUC}, *threshold*]	[0.65,0.58,**0.59**,15]	[0.72,0.59,0.58,13.8]
[MFR-WIT_{sens}, MFR-WIT_{spec}, MFR-WIT_{AUC}, *threshold*]	[0.58,0.57,0.56,16]	[0.59,0.58,0.**62**,16.5]

shown in bold.

5.5. DISCUSSION

In this chapter, the DCEUS acquisitions are denoised using SVD and tr-MLSVD, and their performances are evaluated based on the quality of the TIC modeling. In Section 5.4, two simulations are reported that employ the commonly observed noises in DCEUS acquisitions. The first simulation represents the best-case scenario where only speckle noise exists. The second simulation represents the worst-case scenario where additive WGN and motion exist apart from the speckle noise. The tr-MLSVD mostly performs better at capturing the signal characteristics in both cases, providing a better estimate of dispersion and perfusion

parameters. Both methods perform worse in the worst-case scenario and fail to estimate high κ values precisely. The low SNR might explain the decrease in performance. Another reason is the sampling frequency. We use 0.25 Hz, which is the same frequency that is in a LOGIQ E9 device. High values of κ result in sudden changes around the peak value, which causes an error in LDRW fitting. On the other hand, the estimates $\hat{\alpha}_{i_1 i_2 i_3}$ are observed to be less accurate and precise. The reason is the windowing of the TICs that is proposed in [55]. A window is applied to estimate the TICS between the appearance time and the time that the TICs reach 5% of the peak intensity. This is done to prevent the effect of the recirculation of UCAs in the circulatory system. The fitting of the first-pass is shown to improve the diagnosis. However, in our case, this increases the error in the estimates of $\alpha_{i_1 i_2 i_3}$, which is related to the area under the TICs.

The speckle noise can be modeled as a WGN with outliers [92], which results in the correct estimation of the rank for Scenario 1. The actual rank of the signal is $(2, 3, 3, 3)$, which is the same as the rank that is estimated the majority of the time. For this case, we can see in Fig. 5.4 that the MSE of the best-performing rank and the estimated rank overlap. In the cases where the rank was incorrectly estimated, the error is found to be within 15% of the MSE represented with the line *mlsvd_score*. The estimation of the correct rank fails for Scenario 2. Adding the other noise sources violates the WGN assumption and causes the rank estimator to perform poorly. This suggests that in a practical scenario where the recording suffers from various noise sources apart from the speckle noise, the proposed algorithm might not be suitable. In the figures Fig. 5.4 and Fig. 5.6, it can be observed that the best-performing rank is not necessarily the true rank of the data. The subfigures labeled with (b) in both Plots suggest that the noise, in some cases, might occupy the highest ranks. For this reason, a higher rank might have provided a lower MSE.

Better classification performance of tr-MLSVD over SVD is observed in Table 5.3. The improvement is more significant when a subset of the patients is considered. This might be due to the insignificant malignant voxels of the four patients that are investigated. The significant malignant lesions are expected to have a more distinct characteristic than benign tissue. This behavior is observed in two of the patients who have significant lesions. The AUC is found to be greater than 0.8 for these two patients, whereas a similar performance has not been observed for the others.

5.6. CONCLUSION

In this chapter, we proposed a denoising algorithm for detecting prostate cancer from 4th-order DCEUS acquisitions. Previously, SVD was proposed to denoise TICs by forming a spatiotemporal matrix. Here, we retained the volumetric information by considering the tensor format of the recording and introduced an algorithm based on tr-MLSVD and SCORE. The proposed algorithm outperformed SVD in simulation and in in-vivo experiments. In general, the classification performance of the in-vivo data was poor, which can be due to the insignificant lesions found in four of the six patients who underwent prostatectomy. In Chapter 6, we plan to use a larger dataset, compare the tr-MLSVD with other despeckling methods, and include other features to detect angiogenesis.

ACKNOWLEDGEMENTS

This project is funded in part by Holland High Tech with a PPS supplement for research and development in the Topsector HTSM. We would like to acknowledge Prof. Pintong Huang for carrying out the clinical trials at the Second Affiliated Hospital of Zhejiang University (Hangzhou, Zhejiang, China).

5.7. CHAPTER SUMMARY

In this chapter, we investigated the rank estimation methods for DCEUS acquisitions and compared tr-MLSVD with the SVD-based denoising using the ROC-AUC metric. Only the model-fitting, as explained in Section 3.2.2, was used. For the rank estimation, the approximation of speckle as WGN resulted in high ranks. Therefore, we investigated the SCORE algorithm [109] that is better suited for WGN with sparse outliers. Using the estimated ranks, we used tr-MLSVD to recover the denoised DCEUS acquisitions.

The prostate classification performance improved when compared to SVD. The patient sample size was small. We apply the tr-MLSVD algorithm on a bigger patient cohort in Chapter 6. Furthermore, we compare its performance with OITNN-O and a gradient-based algorithm incorporating the Fisher-Tippet distribution into the low multilinear rank approximation. This chapter and Chapter 6 aim to answer the second research question.

6

SPECKLE DENOISING OF DYNAMIC CONTRAST-ENHANCED ULTRASOUND: A COMPARISON OF NOISE MODELS AND ALGORITHMS

In the previous chapter, we investigated the tr-MLSVD for denoising DCEUS. We incorporated a rank estimator that was suitable for WGN with outliers. The tr-MLSVD did not consider the sparse outliers. In this chapter, we investigate the OITNN-O that is suitable for WGN with sparse outliers. The DCEUS acquisitions are 4th-order tensors. The OITNN-O is an orientation invariant convex optimization method that assumes low-tubal rankness in each mode-$(n, n+1)$ unfolding introduced in Chapter 4. The 1D Fourier transform is always applied to the last mode of the mode-$(n, n+1)$ unfoldings, exploiting the spatial and temporal low tubal rankness of the tensor.

In addition to OITNN-O, we apply a non-convex method that incorporates the log-likelihood of the Fisher-Tippet distribution. The prostate cancer classification performance is compared to other state-of-the-art approaches in the literature using the ROC-AUC metric on annotated data from 94 prostate cancer patients. This chapter partially answers the second research question.

This chapter is based on the article published as "Speckle Denoising of Dynamic Contrast-enhanced Ultrasound using Low- rank Tensor Decomposition." by M. Calis, M. Mischi, A.-J. van der Veen, and B. Hunyadi, In: IEEE Transactions on Medical Imaging (TMI), 2025

6.1. INTRODUCTION

Dynamic contrast-enhanced ultrasound (DCEUS) is an imaging modality that enables the characterization of the blood perfusion patterns through the microvasculature [77]. DCEUS is investigated for many applications, such as the localization of prostate cancer. With the help of intravenously injected microbubbles, 4th-order DCEUS enables the visualization and analysis of the entire prostate gland. The newly formed angiogenic vessels associated with the tumor growth create a distinctive and irregular microvascular architecture that is used as an indicator of prostate cancer [128]. Through analysis of the DCEUS acquisitions by convective dispersion modeling, a number of quantitative dispersion parameters that reflect angiogenic changes in the underlying microvascular architecture have been proposed. The time evolution of the microbubbles inside the vasculature, also called the time-intensity curves (TICs), are fit with the local density random walk model, and distinctive perfusion and dispersion dynamics have been observed in malignant and benign regions [61]. Indeed, a significant amount of work has been published that proves the benefit of using contrast-ultrasound dispersion techniques to identify angiogenesis [55, 65, 69, 72].

The general framework for the classification of angiogenesis consists of several steps. Firstly, the microbubbles are administrated intravenously, and the Digital Imaging and Communications in Medicine (DICOM) data is recorded. Commercial ultrasound scanners utilize various algorithms such as harmonic imaging, phase inversion, or amplitude modulation to enhance the non-linear echoes from the microbubbles while suppressing the approximately linear echoes from the tissue, especially at the employed low ultrasound pressure [129]. The log-compressed envelope of the resulting radio-frequency data is exported for further preprocessing, feature extraction, and classification. Even after the aforementioned contrast enhancement techniques, DCEUS acquisitions suffer from speckle noise. This noise results from the coherent imaging of microbubbles in a resolution cell. Assuming that a large number of randomly distributed scatterers exist in this cell and there are no strong reflectors, the real and the imaginary parts of the complex echo can be modeled by a zero-mean Gaussian density [87, 92, 126]. The magnitude of this echo is Rayleigh distributed, where the mean is proportional to the standard deviation of the complex echo. This creates multiplicative noise, which becomes additive in the log domain after the log compression performed by the scanner. At lower bubble concentrations, noise is more dominated by the pharmacokinetic statistics of the bubbles, i.e., their probability of being in or outside the cell [58].

Many existing features are initially developed and tested in 2D data [130] [57]. Although most of the features are extended for the 3rd-order space [65] [69] [67], the denoising algorithms do not take the higher-order structure of the recording into consideration. The two most widely used denoising techniques for DCEUS acquisitions include a wavelet [87], and an SVD-based [93] algorithm. The authors in [87] model the log-transformed noise as WGN with outliers. They apply the robust smoother-cleaner [131] to the finest level of decomposition using a median filter with a window of five and biorthogonal wavelets with three and nine vanishing moments. The resulting signal is denoised using a regular wavelet shrinkage algorithm with soft thresholding. In [74], a low-pass filter is applied to TICs where the cut-off frequency is set to be 0.5 Hz. In [93], the authors investigated various matrix decomposition techniques and found that

the truncated singular value decomposition (SVD) gives the best results for the localization of prostate cancer. The assumption is that distinct columns of the resulting factor matrices will capture the signal and noise subspace. Denoising is achieved by reconstructing the data after setting the singular values of the noise subspace to zero. The DCEUS acquisition is flattened into a spatiotemporal matrix where the columns correspond to time, and the rows correspond to space to apply SVD. However, this flattening removes spatial information and destroys the 3D structure. Therefore, previously in Chapter 5, we applied the multilinear singular value decomposition to the DCEUS acquisition and truncated the factor matrices at each unfolding by ranks estimated by a robust information-theoretic method [109]. When the dispersion and perfusion features were used, a slightly better separation between the malignant and benign regions was observed compared to [93]. We further showed in [2] an improved classification performance in a multi-parametric setting. The multilinear singular value decomposition aims to decompose the tensor into its orthonormal basis through least squares. This type of decomposition is suitable for WGN [13].

We implement and compare various low-rank tensor-based denoising techniques that incorporate the Fisher-Tippett assumption on the speckle noise. First, we propose a tensor-based denoising algorithm that uses the prior noise distribution to estimate a low-rank tensor via gradient descent by utilizing the general estimation framework (GTE) [132]. The difference between the proposed algorithm and [132] is twofold. The derivatives are calculated more efficiently, enabling their application for large recordings, and the noise is distributed by Fisher-Tippett. In Chapter 4, we investigated several low-rank tensor denoising techniques for denoising Fisher-Tippett noise through simulation and found out that orientation invariant tensor nuclear norm [30] performs well, especially for low SNR scenarios. Hence, we apply the orientation invariant tensor nuclear norm (OITNN) algorithm to model the log-transformed DCEUS acquisitions as a low-rank tensor with sparse and WGN, which has not been done before. The performance of the aforementioned tensor techniques will be compared with the denoising techniques proposed in the literature, and the performance of prostate cancer localization will be reported. We incorporate a linear classifier for each feature and report the area under the ROC curve.

In the literature, low-rank despeckling techniques are commonly applied to the nonlocal patches extracted from the ultrasound recording [100, 133, 134]. The DCEUS acquisitions are different than the fundamental mode ultrasound images. Rather than imaging a static morphology, in DCEUS, moving bubbles inside the vasculature are imaged. Therefore, we take a different approach motivated by [93] where SVD is applied to the spatiotemporal matrix generated by flattening the whole DCEUS acquisition. We are motivated to model low-rankness in spatial modes separately for several reasons. The voxels around a cancerous region show similar dispersion and perfusion characteristics [56][65]. Significantly grown cancerous tumors will create spatial regions that can be captured by the low-rank approximation. In addition, we postulate that the orthogonality assumption in the factor matrices will aid the separation between the tissue and the microvasculature, as well as the cancerous TICs and the benign TICs.

The chapter is organized as follows. Section 6.2 formulates the problem and describes the tensor-based denoising algorithms. Section 6.3 introduces the simulation, in-vitro, and in-vivo setups. Section 6.4 reports the simulation and in-vitro

results and the discriminative power of the features extracted from the DCEUS acquisitions of the prostate. We elaborate on the results in Section 6.5 and conclude with possible future work in Section 6.6.

6.2. LOW-RANK APPROXIMATION OF DCEUS ACQUISITIONS

6.2.1. SIGNAL MODEL

The DCEUS acquisitions are 4th-order recordings where the first three modes represent the spatial domain, and the last mode represents the time domain. Due to the difference between the resolution cell and the size of the microbubbles, the recordings suffer from speckle noise. Let $\underline{\tilde{\mathbf{M}}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ represent the speckle noise where each entry is Rayleigh distributed with a scaling parameter of 1. The DCEUS acquisition, represented by $\underline{\tilde{\mathbf{Y}}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$, is the multiplication of the parameter of interest $\underline{\tilde{\mathbf{L}}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ with $\underline{\tilde{\mathbf{M}}}$. This can be shown as

$$\underline{\tilde{\mathbf{Y}}} = \underline{\tilde{\mathbf{L}}} \odot \underline{\tilde{\mathbf{M}}}. \tag{6.1}$$

It has been shown in [55] [87] that there is a linear relation between the microbubble concentration and $\underline{\tilde{\mathbf{L}}}$. Hence, the magnitude of each entry at $\underline{\tilde{\mathbf{L}}}$ can be used as an indirect measure of the bubble concentration.

Commonly, in ultrasound devices, logarithmic compression is applied for visualization purposes. This operation is given by

$$a^{(1)} \ln \underline{\tilde{\mathbf{Y}}}, \tag{6.2}$$

where

$$a^{(1)} = 255 \log_{10}(e) 10 / a^{(3)} \tag{6.3}$$

is a function of the dynamic range $a^{(3)}$. We ignored the gain-related transformation $a^{(2)}$ since it does not affect the probability density function apart from a shift. The logarithmic compression and the signal model are explained in Section 3.3. This operation affects the probability density function of the noise. We have

$$a^{(1)} \ln \underline{\tilde{\mathbf{Y}}} = a^{(1)} \ln \underline{\tilde{\mathbf{L}}} + a^{(1)} \ln \underline{\tilde{\mathbf{M}}}. \tag{6.4}$$

Let $a^{(1)} \ln \underline{\tilde{\mathbf{M}}} = \underline{\mathbf{M}}$, $a^{(1)} \ln \underline{\tilde{\mathbf{L}}} = \underline{\mathbf{L}}$ and $a^{(1)} \ln \underline{\tilde{\mathbf{Y}}} = \underline{\mathbf{Y}}$. With this notation, we will have

$$\underline{\mathbf{Y}} = \underline{\mathbf{L}} + \underline{\mathbf{M}}. \tag{6.5}$$

The log-transformed Rayleigh noise follows the Fisher-Tippett distribution [92]. The goal of this chapter is to recover the DCEUS acquisition $\underline{\mathbf{L}}$ from the signal model given in (6.5), assuming that the recording shows a low-rank structure. We introduce two algorithms and compare them with the literature on DCEUS denoising. These algorithms differ in the assumption of the noise and the low-rank structure. These algorithms are

- general tensor estimation framework (GTE) [132] that assumes $\underline{\mathbf{M}}$ follows Fisher-tippet distribution and $\underline{\mathbf{L}}$ is low-rank in mode-n unfoldings,
- orientation invariant tubal nuclear norm (OITNN) [30] that assumes $\underline{\mathbf{L}}$ is low tubal rank in mode-$(n, n+1)$ unfoldings and models the noise $\underline{\mathbf{M}}$ as the summation of a WGN and sparse outliers.

The truncated multilinear singular value decomposition described in Chapter 5 is used as a warm initialization for the GTE algorithm.

6.2.2. GENERAL TENSOR ESTIMATION FRAMEWORK

The authors in [132] proposed a low-rank tensor estimation framework that incorporates the prior distribution of the noise. They solve

$$\hat{\underline{\mathbf{L}}} = \underset{\underline{\mathbf{L}}\ is\ low\ rank}{\operatorname{argmin}} \; -\ln(p(\underline{\mathbf{Y}};\underline{\mathbf{L}})), \tag{6.6}$$

where $p(\underline{\mathbf{Y}};\underline{\mathbf{L}})$ is the probability distribution function of the tensor $\underline{\mathbf{Y}}$ parameterized by the low multilinear rank tensor $\underline{\mathbf{L}}$. This algorithm is named the general tensor estimation framework (GTE). In our case, the $\underline{\mathbf{Y}}$ is the 4th-order DCEUS acquisition. The authors in [132] solve the problem for Poisson, Gaussian, and Binomial noise. The authors initialize the algorithm with the tr-MLSVD estimate using ranks R_n for $n \in \{1,\dots,N\}$, where tr-MLSVD is explained in Section 2.3.1. The initial factor matrices and the core tensor are denoted with $\mathbf{U}^{(1)^0}$, $\mathbf{U}^{(2)^0}$, $\mathbf{U}^{(3)^0}$, $\mathbf{U}^{(4)^0}$ and $\underline{\mathbf{S}}^0$. In addition to the maximum likelihood, the authors in [132] proposed to add the term

$$\frac{a}{2}\sum_{n=1}^{4}\left\|\left(\mathbf{U}^{(n)}\right)^{\mathsf{T}}\mathbf{U}^{(n)} - b^2\,\mathbf{I}\right\|_F^2, \tag{6.7}$$

to prevent the factor matrices from being singular throughout the gradient descent and enforce the low multilinear rankness with ranks R_n for $n \in \{1,\dots,N\}$. Let the loss function be denoted by F and the iteration number by k. With the addition of these terms, the loss function F becomes

$$F(\underline{\mathbf{Y}},\underline{\mathbf{L}}^k) = -\ln p\left(\underline{\mathbf{Y}};\underline{\mathbf{L}}^k\right) + \frac{a}{2}\sum_{n=1}^{4}\left\|\left(\mathbf{U}^{(n)^{\mathsf{T}}}\right)^k\mathbf{U}^{(n)^k} - b^2\,\mathbf{I}\right\|_F^2, \tag{6.8}$$

where $\underline{\mathbf{L}}^{k+1} = \underline{\mathbf{S}}^k \times_1 \mathbf{U}^{(1)^k} \times_2 \mathbf{U}^{(2)^k} \times_3 \mathbf{U}^{(3)^k} \times_4 \mathbf{U}^{(4)^k}$.

The two regularization weights a and b given in (6.8) are selected using the spectral norm of the initial estimate $\underline{\mathbf{L}}^0$. Four spectral norms can be defined using the four unfoldings of $\underline{\mathbf{L}}^0$, which is the tensor formed from the initialized factor matrices and the core tensor. Let q denote the maximum spectral norm, that is,

$$q = \max\left(\|\,\mathbf{L}^0_{(1)}\,\|,\ \|\,\mathbf{L}^0_{(2)}\,\|,\ \|\,\mathbf{L}^0_{(3)}\,\|,\ \|\,\mathbf{L}^0_{(4)}\,\|\right). \tag{6.9}$$

Two regularization weights are assigned as $b = q^{1/4}$ and $a = q$. The initial factor matrices $\{\mathbf{U}^{(1)^0}\dots\mathbf{U}^{(4)^0}\}$ are multiplied by $q^{1/4}$, whereas the core tensor $\underline{\mathbf{S}}^0$ is divided by q. This is done to guarantee the local convergence with a high probability [132]. The authors update the factor matrices and the core tensor through gradient descent with a step size η. Define the element-wise derivative operator as $\frac{\partial F}{\partial \underline{\mathbf{J}}} : \mathbb{R}^{I_1\times I_2\times I_3\times I_4} \to \mathbb{R}^{I_1\times I_2\times I_3\times I_4}$ for the derivative of the scalar valued

function F with respect to $\underline{\mathbf{J}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$. We have the gradient steps of GTE as

$$\begin{aligned}
\mathbf{U}^{(1)^{k+1}} &= \mathbf{U}^{(1)^k} - \eta \frac{\partial F}{\partial \mathbf{U}^{(1)^k}}, \\
\mathbf{U}^{(2)^{k+1}} &= \mathbf{U}^{(2)^k} - \eta \frac{\partial F}{\partial \mathbf{U}^{(2)^k}}, \\
\mathbf{U}^{(3)^{k+1}} &= \mathbf{U}^{(3)^k} - \eta \frac{\partial F}{\partial \mathbf{U}^{(3)^k}}, \\
\mathbf{U}^{(4)^{k+1}} &= \mathbf{U}^{(4)^k} - \eta \frac{\partial F}{\partial \mathbf{U}^{(4)^k}}, \\
\underline{\mathbf{S}}^{k+1} &= \underline{\mathbf{S}}^k - \eta \frac{\partial F}{\partial \underline{\mathbf{S}}^k}.
\end{aligned} \tag{6.10}$$

The derivatives in (6.10) can be expressed with

$$\begin{aligned}
\frac{\partial F}{\partial \mathbf{U}^{(1)^k}} &= \left(\frac{\partial F}{\partial \underline{\mathbf{L}}^k}\right)_{(1)} \left(\mathbf{U}^{(4)^k} \otimes \mathbf{U}^{(3)^k} \otimes \mathbf{U}^{(2)^k}\right) \left(\mathbf{S}_{(1)}{}^k\right)^\top + a\mathbf{U}^{(1)^k} \left(\left(\mathbf{U}^{(1)^k}\right)^\top \mathbf{U}^{(1)^k} - b^2 \mathbf{I}\right), \\
\frac{\partial F}{\partial \mathbf{U}^{(2)^k}} &= \left(\frac{\partial F}{\partial \underline{\mathbf{L}}^k}\right)_{(2)} \left(\mathbf{U}^{(4)^k} \otimes \mathbf{U}^{(3)^k} \otimes \mathbf{U}^{(1)^k}\right) \left(\mathbf{S}_{(2)}{}^k\right)^\top + a\mathbf{U}^{(2)^k} \left(\left(\mathbf{U}^{(2)^k}\right)^\top \mathbf{U}^{(2)^k} - b^2 \mathbf{I}\right), \\
\frac{\partial F}{\partial \mathbf{U}^{(3)^k}} &= \left(\frac{\partial F}{\partial \underline{\mathbf{L}}^k}\right)_{(3)} \left(\mathbf{U}^{(4)^k} \otimes \mathbf{U}^{(2)^k} \otimes \mathbf{U}^{(1)^k}\right) \left(\mathbf{S}_{(3)}{}^k\right)^\top + a\mathbf{U}^{(3)^k} \left(\left(\mathbf{U}^{(3)^k}\right)^\top \mathbf{U}^{(3)^k} - b^2 \mathbf{I}\right), \\
\frac{\partial F}{\partial \mathbf{U}^{(4)^k}} &= \left(\frac{\partial F}{\partial \underline{\mathbf{L}}^k}\right)_{(4)} \left(\mathbf{U}^{(3)^k} \otimes \mathbf{U}^{(2)^k} \otimes \mathbf{U}^{(1)^k}\right) \left(\mathbf{S}_{(4)}{}^k\right)^\top + a\mathbf{U}^{(4)^k} \left(\left(\mathbf{U}^{(4)^k}\right)^\top \mathbf{U}^{(4)^k} - b^2 \mathbf{I}\right), \\
\frac{\partial F}{\partial \underline{\mathbf{S}}^k} &= \frac{\partial F}{\partial \underline{\mathbf{L}}^k} \times_1 \left(\mathbf{U}^{(1)^k}\right)^\top \times_2 \left(\mathbf{U}^{(2)^k}\right)^\top \times_3 \left(\mathbf{U}^{(3)^k}\right)^\top \times_4 \left(\mathbf{U}^{(4)^k}\right)^\top.
\end{aligned} \tag{6.11}$$

An additional projection step after each iteration (6.10) is proposed in [132]. However, the authors claim that for a small enough step size η, the projection step can be avoided. The difference between [132] and the proposed method is the characteristics of the noise and the calculation of derivatives. We extend the analysis to the Fisher-Tippet noise and calculate the derivatives of the loss function without Kronecker products, which are shown in Appendix 6.8.2.

PROBABILITY DISTRIBUTION OF THE FISHER-TIPPET NOISE

The Fisher-Tippett distribution is explained in Section 3.3. Let the probability density function of $\underline{\mathbf{M}}$ be the Rayleigh distribution with a scaling parameter of 1, that is,

$$p(\underline{\mathbf{M}}; 1) = \prod_{i_1=1}^{I_1} \prod_{i_2=1}^{I_2} \prod_{i_3=1}^{I_3} \prod_{i_4=1}^{I_4} m_{i_1 i_2 i_3 i_4} \exp \frac{-m_{i_1 i_2 i_3 i_4}^2}{2}, \tag{6.12}$$

assuming that the noise is independent between voxels. The log compression, as described in (6.4), changes the probability density function of the noise. The

Fisher-Tippett distribution [92] with the log compression is

$$p(\underline{\mathbf{Y}};\underline{\mathbf{L}}) = \prod_{i_1=1}^{I_1}\prod_{i_2=1}^{I_2}\prod_{i_3=1}^{I_3}\prod_{i_4=1}^{I_4} \frac{\exp(2(y_{i_1i_2i_3i_4} - l_{i_1i_2i_3i_4})/a^{(1)} - \frac{\exp 2(y_{i_1i_2i_3i_4} - l_{i_1i_2i_3i_4})/a^{(1)}}{2})}{a^{(1)}}. \tag{6.13}$$

The argument that minimizes the negative log-likelihood of (6.13) is the maximum likelihood estimate of $\underline{\mathbf{L}}$ with Fisher-tippet noise.

GRADIENT DESCENT

If we incorporate the Fisher-Tippett probability distribution into (6.8), we have

$$\begin{aligned} F(\underline{\mathbf{Y}},\underline{\mathbf{L}}) &= -\ln p\left(\underline{\mathbf{Y}};\underline{\mathbf{L}}\right) + \frac{a}{2}\sum_{n=1}^{4}\left\|\left(\mathbf{U}^{(n)}\right)^{\top}\mathbf{U}^{(n)} - b^2\mathbf{I}\right\|_F^2 \\ &= \sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\sum_{i_3=1}^{I_3}\sum_{i_4=1}^{I_4}\left[\ln\left(a^{(1)}\right) - \frac{2\left(y_{i_1i_2i_3i_4} - l_{i_1i_2i_3i_4}\right)}{a^{(1)}}\right. \\ &\quad \left. + \frac{1}{2}\exp\left(\frac{2\left(y_{i_1i_2i_3i_4} - l_{i_1i_2i_3i_4}\right)}{a^{(1)}}\right)\right] + \frac{a}{2}\sum_{n=1}^{4}\left\|\left(\mathbf{U}^{(n)}\right)^{\top}\mathbf{U}^{(n)} - b^2\mathbf{I}\right\|_F^2. \end{aligned} \tag{6.14}$$

The derivative of F with respect to $\mathbf{U}^{(1)}$ follows the chain rule

$$\frac{\partial F}{\partial \mathbf{U}^{(1)}} = \frac{\partial F}{\partial \underline{\mathbf{L}}}\frac{\partial \underline{\mathbf{L}}}{\partial \mathbf{U}^{(1)}}. \tag{6.15}$$

We can write

$$\frac{\partial F}{\partial \underline{\mathbf{L}}} = \frac{2 - \exp\left(\frac{2(\underline{\mathbf{Y}} - \underline{\mathbf{L}})}{a^{(1)}}\right)}{a^{(1)}}, \tag{6.16}$$

where $\exp(\cdot)$ is the element-wise exponentiation operator.

We modified the derivatives $\frac{\partial \underline{\mathbf{L}}}{\partial \mathbf{U}^{(1)}}$ given in (6.11) as described in Appendix 6.8.2. Similar to (6.11), we can define the iteration parameter k and calculate

the derivative of F with respect to the factor matrices as

$$\begin{aligned}
\frac{\partial F}{\partial \mathbf{U}^{(1)^k}} &= \Big(\frac{\partial F}{\partial \underline{\mathbf{L}}^k} \times_2 \left(\mathbf{U}^{(2)^k}\right)^\top \times_3 \left(\mathbf{U}^{(3)^k}\right)^\top \times_4 \left(\mathbf{U}^{(4)^k}\right)^\top\Big)_{(1)} \left(\mathbf{S}_{(1)}{}^k\right)^\top \\
&\quad + a\,\mathbf{U}^{(1)^k}\left(\left(\mathbf{U}^{(1)^k}\right)^\top \mathbf{U}^{(1)^k} - b^2\,\mathbf{I}\right), \\
\frac{\partial F}{\partial \mathbf{U}^{(2)^k}} &= \Big(\frac{\partial F}{\partial \underline{\mathbf{L}}^k} \times_1 \left(\mathbf{U}^{(1)^k}\right)^\top \times_3 \left(\mathbf{U}^{(3)^k}\right)^\top \times_4 \left(\mathbf{U}^{(4)^k}\right)^\top\Big)_{(2)} \left(\mathbf{S}_{(2)}{}^k\right)^\top \\
&\quad + a\,\mathbf{U}^{(2)^k}\left(\left(\mathbf{U}^{(2)^k}\right)^\top \mathbf{U}^{(2)^k} - b^2\,\mathbf{I}\right), \\
\frac{\partial F}{\partial \mathbf{U}^{(3)^k}} &= \Big(\frac{\partial F}{\partial \underline{\mathbf{L}}^k} \times_1 \left(\mathbf{U}^{(1)^k}\right)^\top \times_2 \left(\mathbf{U}^{(2)^k}\right)^\top \times_4 \left(\mathbf{U}^{(4)^k}\right)^\top\Big)_{(3)} \left(\mathbf{S}_{(3)}{}^k\right)^\top \\
&\quad + a\,\mathbf{U}^{(3)^k}\left(\left(\mathbf{U}^{(3)^k}\right)^\top \mathbf{U}^{(3)^k} - b^2\,\mathbf{I}\right), \\
\frac{\partial F}{\partial \mathbf{U}^{(4)^k}} &= \Big(\frac{\partial F}{\partial \underline{\mathbf{L}}^k} \times_1 \left(\mathbf{U}^{(1)^k}\right)^\top \times_2 \left(\mathbf{U}^{(2)^k}\right)^\top \times_3 \left(\mathbf{U}^{(3)^k}\right)^\top\Big)_{(4)} \left(\mathbf{S}_{(4)}{}^k\right)^\top \\
&\quad + a\,\mathbf{U}^{(4)^k}\left(\left(\mathbf{U}^{(4)^k}\right)^\top \mathbf{U}^{(4)^k} - b^2\,\mathbf{I}\right).
\end{aligned} \tag{6.17}$$

Finally, the derivative with respect to $\underline{\mathbf{S}}^k$ has the same form given in (6.11), that is,

$$\frac{\partial F}{\partial \underline{\mathbf{S}}^k} = \frac{\partial F}{\partial \underline{\mathbf{L}}^k} \times_1 \left(\mathbf{U}^{(1)^k}\right)^\top \times_2 \left(\mathbf{U}^{(2)^k}\right)^\top \times_3 \left(\mathbf{U}^{(3)^k}\right)^\top \times_4 \left(\mathbf{U}^{(4)^k}\right)^\top. \tag{6.18}$$

In our proposed method, we start with a debiasing step. Fisher-Tippett noise has a non-zero mean. We subtract the theoretical mean $\mathbb{E}[\underline{\mathbf{M}}]$ from the tensor $\underline{\mathbf{Y}}$. Additionally, we subtract the gain of the system $a^{(2)}$, which is estimated as the median of the first twelve seconds. Note that the subtraction of the gain is not applied in the simulation. We initialize $\mathbf{U}^{(1)^0}$, $\mathbf{U}^{(2)^0}$, $\mathbf{U}^{(3)^0}$, $\mathbf{U}^{(4)^0}$ and $\underline{\mathbf{S}}^0$ with the tr-MLSVD estimate components given in Chapter 5. The spectral norm q is calculated as described in (6.9). Consequently, the core tensor and the factor matrices are scaled by $q^{1/4}$ and q, respectively. The derivative of the loss function F with respect to the tensor $\underline{\mathbf{L}}^k$ given in (6.16) is plugged into (6.17) and (6.18). The gradient descent is calculated using the steps defined in (6.10) until a maximum iteration number K or an early stopping criterion is reached. In this dissertation, we refer to this algorithm as GTE.

An early stopping condition is defined as the relative change of the $\underline{\mathbf{L}}^k$ at each iteration to the change of the first iteration, that is,

$$\Delta\underline{\mathbf{L}}^{k+1} < \epsilon = \frac{\|\underline{\mathbf{L}}^{k+1} - \underline{\mathbf{L}}^k\|_F}{\|\underline{\mathbf{L}}^1 - \underline{\mathbf{L}}^0\|_F}. \tag{6.19}$$

The iterations stop when $\Delta\underline{\mathbf{L}}^k$ is smaller than ϵ or the maximum number of iterations K is reached. We modify the step size η to $\eta/10$ whenever the algorithm diverges and start the algorithm again with $k = 0$. The divergence is captured

in an increasing relative loss (6.19). The GTE estimate is denoted with $\hat{\underline{\mathbf{L}}}$. The pseudo-code for the algorithm is shown in Algorithm 4.

Algorithm 4 The GTE framework for estimating a low-rank tensor with Fisher-Tippett distribution

Input : $\underline{\mathbf{Y}}$, $\mathbf{E}\{\underline{\mathbf{M}}\}, \eta, K, \epsilon, (R_1, \ldots, R_N)$
Debias : $\underline{\mathbf{Y}} \leftarrow \underline{\mathbf{Y}} - \mathbf{E}\{\underline{\mathbf{M}}\}$
Initialization : $\underline{\mathbf{L}} \leftarrow \underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \times_4 \mathbf{U}^{(4)}$ ▷ Using tr-MLSVD estimate in Chapter 5
$\mathbf{U}^{(n)^0} \leftarrow q^{1/4}\,\mathbf{U}^{(n)}$ for $n \in \{1 \ldots 4\}$
$\underline{\mathbf{S}}^0 \leftarrow \underline{\mathbf{S}}/q$ see (6.9)
$k \leftarrow 0$
while $k < K$ or $\Delta \underline{\mathbf{L}}^k > \epsilon$ **do**
  $\mathbf{U}^{(n)^{k+1}} \leftarrow \mathbf{U}^{(n)^k} - \eta \frac{\partial F}{\partial \mathbf{U}^{(n)^k}}$ for $n \in \{1 \ldots 4\}$ see (6.17).
  $\underline{\mathbf{S}}^{k+1} \leftarrow \underline{\mathbf{S}}^k - \eta \frac{\partial F}{\partial \underline{\mathbf{S}}^k}$ see (6.18)
  $\underline{\mathbf{L}}^{k+1} \leftarrow \underline{\mathbf{S}}^k \times_1 \mathbf{U}^{(1)^k} \times_2 \mathbf{U}^{(2)^k} \times_3 \mathbf{U}^{(3)^k} \times_4 \mathbf{U}^{(4)^k}$
  if $k > 1$ **then**
    $\Delta \underline{\mathbf{L}}^{k+1} \leftarrow \frac{\|\underline{\mathbf{L}}^{k+1} - \underline{\mathbf{L}}^k\|_F}{\|\underline{\mathbf{L}}^1 - \underline{\mathbf{L}}^0\|_F}$ see (6.19)
  end if
  if $\Delta \underline{\mathbf{L}}^{k+1} > \Delta \underline{\mathbf{L}}^k$ **and** $k > 1$ **then**
    $\eta \leftarrow \eta/10$
    $k \leftarrow 0$
  end if
  $k \leftarrow k + 1$
end while
Return : $\hat{\underline{\mathbf{L}}} = \underline{\mathbf{L}}^k$

6.2.3. ORIENTATION INVARIANT TENSOR NUCLEAR NORM

As an alternative to GTE, we next describe the OITNN algorithm for denoising DCEUS acquisitions. OITNN considers low-rankness in mode-$(n, n+1)$ unfoldings, and it is based on a framework called t-SVD [8] that is described in Chapter 2. We define an additional unfolding operation such that the TNN explained in (2.47) can be applied to the 4th-order DCEUS sequences, similar to Chapter 4.

Definition 6.2.1. Mode-$(n, n+1)$ unfolding: The mode-$(n, n+1)$ unfolding of $\underline{\mathbf{L}} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ creates a 3D tensor $\underline{\mathbf{L}}_{(n,n+1)} \in \mathbb{R}^{I_n \times D/(I_n I_{n+1}) \times I_{n+1}}$ by permuting the nth mode of $\underline{\mathbf{L}}$ to the first dimension, the $(n+1)$st mode to the last dimension, and grouping the remaining modes. Here, D is defined as $\prod_{n=1}^{4} I_n$. This unfolding is also denoted as $\underline{\mathbf{L}}_{[n]}$.

Definition 6.2.2. The OITNN is defined using the mode-$(n, n+1)$ operation as

$$\|\underline{\mathbf{L}}\|_{\text{OITNN}} = \frac{1}{4} \sum_{n=1}^{4} \|\underline{\mathbf{L}}_{[n]}\|_{\text{TNN}}. \tag{6.20}$$

We refer back to the original problem formulation given in (6.5). The authors in [92] approximate the Fisher-tippet noise $\underline{\mathbf{M}}$ as the summation of WGN $\underline{\mathbf{W}} \in \mathbb{R}^{I_1 \times \cdots \times I_4}$ and sparse outliers $\underline{\mathbf{O}} \in \mathbb{R}^{I_1 \times \cdots \times I_4}$, that is,

$$\underline{\mathbf{Y}} = \underline{\mathbf{L}} + \underline{\mathbf{W}} + \underline{\mathbf{O}} \,. \tag{6.21}$$

The OITNN considers the low-rankness in all orientations mode-$(n, n+1)$ for $n \in \{1, 2, 3, 4\}$ and solves the optimization problem that is defined as

$$\min_{\underline{\mathbf{L}}, \underline{\mathbf{O}}} \left\{ \frac{1}{2} \| \underline{\mathbf{L}} + \underline{\mathbf{O}} - \underline{\mathbf{Y}} \|_F^2 + \gamma_L \| \underline{\mathbf{L}} \|_{\text{OITNN}} + \gamma_O \| \underline{\mathbf{O}} \|_1 \right\}, \quad \text{s.t.} \quad \| \underline{\mathbf{L}} \|_\infty \leq \alpha \,. \tag{6.22}$$

Similar to GTE, the DCEUS acquisition is first preprocessed by subtracting the median of the first twelve seconds from the signal. The infinity norm α in (6.22) is assigned by median filtering the preprocessed $\underline{\mathbf{Y}}$ with a window of 5 seconds and taking the maximum. This algorithm is solved by using the algorithm in [30] until the stopping convergence given in (6.19) or the maximum iteration number K is reached. An overview of the methods is shown in Fig. 6.1.

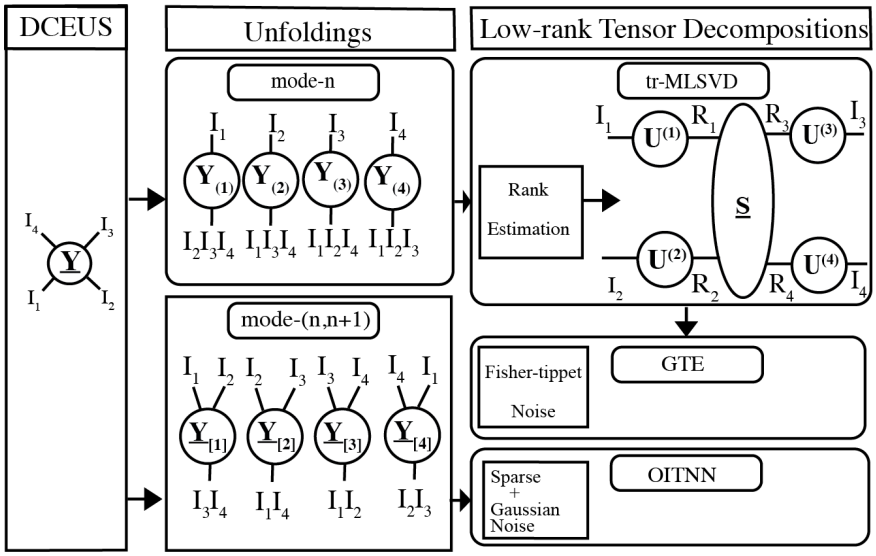


Figure 6.1: Overview of the tensor-based DCEUS denoising methods.

6.3. VALIDATION METHODOLOGY

In this section, we introduce the setup for simulation, in-vitro, and in-vivo studies, and report the results in Section 6.4. We first compare the denoising performance of tr-MLSVD, GTE, and OITNN using synthetic data. Previously, in Chapter 5, we compared the performance of SVD and tr-MLSVD methods through simulation,

where tr-MLSVD was found to perform better. For that reason, the SVD method is omitted from the simulation. Additionally, wavelet-based denoising [126] is omitted from the simulation because the assumption that the time evolutions are smooth does not hold for the general case. For this reason, we only compare the tensor-based denoising techniques. Following the simulations, we will conduct an in-vitro study, report the model-fitting performance, and visualize the phantom recordings after denoising. Finally, we will compare the single-feature classification performance of the tensor-based methods with the state-of-the-art DCEUS denoising techniques.

6.3.1. SIMULATION

We generated three 4th-order tensors $\underline{\mathbf{L}} \in \mathbb{R}^{20\times20\times20\times20}$ with three different ranks. First the core tensor was generated $\underline{\mathbf{S}} \in \mathbb{R}^{R_1\times R_2\times R_3\times R_4}$ from the normal distribution with sizes $[6,6,6,6]$, $[4,8,12,16]$ and $[12,12,12,12]$. The core matrix was multiplied in each mode with the orthonormal matrices to get $\underline{\tilde{\mathbf{L}}} = \underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \times_4 \mathbf{U}^{(4)}$. The orthonormal matrices $\mathbf{U}^{(1)} \in \mathbb{R}^{20\times R_1}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{20\times R_2}$, $\mathbf{U}^{(3)} \in \mathbb{R}^{20\times R_3}$ and $\mathbf{U}^{(4)} \in \mathbb{R}^{20\times R_4}$ were generated according to the Haar measure as described in [135]. We selected such ranks to cover three cases that might occur in actual DCEUS acquisitions. An actual DCEUS acquisition might have low or high multilinear ranks in all modes, or it can have a lower rank in some of the modes than others. Since we expect the recording to be low-rank, we selected a value close to half of the tensor size for high-rank cases.

The $\underline{\tilde{\mathbf{L}}}$ values were scaled to the range $[10,255]$. The noisy tensor $\underline{\tilde{\mathbf{Y}}}$ was obtained by multiplying $\underline{\tilde{\mathbf{L}}}$ with Rayleigh-shaped $\underline{\tilde{\mathbf{M}}}$ with scaling parameter 1 as shown in (6.12). The logarithmically transformed tensor $\underline{\mathbf{L}}$ was obtained using (6.4) with a dynamic range of 42. The denoising algorithm given in Algorithm 4 was applied to $\underline{\mathbf{L}}$. The theoretical noise statistics were taken from [136] and assigned as $\mathbb{E}[\underline{\mathbf{M}}] = 0.0579\alpha^{(1)}$, where $\alpha^{(1)}$ was given in (6.3). The preprocessing step was skipped. The ranks were assumed to be known. The step size was $\eta = 10^{-7}$. The total number of iterations was defined as $K = 10^5$. The iterations stopped with the condition $\Delta\underline{\mathbf{L}}^k < \epsilon = 0.1$. The true ranks were used for tr-MLSVD and GTE. For OITNN, four values for γ_L and γ_O in the range $(1000, 4000)$ and $(10, 40)$, respectively, have been traversed. Furthermore, the true infinity norm of the original tensor is used. The minimum normalized mean error is reported.

6.3.2. IN-VITRO

We conducted a phantom study using the LOGIQ E10 scanner equipped with a RIC5-9-D endocavity transducer driven at 3.5 MHz, a porous medium, and a sponge to prevent reverberations. The setup can be seen in Fig. 6.2. The porous media phantom was built by packing alginate beads of size 2.5 mm in a polyurethane tube with a diameter of 20 mm. The obtained alginate beads of the same size were packed into a polyurethane tube, and the cylinder shape was fixed with two circular nets on two sides of the phantoms. After that, we gently squeezed and shook the phantom to ensure a more homogeneous packing structure. The length of the phantoms was about 43 mm. The water pump is set to a flow rate of 0.22 mL/s. We mixed 1 mL of SonoVue with 100 mL of water and injected it into the tube before imaging. The model is fitted to each voxel

using (6.24). The MFR-RMSE is calculated by taking the root mean square of the difference between the fitted and filtered TICs and multiplying it with a sigmoid-shaped weighting function that exponentially penalizes the error starting from 20 seconds after the peak time.

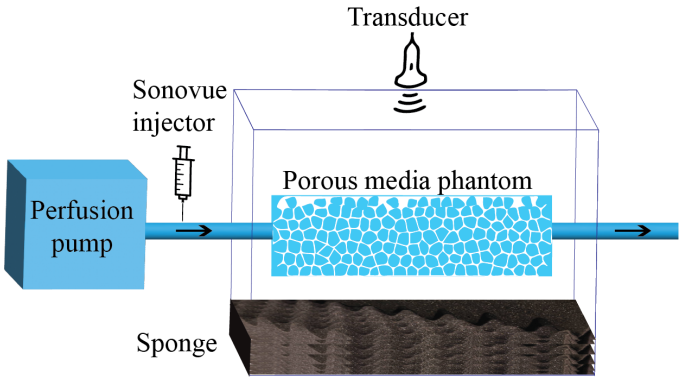


Figure 6.2: Phantom setup.

6.3.3. IN-VIVO

Recordings of 32 patients from the Amsterdam University Medical Center and 62 patients from the Netherlands Cancer Institute were obtained by Auke Jager [85]. This study obtained IRB approval, and the patients provided written consent to be enrolled in the study. A 4th-order recording in contrast mode was obtained with the LOGIQ E10 scanner equipped with an RIC5-9-D endocavity transducer driven at 3.5 MHz. The volume rate was fixed to 0.9 Hz by setting the image quality to BQMid1, and a low mechanical index of 0.1 was employed to minimize the bubble destruction. The patients went through radical prostatectomy after the recording because of biopsy-proven prostate cancer. The prostate was sliced with 4 mm thickness, and for each slice, an annotation was made by the pathologist. The annotations were registered back to the domain of the recording, and the ground truth was obtained. Significant malignant voxels, at least with a grade of $3+4$ are selected [137]. There were approximately seven million benign voxels and two hundred thousand malignant voxels.

The DCEUS acquisition was transformed from spherical to Cartesian domain with a voxel size of 0.25 mm × 0.25 mm × 0.25 mm. The spatial resolution was regularized across space through a dedicated Wiener filter [57], and the data was downsampled by 3 such that a voxel size of 0.75 mm was obtained. The warm initialization is obtained using tr-MLSVD. The gain is estimated as the median of the first 12 seconds, which is subsequently subtracted from the TICs of both the tr-MLSVD and the noisy tensor. A step size of $\eta = 10^{-7}$ was used. The maximum number of iterations was set to $K = 10^5$. The stopping condition was set at $\Delta \underline{\mathbf{L}}^k < \epsilon = 0.1$.

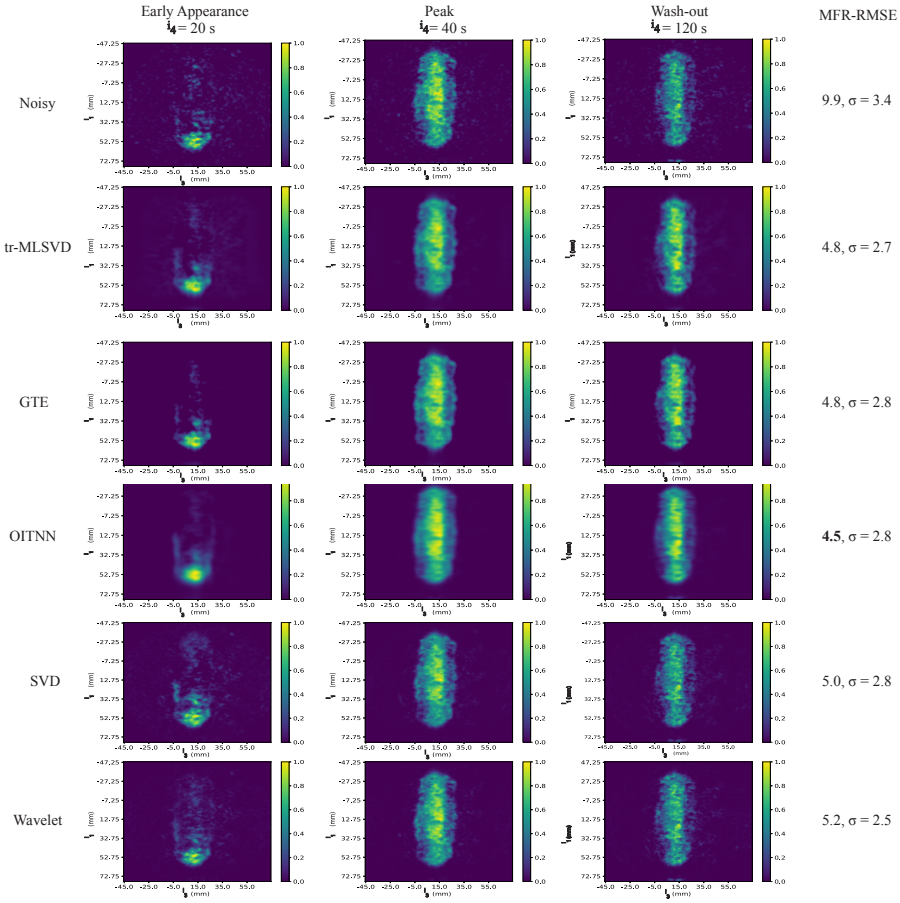


Figure 6.3: The normalized amplitudes of noisy and denoised phantom recordings during the early appearance, peak, and wash-out times. The (i_1, i_3) slice at the middle of the phantom is plotted.

The features given in Chapter 3 were extracted from each voxel after denoising with either of the five denoising methods, i.e., OITNN, GTE, tr-MLSVD, SVD, and wavelet-based denoising and without denoising. The extracted features were provided in Table 3.3. Among all the features, computing the model fitting took the most time. Therefore, the model fitting was approximated using the Exponential Linear Unit (ELU) function, which resulted in faster processing. The approximation is shown in Appendix 6.8.1. Similarly to the process proposed in the literature [55], a windowing was applied where a higher weight was given to

the first pass of the microbubbles. Adam optimizer [138] was used for fitting the model.

6.4. RESULTS

6.4.1. SIMULATION

The performance metric was the normalized mean error (NME) given by

$$\text{NME} = \frac{\|\hat{\underline{\mathbf{L}}} - \underline{\mathbf{L}}\|_F}{\|\underline{\mathbf{L}}\|_F}, \tag{6.23}$$

where $\hat{\underline{\mathbf{L}}}$ was the estimate of $\underline{\mathbf{L}}$. For both GTE and OITNN, the convergence is achieved with $\Delta\underline{\mathbf{L}}^k < \epsilon = 0.1$ defined in (6.19). We compared the performance of tr-MLSVD, GTE, and OITNN and reported the results in Fig. 6.4. We ran pairwise t-tests to compare the denoising performances. Only tr-MLSVD and OITNN for the rank (6,6,6,6) showed insignificant differences with $p < 0.0001$. For rank (6,6,6,6), the GTE gave the best NME. For ranks (4,8,12,16) and (12,12,12,12), OITNN performed better.

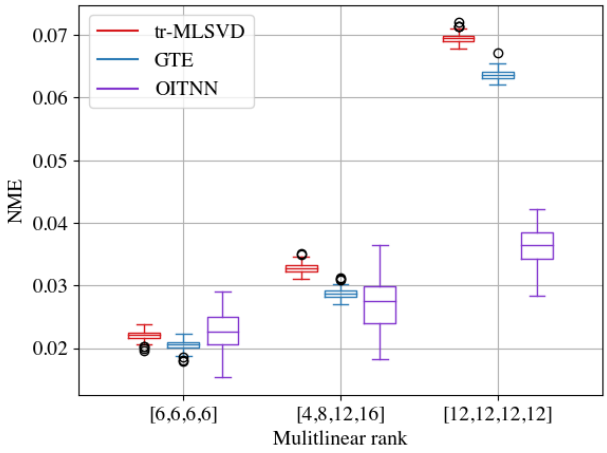


Figure 6.4: The normalized mean error for random tensors of size $[20, 20, 20, 20]$ and multilinear rank $[6, 6, 6, 6]$, $[4, 8, 12, 16]$ and $[12, 12, 12, 12]$. The median NME over 10^5 Monte Carlo simulations are shown with a flat line, the box ranges represent the 25th and 75th quantiles, and the whiskers represent the inter-quartile ranges.

6.4.2. IN-VITRO

We applied the tensor-based speckle denoising algorithms to the phantom recordings. The rank estimation parameter for tr-MLSVD is selected to be $\rho = 10^{-5}$. The sparsity and low-rankness related parameters γ_L and γ_O are selected after a 100 values between $(1, 20000)$ and $(1, 100)$, respectively, are swept. The value that gave the least MFR-RMSE is found to be 10000 and 30 for γ_L and γ_O, respectively. The GTE algorithm is run with $\eta = 10^{-7}$. In Fig. 6.3, the early appearance, peak, and wash-out times of the phantom recordings are shown on an (i_1, i_3) slice at the middle of the phantom, and the averaged MFR-RMSE values are reported. Only the voxels inside the phantom are fitted with the model. The OITNN method resulted in the best fit of the model described in Appendix 6.8.1. The tr-MLSVD and GTE performed similarly, while a better suppression of the speckle artifacts in the background surrounding the phantom was observed for GTE and OITNN. Finally, SVD, Wavelet methods showed worse model-fitting performance than the tensor-based counterparts. We ran pairwise t-tests, and only tr-MLSVD and GTE gave statistically insignificant MFR-RMSE distributions inside the phantom with $p < 0.0001$.

6.4.3. IN-VIVO

We used the same parameters as the in-vitro for the tensor-based denoising methods (MLSVD, GTE, OITNN), i.e. the rank selection parameter $\rho = 10^{-5}$ for tr-MLSVD, and the $\gamma_O = 30$, $\gamma_L = 10^4$ for the OITNN. The features extracted after tensor-based denoising methods were compared against features extracted after matrix-based (SVD) denoising and a state-of-the-art wavelet-based denoising, as well as no denoising. The denoising results are illustrated for an arbitrarily selected malignant and benign voxel in Fig. 6.6 (a) and (c). In the same figure, the results of the modified linear random walk model fitting described in Appendix 6.8.1 are shown in the subplots (b) and (d). Furthermore, in Fig. 6.7, we illustrate the model fit parameter MFR-κ on a slice with the highest number of malignant voxels of an arbitrarily selected patient.

The mean and the standard deviation of all features are given in Table 6.2. The mean and the standard deviation of the model fit-related features did not change after the various denoising algorithms. To investigate further, we also compared the goodness of the model fit in terms of the weighted root mean square error of the fit (MFR-RMSE), which is explained in Section 6.3.2. The mean and standard deviation of MFR-RMSE are shown in the third column of Table 6.2. MFR-RMSE is found to be the lowest for the OITNN method, with an average of 3.84 over all the malignant and benign voxels. The OITNN method is followed by GTE, SVD, MLSVD, and Wavelet methods with average RMSE of 4.11, 4.14, 4.80, and 5.12, respectively. The noisy RMSE fit has an average RMSE of 7.81. Although the RMSE changed, the model fit parameters stayed fairly close to each other, only with an increase in the variance of the benign features in GTE and OITNN. The similarity metrics SA-MI, SA-ρ, and SA-r increased significantly when a low-rank decomposition is applied. We ran pairwise t-tests to analyze the statistics of the features before and after filtering. The t-test is a hypothesis testing method that determines whether there is a significant difference between the means of two groups. The t-test assumes that the data are approximately normally distributed and that the samples are independent. The result of a t-test is reported along

with a p-value. A low p-value indicates that the observed difference is unlikely to have occurred by chance. A high p-value does not necessarily indicate that the group means are similar or different. In this chapter, statistical significance was determined at a threshold of $p < 0.0001$, indicating strong evidence that the means of the groups are different. We call t-tests with $p < 0.0001$ statistically significant results. This means that the features have different means after filtering that cannot occur by chance.

We ran t-tests to assess the two categories: the effect of filtering methods on the distribution of the features, and the difference between malignant and benign features. For the first one, we stacked the benign and malignant features and ran a pair-wise t-test to compare the filtering methods. All the filtering methods had statistically significant differences with $p < 0.0001$ except for two cases: CD-V and SA-MI features for GTE and SVD and MFR-κ feature for GTE and OITNN. This is shown in Table 6.1. Important to note that the features that *lack* statistical significance are shown in Table 6.1. For the second assessment, we ran Welch's t-test between the malignant and benign features generated through each filtering method, including the noisy one. Welch's t-test was chosen due to the difference in the number of malignant and benign voxels. In all cases, we can differentiate the benign and malignant voxels with statistical significance ($p < 0.0001$).

Table 6.1: The statistical significance test results with (p<0.0001) between the features extracted from the noisy and filtered DCEUS acquisitions. Only the ones that *lack* statistical significance are shown.

	Noisy	tr-MLSVD	GTE	OITNN	SVD	Wavelet
Noisy		-	-	-	-	-
tr-MLSVD	-		-	-	-	-
GTE	-	-		-	CD-V, SA-MI	-
OITNN	-	-	MFR-κ		-	-
SVD	-	-	-	-		-
Wavelet	-	-	-	-	-	

To quantify the discriminative power of the features, the malignant and benign features from all the patients were stacked, and simple single-feature classification was done by varying a threshold over the entire range of each feature. The values that were greater or lower than the threshold were labeled as malignant or benign, and the receiver operating characteristic (ROC) curve was generated. The area under the ROC curve (AUC) for the various methods and features is shown in Table 6.3. Five out of nine features (MFR-μ, CD-v, SA-MI, VE-CEv, VE-EV) showed better separation between malignant and benign voxels for tr-MLSVD. Three out of the nine features (MFR-κ, CD-D, SA-ρ) had a better classification result for OITNN. Only one feature (SA-r) showed improved performance for SVD.

6.4.4. MEMORY AND COMPUTATIONAL REQUIREMENTS

Low-rank decomposition methods enable the representation of any data with fewer parameters. For tr-MLSVD and GTE, compression is achieved by saving the

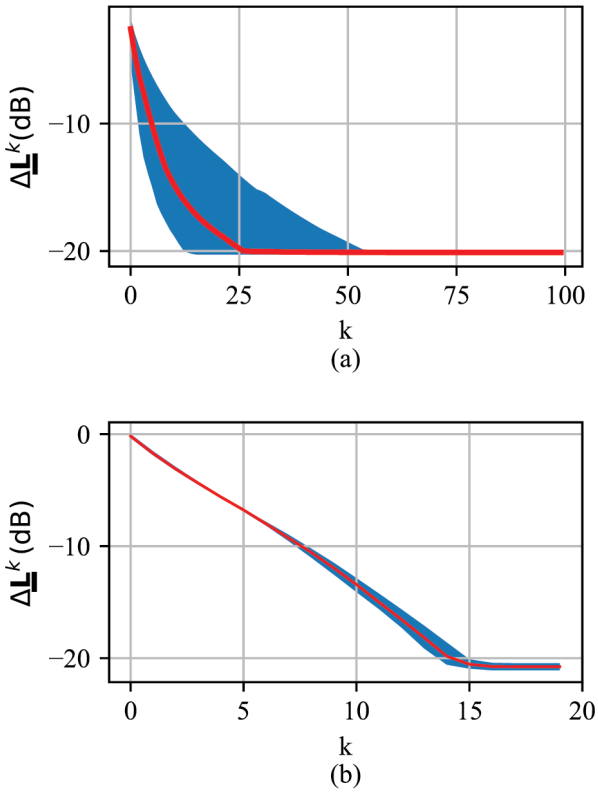


Figure 6.5: The median, 25th and 75th percentiles of loss $\Delta\underline{\mathbf{L}}^k$ that is defined in (6.19) calculated for all the patients when the GTE and OITNN algorithm given at Algorithm 4 is run. Subplot (a) represents the GTE, and subplot (b) represents the OITNN. The red line represents the median, and the blue shade represents the percentiles, and the y-axis is shown in dB.

factor matrices $\hat{\mathbf{U}}^{(n)}$ for $n \in \{1, 2, 3, 4\}$ and the core tensor $\hat{\underline{\mathbf{S}}}$; for SVD, the left and right singular vectors, and the singular values; for OITNN t-compress algorithm from [8]. The t-compress algorithm is applied to mode-(1,2) unfolding. The other possible mode-$(n, n+1)$ unfoldings resulted in similar results. Significantly small singular values and their corresponding factors are discarded, where the tolerance is set to 1/100th of the highest singular value. The highest compression is achieved for the tr-MLSVD and GTE algorithms. This is followed by SVD and OITNN. Calculation of the features and the denoising is done on a server with an Intel 2 x 10 Xeon CPU and 256 GB RAM. SVD was the fastest algorithm, with 7 hours of computation time. The tr-MLSVD, Wavelet, GTE, and OITNN followed

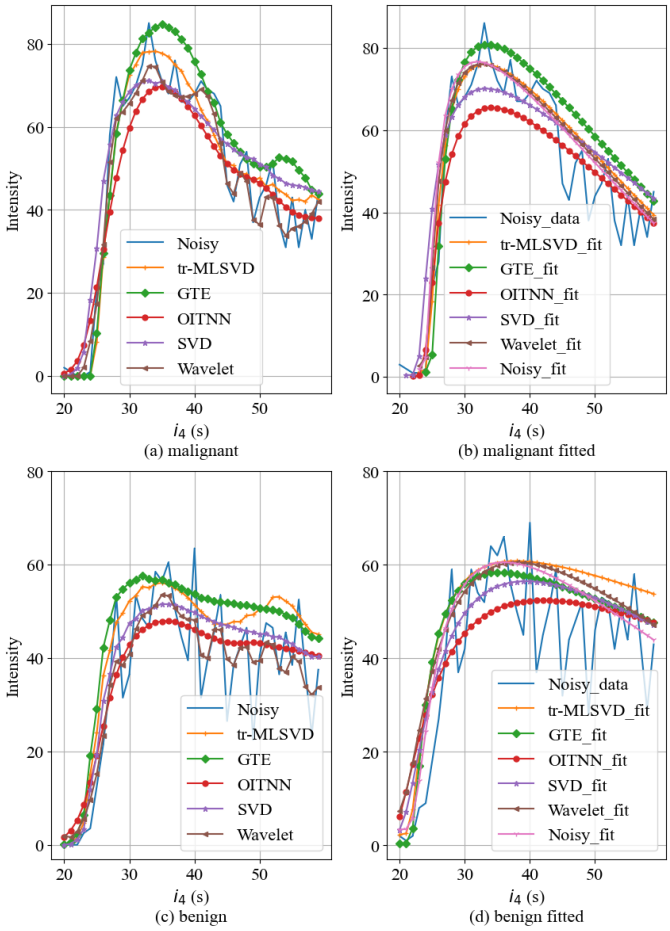


Figure 6.6: The denoised TICs $\mathbf{I}_{i_1 i_2 i_3}(i_4)$ for a malignant and benign voxel and the corresponding fitted models. The noisy data $\mathbf{y}_{i_1 i_2 i_3}(i_4)$ follows the signal model in (6.5). The upper row, defined with labels (a) and (b), represents the malignant voxels. The bottom row, defined with the labels (c) and (d), represents the benign voxels. Only the wash-in period is shown.

it with computation times of 8.3, 9.5, 17, and 42 hours, respectively. The full recordings of 94 patients were 68 GB. After the compression, the reduction is shown in Table 6.4.

In addition, the computational complexities of the algorithms are presented in Table 6.4. All algorithms except wavelet include SVD as their foundation. Here, we assume the Golub-Reinsch algorithm is used to calculate the SVD [139]. We point out that with known ranks, iterative algorithms such as [140] will have a reduced complexity. We consider the worst-case scenario, where the ranks are

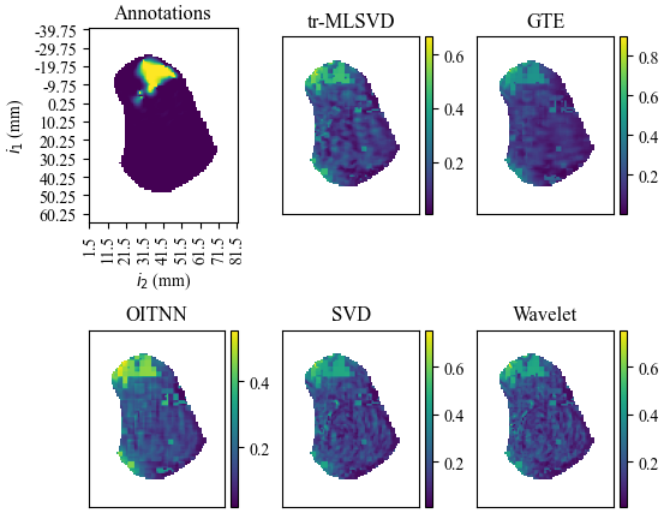


Figure 6.7: The κ values for a i_3 slice that has the most malignant voxels for various denoising schemes of a random patient.

as high as the sizes of the corresponding mode. In [139], the complexity of SVD for an $I \times J$ matrix is described as $O(I^2J)$ given $I < J$. Therefore, the SVD of the spatiotemporal matrix with size $I_4 \times I_1I_2I_3$ has the complexity $O(I_1I_2I_3I_4^2) = O(DI_4)$. For tr-MLSVD, the SVD of the tensor $\underline{\mathbf{Y}}$ in the unfoldings dominates the complexity and results in a complexity of $O(D(\sum_{n=1}^{4} I_n))$. For GTE, the algorithm is initialized by the tr-MLSVD output. The complexity per iteration is due to the tensor-matrix products that generate the low-rank tensor $\underline{\mathbf{L}}^k$ given in Algorithm 4. Each mode-n tensor-matrix multiplication is a matrix multiplication of sizes $D/I_n \times I_n$, and $I_n \times I_n$, which results in a complexity of $D\max\{I_1, \ldots, I_4\}$. For OITNN, the Fourier transform, inverse Fourier transform, and the SVDs of four unfoldings $\underline{\mathbf{Y}}_{[n]}$ dominate the complexity of each iteration. The Fourier transform of unfoldings $\underline{\mathbf{Y}}_{[n]}$ for $n \in \{1, \ldots, 4\}$ results in the complexity of $O(D\log(D))$, and their SVD results in the complexity of $O(D(\sum_{n=1}^{4} I_n))$. Finally, the wavelet decomposition has a complexity of $O(D\log(I_4))$ due to the 1D Fourier transform

Table 6.2: The mean and the standard deviation of the DCEUS features after the signal is denoised by Wavelet, SVD, MLSVD, GTE algorithms.

		MFR-κ	MFR-μ	MFR-RMSE	CD-D	CD-v	SA-MI	SA-r	SA-ρ	VE-CEv	VE-Ev
Noisy	Malignant	0.15, $\sigma = 0.06$	25.99, $\sigma = 15.51$	9.11, $\sigma = 2.58$	0.05, $\sigma = 0.05$	0.28, $\sigma = 0.20$	1.19, $\sigma = 0.18$	0.45, $\sigma = 0.13$	0.76, $\sigma = 0.14$	2.00, $\sigma = 0.57$	2.92, $\sigma = 0.92$
	Benign	0.13, $\sigma = 0.09$	42.13, $\sigma = 18.83$	7.78, $\sigma = 2.87$	0.01, $\sigma = 0.02$	0.09, $\sigma = 0.11$	0.83, $\sigma = 0.48$	0.31, $\sigma = 0.19$	0.51, $\sigma = 0.30$	1.32, $\sigma = 0.39$	1.85, $\sigma = 0.60$
tr-MLSVD	Malignant	0.15, $\sigma = 0.06$	28.66, $\sigma = 15.42$	5.17, $\sigma = 2.88$	0.07, $\sigma = 0.08$	0.34, $\sigma = 0.26$	1.56, $\sigma = 0.38$	0.80, $\sigma = 0.19$	0.96, $\sigma = 0.06$	2.24, $\sigma = 0.67$	3.27, $\sigma = 1.06$
	Benign	0.13, $\sigma = 0.09$	44.83, $\sigma = 18.77$	4.79, $\sigma = 2.88$	0.02, $\sigma = 0.03$	0.10, $\sigma = 0.13$	0.92, $\sigma = 0.56$	0.58, $\sigma = 0.28$	0.85, $\sigma = 0.18$	1.37, $\sigma = 0.51$	1.93, $\sigma = 0.75$
GTE	Malignant	0.15, $\sigma = 0.07$	29.08, $\sigma = 15.94$	5.00, $\sigma = 4.77$	0.08, $\sigma = 0.10$	0.33, $\sigma = 0.26$	1.69, $\sigma = 0.40$	0.87, $\sigma = 0.14$	0.98, $\sigma = 0.06$	2.22, $\sigma = 0.68$	3.23, $\sigma = 1.07$
	Benign	0.13, $\sigma = 0.32$	45.65, $\sigma = 19.93$	4.09, $\sigma = 2.91$	0.02, $\sigma = 0.04$	0.10, $\sigma = 0.14$	0.99, $\sigma = 0.67$	0.62, $\sigma = 0.35$	0.75, $\sigma = 0.37$	1.38, $\sigma = 0.56$	1.96, $\sigma = 0.83$
OITNN	Malignant	0.15, $\sigma = 0.07$	33.83, $\sigma = 16.02$	**4.33**, $\sigma = 2.69$	0.08, $\sigma = 0.09$	0.35, $\sigma = 0.29$	1.92, $\sigma = 0.40$	0.91, $\sigma = 0.07$	0.99, $\sigma = 0.02$	2.07, $\sigma = 0.76$	3.08, $\sigma = 1.18$
	Benign	0.13, $\sigma = 0.35$	48.78, $\sigma = 18.86$	**3.83**, $\sigma = 2.67$	0.02, $\sigma = 0.04$	0.10, $\sigma = 0.14$	1.21, $\sigma = 0.72$	0.73, $\sigma = 0.30$	0.84, $\sigma = 0.27$	1.25, $\sigma = 0.54$	1.80, $\sigma = 0.82$
SVD	Malignant	0.16, $\sigma = 0.07$	30.05, $\sigma = 16.03$	4.55, $\sigma = 2.99$	0.07, $\sigma = 0.08$	0.33, $\sigma = 0.25$	1.65, $\sigma = 0.35$	0.86, $\sigma = 0.14$	0.98, $\sigma = 0.03$	2.25, $\sigma = 0.65$	3.26, $\sigma = 1.04$
	Benign	0.13, $\sigma = 0.09$	45.69, $\sigma = 18.67$	4.13, $\sigma = 3.00$	0.02, $\sigma = 0.03$	0.10, $\sigma = 0.14$	0.99, $\sigma = 0.62$	0.62, $\sigma = 0.32$	0.88, $\sigma = 0.20$	1.40, $\sigma = 0.53$	1.96, $\sigma = 0.77$
Wavelet	Malignant	0.15, $\sigma = 0.07$	31.32, $\sigma = 16.45$	5.77, $\sigma = 2.71$	0.05, $\sigma = 0.06$	0.28, $\sigma = 0.21$	1.33, $\sigma = 0.25$	0.64, $\sigma = 0.18$	0.93, $\sigma = 0.06$	2.04, $\sigma = 0.61$	2.96, $\sigma = 0.97$
	Benign	0.13, $\sigma = 0.09$	46.83, $\sigma = 18.63$	5.10, $\sigma = 2.77$	0.02, $\sigma = 0.03$	0.09, $\sigma = 0.11$	0.89, $\sigma = 0.55$	0.50, $\sigma = 0.29$	0.77, $\sigma = 0.24$	1.32, $\sigma = 0.43$	1.84, $\sigma = 0.65$

Table 6.3: The voxel-based ROC-AUC metric is calculated for each denoising method when a single feature classifier is used across 94 patients. The threshold is selected such that the highest AUC is achieved.

	Noisy	tr-MLSVD	GTE	OITNN	SVD	Wavelet
MFR-κ	0.587	0.638	0.628	**0.649**	0.628	0.617
MFR-μ	0.761	**0.767**	0.765	0.749	0.761	0.752
CD-D	0.771	0.788	0.781	**0.793**	0.780	0.777
CD-v	0.832	**0.837**	0.819	0.826	0.828	0.829
SA-MI	0.728	**0.827**	0.803	0.798	0.824	0.744
SA-r	0.717	0.756	0.765	0.721	**0.779**	0.626
SA-ρ	0.774	0.801	0.783	**0.805**	0.782	0.752
VE-CEv	0.828	**0.840**	0.822	0.797	0.834	0.825
VE-Ev	0.831	**0.842**	0.822	0.803	0.836	0.827

of D/I_4 voxels with I_4 samples.

Avoiding the Kronecker products, which are described in Section 6.2.2, made it possible for the GTE method to be run with less random access memory (RAM). Assume the calculation of the derivative with respect to the first-factor matrix $\frac{\partial F}{\partial \mathbf{U}^{(1)}}$. Calculation of the Kronecker product given in (6.28) creates $(\mathbf{U}^{(4)} \otimes \mathbf{U}^{(3)} \otimes \mathbf{U}^{(2)}) \in \mathbb{R}^{I_2 I_3 I_4 \times R_2 R_3 R_4}$. For DCEUS acquisitions of size $100 \times 80 \times 100 \times 120$, and ranks of $30 \times 20 \times 30 \times 8$, the resulting matrix is of size 960000×4800, creating 34.3 GB of data. On the other hand, consecutive mode-n multiplication in each mode, that is, $\frac{\partial F}{\partial \underline{\mathbf{L}}} \times_2 (\mathbf{U}^{(2)})^T \times_3 (\mathbf{U}^{(3)})^T \times_4 (\mathbf{U}^{(4)})^T$ reduces the size to $100 \times 20 \times 30 \times 8$. The mode-1 unfolding of the resulting tensor is of size 100×4800, creating 0.00384 GB of data.

Table 6.4: The computational time of running the denoising methods and the memory required to save 94 patient recordings, along with theoretical computational complexities. The total total iterations of GTE and OITNN are set to K.

	Computation Time	Memory	Computational Complexity
tr-MLSVD	8.3 h	840 MB	$\mathrm{O}\left(D(\sum_{n=1}^{4} I_n)\right)$
GTE	17 h	840 MB	$\mathrm{O}\left(KD\max\{I_1, \ldots, I_4\}\right)$
OITNN	42 h	24.94 GB	$\mathrm{O}\left(K(D\log(D) + D(\sum_{n=1}^{4} I_n))\right)$
SVD	7 h	2.87 GB	$\mathrm{O}\left(DI_4\right)$
Wavelet	9.5 h	68 GB	$\mathrm{O}\left(D\log(I_4)\right)$

6.5. DISCUSSION

We introduced two tensor-based algorithms for denoising DCEUS data that incorporate two approaches to model the Fisher-tippet speckle noise. The GTE algorithm models low-rankness in the mode-n unfoldings and uses the log-likelihood to reduce noise. On the other hand, the OITNN algorithm considers low-rankness in the mode-$(n, n+1)$ unfoldings and models the speckle noise as WGN and sparse outliers. In this section, we discuss the results and propose research directions for the future.

Inspired by the success of the application of SVD [93] for denoising DCEUS data, we extended the idea of low-rank approximation for denoising DCEUS data to high orders. Previously, we introduced the tr-MLSVD algorithm in [1] [2] and showed an improved performance of prostate cancer classification. The tr-MLSVD algorithm is more suitable for denoising WGN but not for speckle noise. This introduces artifacts that can be seen in the phantom slice depicted at the early appearance time of tr-MLSVD in Fig. 6.3. The speckles around the phantom at tr-MLSVD and the noisy recording do not exist in GTE and OITNN. We believe this can aid the visualization of bubble movement. Furthermore, in the simulation given in Fig. 6.4, we observed that OITNN and GTE perform better than tr-MLSVD for denoising speckle noise in nearly all cases. This justifies the benefit of incorporating the prior distribution and the assumption regarding the orthogonality of the factor matrices.

On the other hand, improved performance was not observed for the in-vivo classification. We made several assumptions that might affect the performance. Noise is assumed to be independent between voxels. We deconvolved the recording with a Wiener filter such that a resolution of 0.8 mm is obtained as proposed in [65]. We sampled the recording with a voxel size of 0.75 mm. Since each voxel is comparable to the resolution of the system, we assumed that the assumption of independence holds. There can be other factors, such as the movement of the probe, that can violate the independence. Without independence, the probability density function of Fisher-Tippett noise given in (6.12) cannot be multiplied across different elements of $\underline{\mathbf{M}}$. This implies that the statistical model becomes invalid, potentially detecting spurious structures and increasing the estimation error.

A reason for the inferior performance of the GTE algorithm could be the mismatch of the assumed noise characteristics. The DCEUS acquisitions suffer from various noise types such as clutter, shadowing, ring-comet, reflection, refraction, and reverberation artifacts [141]. The GTE algorithm aimed to remove the speckle noise with the assumption of the Fisher-Tippett noise. At lower bubble concentrations, this assumption is known to be invalid, as reported in [58]. Furthermore, distributions such as Gamma [126] or Nakagami [142] have been reported to be a better choice for describing the speckle statistics. Incorporating such a model could improve the results. On the other hand, OITNN assumes that the Fisher-Tippett noise can be modeled as WGN with sparse outliers. The aforementioned artifacts might violate this assumption. However, the two regularization parameters γ_L and γ_O allow a direction of future work where each parameter can be tweaked for the removal of different artifacts. The reduced performance of the wavelet-based denoising could be due to the reduced temporal frequency. The results in the literature have been reported for 25 frames/s, whereas the 4D DCEUS acquisitions provide 1 frame/s. The assumption that the

noise and the signal have separate subspaces in the wavelet domain might fail for such temporal frequencies.

The new model fitting algorithm described in Appendix 6.8.1 is found to be fairly robust to noise. As supported by Fig. 6.6 and Table 6.2, the model fitting algorithm gave similar features for all the denoised signals. This is further supported by the close performance of the model fit features in Table 6.3. When the noisy data is used to extract the model fit features, the performance of the voxel-based classifier is found to be similar to the features extracted from the denoised signals. In the literature, various fitting algorithms were compared for model fitting, such as the maximum likelihood and non-linear least squares [58]. In this chapter, we used an ADAM optimizer for faster processing with the relaxation proposed in Appendix 6.8.1. This could explain the closeness between the noisy and the denoised model fitting features. A significant increase in the similarity metrics (SA-ρ, SA-r, SA-MI) can be observed in Table 6.2 after the low-rank decompositions. This is expected since the low-rank decompositions explain the data using the components that describe the majority of the variance. Using a few components results in mostly similar TICs. This is especially observed in malignant regions. Finally, only temporal correlation SA-r gave better classification performance compared to tensor approaches. In SVD, low-rankness is considered only in the fourth unfolding. In such a spatiotemporal matrix, temporal information gets smoothened out, resulting in similar TIC appearance times across the whole DCEUS acquisition. Hence, the temporal correlation SA-ρ is performing worse for SVD, and the spectral correlation SA-r, which only considers the frequency content, performs better.

In our study, we have used the same LOGIQ E10 and the probe in both Amsterdam University Medical Center and the Netherlands Cancer Institute. All the machine settings were fixed, and the outcome of the single feature classification was consistent across both datasets. We have not included a study regarding the generalizability of the proposed algorithms in diverse clinical settings. The generalizability of the proposed tensor-based techniques is dependent on several parameters, such as the variability in equipment settings, patient demographics, and disease presentations. We selected the low-rank-related γ_L and the sparsity-related γ_O hyperparameters defined in Section 6.2.3, as well as the rank-related parameter ρ defined in Section 5 using phantom studies. We recommend a study that explores variability between different machines, which can shed light on good parameter settings for all future users. An analysis of model-fitting performance, such as the one given in Fig. 6.3, can be conducted. If a phantom study is not available, we recommend an analysis of model fitting or the classification performance on a subset of the patient data. Equipment settings can be incorporated into the classifier so that the classifier can utilize the variability between different devices and settings.

The preprocessing steps, such as subtracting the median of the first few seconds and the varying step size in optimization algorithms, are still applicable to all DCEUS equipment. This is due to the similarity of protocols regarding DCEUS imaging [143]. The gain is increased such that the background noise is observed without microbubbles, and the intravenous injection is done after starting the imaging sequence. Therefore, the first few frames are expected to represent the gain of the system, which can be subsequently subtracted from the TICs. The subtraction of the gain helps the algorithm focus on learning relative changes

related to microbubbles rather than an arbitrary baseline. The algorithm becomes more stable, and faster convergence is achieved. Similarly, the varying step size η given in Algorithm 4 increases the convergence speed and improves the robustness. If the tensor unfoldings are well-conditioned, the algorithms will converge faster with a high step size. On the contrary, if the tensor unfoldings are ill-conditioned, a smaller step size is required to aid the convergence. In this chapter, convergence is achieved for all the patient recordings. However, GTE is a non-convex algorithm, and the convergence is dependent on several factors, such as the amplitude of the noise, the condition number of the tensor unfoldings, and the regularity condition [132]. The OITNN is a convex algorithm with theoretical convergence guarantees [30]. We selected the stopping condition of the iterative algorithms as $\epsilon = 0.1$ given in (6.19). A lower stopping condition will improve the denoising performance while increasing the convergence time.

The patient demographics can affect the performance of the classification. Commonly, the maximum imaging depth is selected on the ultrasound device to cover all possible sizes of prostates. The patient age in our study ranged from 60 to 87. An interesting research direction can be to investigate the effect of age on the extracted DCEUS features. Additionally, the disease presentation might affect the low-rank assumption. Several factors signify prostate cancer's significance: the tumor size, Gleason score (or grade), and extracapsular extension [144]. Tumors require an increased supply of oxygen and nutrients beyond $1-2$ mm in diameter [116]. Clinically significant prostate cancer are tumors with at least size 0.5 cm^3 [144] and a grade of $3+4$ or higher. Subsequently, we selected tumor samples with at least a grade of $3+4$ Gleason score. We considered regions that are 2 mm in diameter, accounting for the system's spatial resolution as described in Section 6.3.3, since the main focus of this chapter was voxel-based classification. We propose a future study to analyze the effect of different Gleason scores and tumor sizes on the classification results. We expect the large regions to be identified more easily using low-rank tensor decomposition methods compared to early-stage tumors. This is due to their relative contribution to explaining the full DCEUS tensor. Regions that are small and have different TICs compared to the majority of the regions are expected to be captured in the smaller singular values. The signal subspace is assumed to be in the highest singular values in all the tensor-based denoising algorithms. Therefore, regions with low spatial structure and low temporal power will be ignored. Experimental validation of this hypothesis was not conducted, and we leave the comparison between the early-stage and significant prostate cancer classification as future work. Such a research direction requires annotations of malignant regions with different Gleason scores, the consideration of tumors with sizes varying between $2-25$ mm in diameter, and the inclusion of cases with extracapsular extension of the tumor. A possible research direction is to apply the aforementioned denoising techniques to subsets of the data rather than the full tensor so that the lower-grade or smaller tumor can be identified. Spatially, the input tensor can be divided into blocks, and the denoising can be applied to their time evolution. An analysis of the low-rank and sparsity-related hyperparameters, the selection of the ranks, and the effect of the block size on the classification results are recommended.

We calculated the DCEUS features by transforming the low-rank tensor back to the original size. Instead, the low-rank format could be kept for calculating the

features without forming the tensor. This will further relax the RAM requirements and speed up the feature calculation time. In addition, low-rank decomposition allows the estimation of possibly missing or corrupted temporal frames [145].

6.6. CONCLUSION

In this chapter, we investigate low-rank tensor decomposition-based denoising of dynamic contrast-enhanced ultrasound data. Besides the use of MLSVD, we introduced a low-rank denoising algorithm suitable for Rayleigh-shaped multiplicative noise based on a gradient descent algorithm and a low-rank denoising algorithm based on the OITNN framework. The proposed algorithms perform better than the truncated MLSVD in the simulation. In the in-vivo recordings, the same improvement was not observed for distinguishing benign and malignant voxels. However, low-rank tensor-based denoising using MLSVD outperformed other state-of-the-art approaches. In addition, the tr-MLSVD resulted in the best compression of the DCEUS acquisitions with a factor of 80. Although the addition of the noise distribution aids the visualization of DCEUS acquisitions, we did not find any improvement in the classification of prostate cancer. Considering these aspects and the added processing time of the GTE and OITNN methods, we believe approximating the noise as WGN, i.e., the use of tr-MLSVD is the best approach for denoising and compressing DCEUS acquisitions.

ACKNOWLEDGEMENTS

This project is funded in part by Holland High Tech with a PPS supplement for research and development in the Topsector HTSM. We would like to acknowledge Prof. Pintong Huang for carrying out the clinical trials at the Second Affiliated Hospital of Zhejiang University (Hangzhou, Zhejiang, China). Furthermore, we would like to thank Auke Jager for conducting clinical trials at the Amsterdam Medical University, Netherlands Cancer Institute in Amsterdam, Netherlands.

6.7. CHAPTER SUMMARY

In this chapter, we investigated the convex optimization OITNN-O, which was found to be suitable for low SNR scenarios in Chapter 4. Furthermore, a non-convex tensor-based denoising algorithm that used the log-likelihood of the Fisher-Tippett distribution in the low multilinear rank approximation was introduced. Five out of nine prostate cancer markers introduced in Chapter 3 gave the best ROC-AUC metric when tr-MLSVD introduced in Chapter 5 was used. Incorporating the added processing times, we found out that tr-MLSVD was the best option for denoising DCEUS acquisitions.

The tr-MLSVD is more suitable for WGN. The lower bounds of estimating tr-MLSVD are not known. In Chapter 7, we investigate such a theoretical bound by calculating the constrained Cramér-Rao bound using the components that make up an MLSVD with additive WGN. We aim to examine if such cases naturally occur in DCEUS sequences and the bound can be used to compare various tensor decomposition algorithms.

6.8. APPENDIX

6.8.1. RELAXATION OF THE MODEL FITTING WITH EXPONENTIAL LINEAR UNIT (ELU) FUNCTION

The model-fitting is the most time-consuming DCEUS feature, as described in Section 6.3.3. Here, we describe the relaxation of the model fitting through the ELU function, which resulted in faster convergence. The modified local density random walk [55] was described as

$$\mathbf{I}_{i_1 i_2 i_3}(i_4) = a^{(1)} \ln\left(\alpha \sqrt{\frac{\kappa}{2\pi(i_4 - t_0)}} \exp\left(-\frac{\kappa(i_4 - t_0 - \mu)^2}{2(i_4 - t_0)}\right) + 1\right). \tag{6.24}$$

where $a^{(1)}$ is the dynamic range-related scaling parameter given in (6.4), κ is the local dispersion-related parameter independent of the injection site's distance, μ is the convective time, and the t_0 is the injection time, α is the area under the time intensity curve. Using the ELU function, we can approximate the modified local density walk model as

$$\mathbf{I}_{i_1 i_2 i_3}(i_4) \approx \mathrm{ELU}\left(\theta_1 - \frac{1}{2} a^{(1)} \ln(i_4 - t_0) - \frac{\theta_2 (i_4 - \theta_3)}{2(i_4 - t_0)}\right), \tag{6.25}$$

with

$$\mathrm{ELU}(x) = \begin{cases} x, & \text{if } x > 0, \\ e^x - 1, & \text{if } x \leq 0, \end{cases} \tag{6.26}$$

and

$$\begin{aligned} \theta_1 &= a^{(1)} \ln \alpha + \frac{a^{(1)}}{2} \ln\left(\frac{\kappa}{2\pi}\right), \\ \theta_2 &= a^{(1)} \kappa, \\ \theta_3 &= \mu + t_0. \end{aligned} \tag{6.27}$$

The model is fitted starting from the appearance time, which is estimated separately.

6.8.2. MODIFICATION OF THE DERIVATIVES OF GTE

The derivative of the loss function in (6.8) with respect to the first factor matrix $\frac{\partial F}{\partial \mathbf{U}^{(1)}}$ is described in [132] as

$$\begin{aligned} \frac{\partial F}{\partial \mathbf{U}^{(1)}} = \left(\frac{\partial F}{\partial \underline{\mathbf{L}}}\right)_{(1)} \left(\mathbf{U}^{(4)} \otimes \mathbf{U}^{(3)} \otimes \mathbf{U}^{(2)}\right) \mathbf{S}_{(1)}^{\top} \\ + a\, \mathbf{U}^{(1)} \left(\left(\mathbf{U}^{(1)}\right)^{\top} \mathbf{U}^{(1)} - b^2 \mathbf{I}\right). \end{aligned} \tag{6.28}$$

We describe the modification of the derivatives on the first-factor matrix, which can be applied to the other factor matrices by changing the unfoldings.

Proposition 6.8.1. The Kronecker products given in (6.28) can be avoided by rewriting it as

$$\begin{aligned}\frac{\partial F}{\partial \mathbf{U}^{(1)}} =& \Big(\frac{\partial F}{\partial \underline{\mathbf{L}}} \times_2 \left(\mathbf{U}^{(2)}\right)^{\mathsf{T}} \times_3 \left(\mathbf{U}^{(3)}\right)^{\mathsf{T}} \times_4 \left(\mathbf{U}^{(4)}\right)^{\mathsf{T}}\Big)_{(1)} \left(\mathbf{S}_{(1)}\right)^{\mathsf{T}} \\ &+ a\, \mathbf{U}^{(1)} \left(\left(\mathbf{U}^{(1)}\right)^{\mathsf{T}} \mathbf{U}^{(1)} - b^2\, \mathbf{I}\right).\end{aligned} \tag{6.29}$$

Proof. Using the relation [14] between

$$\left(\underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \times_4 \mathbf{U}^{(4)}\right)_{(n)}, \tag{6.30}$$

and

$$\mathbf{U}_{(n)}\, \mathbf{S}_{(n)} \left(\mathbf{U}^{(4)} \otimes \cdots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \otimes \cdots \otimes \mathbf{U}^{(1)}\right)^{\mathsf{T}}, \tag{6.31}$$

the equality between the equations (6.28) and (6.29) can be proven. This can be shown by

$$\begin{aligned}&\Big(\frac{\partial F}{\partial \underline{\mathbf{L}}} \times_2 \left(\mathbf{U}^{(2)}\right)^{\mathsf{T}} \times_3 \left(\mathbf{U}^{(3)}\right)^{\mathsf{T}} \times_4 \left(\mathbf{U}^{(4)}\right)^{\mathsf{T}}\Big)_{(1)} \\ =& \Big(\frac{\partial F}{\partial \underline{\mathbf{L}}} \times_1 \mathbf{I} \times_2 \left(\mathbf{U}^{(2)}\right)^{\mathsf{T}} \times_3 \left(\mathbf{U}^{(3)}\right)^{\mathsf{T}} \times_4 \left(\mathbf{U}^{(4)}\right)^{\mathsf{T}}\Big)_{(1)} \\ =& \Big(\frac{\partial F}{\partial \underline{\mathbf{L}}}\Big)_{(1)} \left(\mathbf{U}^{(4)} \otimes \mathbf{U}^{(3)} \otimes \mathbf{U}^{(2)}\right).\end{aligned} \tag{6.32}$$

In a similar fashion, the derivative of the loss function with respect to the other factor matrices can be shown. □

7

CONSTRAINED CRAMÉR-RAO BOUND OF MULTILINEAR SINGULAR VALUE DECOMPOSITION

We investigated the MLSVD framework and t-SVD framework for denoising DCEUS acquisitions. For high SNR simulation scenarios, we found that tr-MLSVD gave the best performance in Chapter 4. In Chapters 5 and 6, we investigated the performance of tr-MLSVD for denoising DCEUS acquisitions. In this chapter, we continue with a theoretical analysis of estimating the components that make up an MLSVD. We calculate the constrained Cramér-Rao bound to further analyze the limitations of denoising. This chapter aims to answer the third research question.

7.1. INTRODUCTION

The analysis of tensors has been an active area of research, ranging from machine learning [146], biomedical signal processing [147] to large-scale optimization and data compression [7]. In many applications, the data tensor exhibits low-rank characteristics and is frequently contaminated by noise [15]. In this chapter, we consider the problem of estimating the components that make up a low multilinear rank tensor from noisy measurements.

There is not one single generalization of matrix rank to tensors. Many definitions of rank that coincide for matrices lead to different notions of rank for tensors. For example, the rank related to the canonical polyadic decomposition (CPD) is the smallest number of rank-one tensors that sum up to tensor $\underline{\mathbf{T}}$, similar to the smallest number of rank-one matrices that sum up to a matrix. Border rank

This chapter is based on the article "Constrained Cramér-Rao Bound of Higher-order Singular Value Decomposition" by M. Calis, R.T.Rajan, M.Mischi, A. van der Veen, B. Hunyadi, IEEE Open Journal of Signal Processing (IEEE-OJSP), 2025

generalizes the description of the set of ranks less than or equal to R matrices as an algebraic variety [148]. Different matricizations of a tensor $\underline{\mathbf{T}}$ will lower bound the tensor rank, such as grouping the first and the last two modes of a 4th-order tensor or the Koszul-Young flattenings [149]. The N different matricizations for an Nth-order tensor, commonly known as the mode-n unfoldings, can be considered as a generalization of modal ranks, i.e., the column and row rank of a matrix. The multilinear rank of a tensor is then an N-tuple with elements defined by the rank of each mode-n unfolding [13]. We focus on such low multilinear rank tensors, the root of which can be traced back to the Tucker decomposition [150].

Any tensor can be expressed as the product of a core tensor and a set of factor matrices, which together constitute the Tucker decomposition [150]. This decomposition is not unique in its unconstrained form: the factor matrices can be right-multiplied by any invertible matrix, provided the core tensor is adjusted accordingly, i.e., multiplied by their inverse, to yield the same reconstructed tensor. To resolve this ambiguity, the authors in [13] proposed the multilinear singular value decomposition (MLSVD) by introducing structural constraints, specifically, the all-orthogonality of the core tensor and the orthonormality of the factor matrices. The factor matrices are obtained by applying the singular value decomposition (SVD) to each mode-n unfolding of the tensor, taking the left singular vectors as the factor matrices. The all-orthogonal core tensor is then computed by multiplying the tensor with the transposes of the factor matrices along the corresponding modes. When the singular values in each mode-n unfolding are distinct and ordered, the MLSVD of a real-valued tensor is unique up to a sign ambiguity [13].

According to the Eckart-Young theorem [21], the best low-rank approximation of a matrix can be obtained through the truncated SVD if we consider the Frobenius norm on the residual error. An analogous extension to the higher-order case is not possible for MLSVD. The truncated multilinear singular value decomposition (tr-MLSVD) does not guarantee the best $(R_1, R_2, \ldots, R_N)$ multilinear rank approximation of a tensor for both noisy and noiseless cases [19]. Higher-order orthogonal iteration (HOOI) improves the estimation by alternating updates of the factor matrices. Note that in both the matrix and the tensor case, noise can perturb the singular values, singular vectors, and the MLSVD factors, respectively [151][152]. An important question remains: How well can any unbiased estimator recover the components of the true MLSVD components?

The Cramér-Rao bound (CRB) [153] is a statistical tool to study the performance of an unbiased estimator. Under certain regularity conditions, the CRB provides a lower bound on the asymptotic performance of locally unbiased estimators. In the presence of additional constraints, the authors in [154] introduced the constrained Cramér-Rao bound (CCRB) that lower bounds the error covariance for constrained, unbiased estimators when the unconstrained model has a nonsingular Fisher information matrix (FIM). This is further generalized in [155], where the authors introduced the CCRB without assuming a full rank FIM [156]. By incorporating the constraints and assumptions that create a unique MLSVD, we aim to calculate a lower bound for estimating the noiseless factor matrices and the all-orthogonal core tensor.

The CRB is calculated for a CPD in [157] and further explored in [146]. In [158], the CCRB of a coupled CPD is calculated. In [159], the CCRB of the delayed exponential fitting problem using MLSVD is explored. In this chapter, we present

the CCRB for the MLSVD of a real tensor under additive white Gaussian noise (WGN). We assume the multilinear ranks are known and there is a unique MLSVD decomposition. We compare our proposed bound with HOOI and tr-MLSVD, which are well-suited for signals in the presence of WGN.

The layout of the chapter is as follows. In Section 7.2, tensor notation and tensor preliminaries are introduced. In Section 7.3, the signal model is introduced, and the problem is formulated. In Section 7.4 and Section 7.5, the CRB and the CCRB are introduced. The supplementary material for calculating the bounds can be found in the Appendix. In Section 7.6, we compare the CRB and CCRB with HOOI and tr-MLSVD through simulation. Finally, we conclude the chapter in Section 7.8.

7.2. TENSOR NOTATION AND PRELIMINARIES

The tensor preliminaries are explained in Chapter 2. We briefly introduce some extra notations used in this chapter. The Kronecker delta $\delta_{i_1 i_2}$ is 1 when $i_1 = i_2$ and zero otherwise. We use selection matrices, permutation matrices, and matrices that define indexing relations. $\mathbf{E}^{(n)} \in \mathbb{R}^{R_n(R_n+1)/2 \times R_n^2}$ and $\mathbf{K}^{(n)} \in \mathbb{R}^{R_n(R_n-1)/2 \times R_n^2}$ for $n \in \{1, \ldots, N\}$ are binary-valued matrices. For $n \in \{1, \ldots, N\}$, the products $\mathbf{E}^{(n)} \operatorname{vec}(\mathbf{C}^{(n)})$ and $\mathbf{K}^{(n)} \operatorname{vec}(\mathbf{C}^{(n)})$ extract the lower triangle and the strict lower triangle (excluding the diagonal) of $\mathbf{C}^{(n)} \in \mathbb{R}^{R_n \times R_n}$, respectively. Permutation matrices $\mathbf{P}^{(n)} \in \mathbb{R}^{D \times D}$ for $n \in \{1, \ldots, N\}$ define a row-wise permutation between the vectorizations of mode-n unfoldings and the vectorization of mode-1 unfolding, where $D = \prod_{n=1}^{N} I_n$. An example is the permutation matrix $\mathbf{P}^{(2)}$ that permutes the vectorization of the mode-2 unfolding of $\underline{\mathbf{Y}}$, such that vectorization of the mode-1 unfolding of $\underline{\mathbf{Y}}$ is obtained, i.e. $\operatorname{vec}(\mathbf{Y}_{(1)}) = \mathbf{P}^{(2)} \operatorname{vec}(\mathbf{Y}_{(2)})$. The matrix $\mathbf{P}^{(2)}$ can be generated by first creating a square matrix of 0s of size $D \times D$ and assigning $p^{(2)}_{\overline{i_1 i_2 \ldots i_N}, \overline{i_2 i_1 \ldots i_N}} = 1$. As an example, we illustrate the structure of the permutation matrix $\mathbf{P}^{(2)}$ in Fig. 2.1 for a tensor $\underline{\mathbf{M}} \in \mathbb{R}^{2 \times 2 \times 2}$.

An entry in the tensor $\underline{\mathbf{Y}}$ is placed in different rows or columns according to the indexing rules of the mode-n unfolding and vectorization. We use the matrix $\mathbf{G}^{(n;i_n,k;i_k)} \in \mathbb{R}^{D/I_n \times D/I_k}$ to define an indexing relation between the i_nth row of $\mathbf{Y}_{(n)}$ and i_kth row of $\mathbf{Y}_{(k)}$ for $n \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, N\}$. Additionally, we use $\mathbf{G}^{(n;i_n)} \in \mathbb{R}^{D/I_n \times D}$ to define the indexing relation between the elements of the i_nth row of $\mathbf{Y}_{(n)}$ and $\operatorname{vec}(\underline{\mathbf{Y}})$. To achieve this, we introduce a new notation to describe a grouping with stride, which is formulated as $i_j \overline{i_1 \ldots i_{j-1} i_{j+1} \ldots i_N} = i_j + i_1 + (i_2 - 1)I_1 + \cdots + (i_{j-1} - 1)I_1 \ldots I_{j-2} + (i_{j+1} - 1)I_1 \ldots I_{j-1} + \cdots + (i_N - 1)I_1 \ldots I_{j-1} I_{j+1} \ldots I_{N-1}$ for $j \in \{1, \ldots, N\}$. For illustration, consider the matrix $\mathbf{G}^{(1;i_1,2;i_2)} \in \mathbb{R}^{I_2 I_3 \times I_1 I_3}$ that defines an indexing correspondence between the i_1th row of $\mathbf{Y}_{(1)}$, and the i_2th row of $\mathbf{Y}_{(2)}$. The element $g_{i_1 \overline{i_2 i_3}, i_2 \overline{i_1 i_3}}$ is 1, if $i_1 \overline{i_2 i_3} = i_2 \overline{i_1 i_3}$ for i_1, i_2 and $i_3 \in \{1, \ldots, I_3\}$ and zero otherwise. This is generated by creating matrix of 0s of size $\mathbf{G} \in \mathbb{R}^{I_2 I_3 \times I_1 I_3}$ and assigning $g_{i_1 \overline{i_2 i_3}, i_2 \overline{i_1 i_3}} = 1$. An example is $\mathbf{G}^{(1;2,2;3)}$ which has 1 at indices that are the same between the 2nd row of the mode-1 unfolding of a 3rd-order tensor of size (3,3,3) and the 3rd row of the mode-2 unfolding of the same tensor and 0s elsewhere. Note that i_1 and i_2 are fixed and i_3 belongs to the set $\{1, \ldots, I_3\}$. The matrix structure is illustrated in Fig. 7.1. Finally, $\mathbf{G}^{(n;i_n)} \in \mathbb{R}^{D/I_n \times D}$ can be created by initializing a matrix of 0s and assigning $g_{i_n \overline{i_1 \ldots i_{n-1} i_{n+1} \ldots i_N}, \overline{i_1 \ldots i_N}} = 1$.

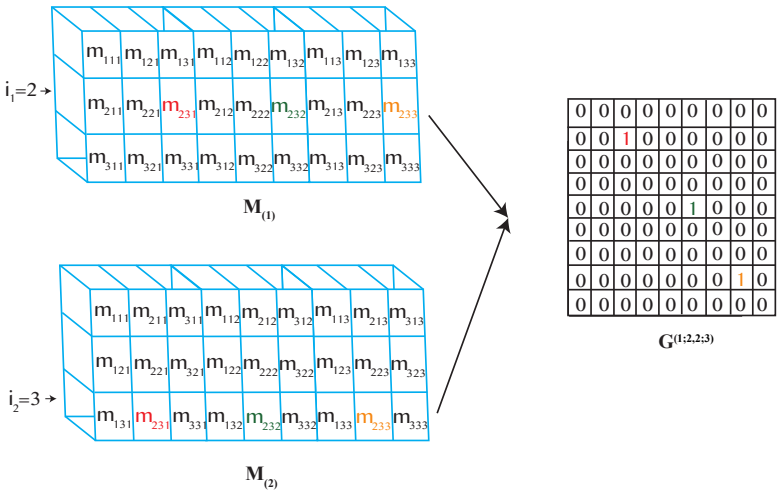


Figure 7.1: The structure of the $\mathbf{G}^{(1;2,2;3)}$ matrix for a tensor $\underline{\mathbf{M}} \in \mathbb{R}^{3\times3\times3}$. $\mathbf{G}$ has 1 at indices that are the same between the 2nd row of the mode-1 unfolding of $\underline{\mathbf{M}}$ and the 3rd row of the mode-2 unfolding of the same tensor and 0s elsewhere.

7.3. PROBLEM FORMULATION

In this chapter, we aim to find the CCRB of the core tensor and the factor matrices of a low multilinear rank tensor from a measurement model with additive noise, which can be formulated by

$$\underline{\mathbf{Y}} = \underline{\mathbf{L}} + \underline{\mathbf{M}}, \tag{7.1}$$

where $\underline{\mathbf{L}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ is a deterministic low multilinear rank tensor, and $\underline{\mathbf{M}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ is assumed to be a zero-mean white Gaussian noise tensor where each entry is independent and identically distributed with variance $\mathcal{N}(0,\sigma^2)$. We introduced the speckle noise in Chapter 3 as WGN with sparse outliers. For high SNR, the benefit obtained from the addition of the sparsity constraint was minimal, as described in Chapter 4. For tractability, we assume the noise is WGN and introduce the bounds accordingly.

The low multilinear rank tensor can be written as

$$\underline{\mathbf{L}} = \underline{\mathbf{S}} \times_1 \mathbf{U}^{(1)} \cdots \times_N \mathbf{U}^{(N)}, \tag{7.2}$$

with orthonormal factor matrices $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n\times R_n}$ for $n \in \{1,\ldots,N\}$ and an all-orthogonal core tensor $\underline{\mathbf{S}} \in \mathbb{R}^{R_1\times\cdots\times R_N}$, where we assume that the ranks R_n are known for $n \in \{1,\ldots,N\}$.

We can now write the deterministic parameters of interest as $\boldsymbol{\theta} \in \mathbb{R}^{N_\theta}$

$$\boldsymbol{\theta} = [(\text{vec}(\mathbf{U}^{(1)}))^T, \ldots, (\text{vec}(\mathbf{U}^{(N)}))^T, (\text{vec}(\underline{\mathbf{S}}))^T]^T, \tag{7.3}$$

where $N_{\boldsymbol{\theta}} = \sum_{n=1}^{N} I_n R_n + \prod_{n=1}^{N} R_n$ is the total size of the factor matrices and the core tensor.

In the following sections, we introduce the CRB and calculate the CCRB of $\boldsymbol{\theta}$ (7.3) from the signal model (7.1), given the tensor structure (7.2).

7.4. CRAMÉR-RAO BOUND

Let $f(\underline{\mathbf{Y}};\boldsymbol{\theta})$ be the likelihood of the observed data $\underline{\mathbf{Y}}$, given the model parameters $\boldsymbol{\theta}$. The log-likelihood $\ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})$ can be written in N different ways. For each mode $n \in \{1,\ldots,N\}$, we have

$$\ln f(\underline{\mathbf{Y}};\boldsymbol{\theta}) = -\left(\prod_{n=1}^{N} I_n\right)\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\left\|\mathbf{Y}_{(n)} - \mathbf{L}_{(n)}\right\|_F^2, \tag{7.4}$$

which is the logarithm of the multiplication of zero-mean Gaussian probability distribution functions. We take the partial derivative of the log-likelihood with respect to each element of (7.3) and create the FIM that is defined as

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \mathbb{E}\left\{\left(\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right\}, \tag{7.5}$$

with size $\mathbb{R}^{N_{\boldsymbol{\theta}} \times N_{\boldsymbol{\theta}}}$.

The FIM is a matrix consisting of submatrices in the following form

$$\boldsymbol{\Omega}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Omega}_{\mathbf{U}^{(1)}\mathbf{U}^{(1)}} & \cdots & \boldsymbol{\Omega}_{\mathbf{U}^{(1)}\mathbf{U}^{(N)}} & \boldsymbol{\Omega}_{\mathbf{U}^{(1)}\underline{\mathbf{S}}} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Omega}^T_{\mathbf{U}^{(1)}\mathbf{U}^{(N)}} & \cdots & \boldsymbol{\Omega}_{\mathbf{U}^{(N)}\mathbf{U}^{(N)}} & \boldsymbol{\Omega}_{\mathbf{U}^{(N)}\underline{\mathbf{S}}} \\ \boldsymbol{\Omega}^T_{\mathbf{U}^{(1)}\underline{\mathbf{S}}} & \cdots & \boldsymbol{\Omega}^T_{\mathbf{U}^{(N)}\underline{\mathbf{S}}} & \boldsymbol{\Omega}_{\underline{\mathbf{S}}\underline{\mathbf{S}}} \end{bmatrix}, \tag{7.6}$$

where the submatrices divide the FIM according to the parts of $\boldsymbol{\theta}$. The FIM is a symmetric matrix. Therefore, only the upper or lower triangle needs to be calculated. The partial derivatives and the elements of the FIM are given in the Appendix from (7.20)-(7.23).

The CRB is the trace of the inverse of the Fisher information matrix. Asymptotically, the mean squared error (MSE) of any locally unbiased estimator is greater than or equal to the CRB, i.e.,

$$\mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_F^2\right] \geq \mathrm{Tr}\left(\boldsymbol{\Omega}(\boldsymbol{\theta})^{-1}\right), \tag{7.7}$$

where $\hat{\boldsymbol{\theta}}$ is an estimate of the true parameter vector $\boldsymbol{\theta}$. The FIM (7.6) is observed to be typically singular, and hence the inverse does not exist. The trace of the pseudo-inverse of the FIM is called the oracle bound (OB) [160], which can be shown as

$$\mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_F^2\right] \geq \mathrm{Tr}\left(\boldsymbol{\Omega}(\boldsymbol{\theta})^{\dagger}\right), \tag{7.8}$$

where $\hat{\boldsymbol{\theta}}$ is an estimate of the true parameter vector $\boldsymbol{\theta}$. The OB cannot be achieved without additional constraints [161]. Given additional constraints, e.g., the orthonormality of the factor matrices and the all-orthogonality of the core tensor, respectively, a modified CRB can be obtained, i.e., the constrained Cramér-Rao bound.

7.5. CONSTRAINED CRAMÉR-RAO BOUND

We have two constraints on the unknown parameters $\boldsymbol{\theta}$. First, the factor matrices are orthonormal, which can be expressed for $n \in \{1, \ldots, N\}$ as

$$\mathbf{E}^{(n)} \text{vec}(\mathbf{U}^{(n)^T} \mathbf{U}^{(n)} - \mathbf{I}) = \mathbf{0}, \tag{7.9}$$

which has $R_n(R_n+1)/2$ independent equations. Second, the core tensor is all-orthogonal. This condition implies that all off-diagonal elements of $\mathbf{S}_{(n)} \mathbf{S}_{(n)}^T$ for $n \in \{1, \ldots, N\}$ are zero, that is,

$$\mathbf{K}^{(n)} \text{vec}(\mathbf{S}_{(n)} \mathbf{S}_{(n)}^T) = \mathbf{0}, \tag{7.10}$$

which has $R_n(R_n-1)/2$ independent equations. Since both $\mathbf{S}_{(n)} \mathbf{S}_{(n)}^T$ and $\mathbf{U}^{(n)^T} \mathbf{U}^{(n)}$ are symmetric matrices, we extract the strictly triangular and the triangular parts, respectively, using the selection matrices. The selection matrices $\mathbf{E}^{(n)}$ and $\mathbf{K}^{(n)}$ are introduced in Section 7.2. In total, there are $N_c = \sum_{n=1}^{N}(R_n(R_n+1))/2 + \sum_{n=1}^{N}(R_n(R_n-1))/2 = \sum_{n=1}^{N} R_n^2$ number of constraints. The constraints in (7.9) and (7.10) can be stacked into a column vector $\mathbf{c}(\boldsymbol{\theta}) = \mathbf{0} \in \mathbb{R}^{N_c}$.

The partial derivative of $\mathbf{c}(\boldsymbol{\theta})$ with respect to the unknown $\boldsymbol{\theta}$ yields

$$\mathbf{C}(\boldsymbol{\theta}) = \frac{\partial \mathbf{c}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}, \tag{7.11}$$

where $\mathbf{C}(\boldsymbol{\theta}) \in \mathbb{R}^{N_c \times N_\theta}$. The matrix $\mathbf{C}(\boldsymbol{\theta})$ consists of several submatrices in the following form

$$\mathbf{C}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{C}_{\mathbf{U}^{(1)}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{\mathbf{U}^{(N)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{\mathbf{S}_{(1)}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{\mathbf{S}_{(2)}} \mathbf{P}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}_{\mathbf{S}_{(N)}} \mathbf{P}^{(N)} \end{bmatrix}. \tag{7.12}$$

The elements of the submatrices in (7.12) are described in (7.25) in the Appendix. The submatrix $\mathbf{C}_{\mathbf{U}^{(n)}} \in \mathbb{R}^{R_n(R_n+1)/2 \times I_n R_n}$ is the partial derivative of (7.9) with respect to $(\text{vec}(\mathbf{U}^{(n)}))^T$ for $n \in \{1, \ldots, N\}$. Additionally, the submatrix $\mathbf{C}_{\mathbf{S}_{(n)}} \in \mathbb{R}^{R_n(R_n-1)/2 \times \prod_{n=1}^{N} R_n}$ is the partial derivative of (7.10) with respect to $(\text{vec}(\mathbf{S}_{(n)}))^T$ for $n \in \{1, \ldots, N\}$. The partial derivatives $\mathbf{C}_{\mathbf{S}_{(n)}}$ are calculated in (7.26), however the $\text{vec}(\mathbf{S}_{(n)})$ for $n \in \{2, \ldots, N\}$ are not consistent with the $\text{vec}(\mathbf{S}_{(1)}) = \text{vec}(\underline{\mathbf{S}})$. Due to the indexing difference between the vectorization of the different unfoldings of $\underline{\mathbf{S}}$, we use the permutation matrices $\mathbf{P}^{(\cdot)}$.

Define $\mathbf{V}$ as an orthonormal basis for the null space of $\mathbf{C}(\boldsymbol{\theta})$, which can be found through SVD [14]. The CCRB [155] can then be calculated by

$$\mathbb{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_F^2\right] \geq \text{Tr}\left(\mathbf{V}\left(\mathbf{V}^\mathsf{T} \boldsymbol{\Omega}(\boldsymbol{\theta}) \mathbf{V}\right)^{-1} \mathbf{V}^\mathsf{T}\right), \tag{7.13}$$

where $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is the Fisher information matrix defined in (7.5). The CCRB only exists if $N_c \geq N_\theta - \text{rank}(\boldsymbol{\Omega}(\boldsymbol{\theta}))$ and the constraints are linearly independent, which are

sufficient conditions for the inverse of $\mathbf{V}^T \boldsymbol{\Omega}(\boldsymbol{\theta}) \mathbf{V}$ to exist as the constrained FIM then becomes full-rank. The matrix on the right-hand side in (7.13) becomes the OB if $\mathbf{V}$ is a basis for the null space of the FIM in (7.8). With unknown parameters, we do not know constraints that give such a $\mathbf{V}$. Since OB has a larger parameter space, including those provided in (7.9) and (7.10), the OB is less than or equal to the CCRB [161], that is,

$$\mathrm{Tr}\left(\boldsymbol{\Omega}(\boldsymbol{\theta})^{\dagger}\right) \leq \mathrm{Tr}\left(\mathbf{V}\left(\mathbf{V}^{\top}\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{V}\right)^{-1}\mathbf{V}^{\top}\right). \tag{7.14}$$

7.6. SIMULATION

In this section, we analyze the performance of existing low multilinear rank tensor estimation methods, tr-MLSVD [13] and HOOI [19], against the computed lower bounds. We make sure that the simulated tensors have unique mode-n singular values by following [162], where a tensor that satisfies the prescribed mode-n singular values is generated. Let $\boldsymbol{\Sigma}^{(n)}$ represent the mode-n singular values of $\underline{\mathbf{L}}$. Specifically, we select mode-n singular values that are separated by a $d \in \{2, 1, 0.5, 0.25\}$ in each mode, that is,

$$\boldsymbol{\Sigma}^{(n)}(i_n - 1, i_n - 1) - \boldsymbol{\Sigma}^{(n)}(i_n, i_n) = d, \tag{7.15}$$

for $n \in \{1, \ldots, N\}$ and $i_n \in \{2, \ldots, I_n\}$. We investigate two scenarios. For both scenarios, we fix the squared sum of the mode-n singular values to 200, that is,

$$\mathrm{Tr}(\boldsymbol{\Sigma}^{(n)}{\boldsymbol{\Sigma}^{(n)}}^T) = \mathrm{Tr}(\mathbf{S}_{(n)}{\mathbf{S}_{(n)}}^T) = 200, \tag{7.16}$$

for $n \in \{1, \ldots, N\}$ and generate equally spaced singular values according to the desired multilinear ranks defined therein.

In the first scenario, the effect of the order on CCRB, OB, and MSE is observed. We increase the order of the tensor from 4 to 7, without changing the rank or the singular value distribution in each mode. The setup of Scenario 1 is summarized in Table 7.1, and the mode-n singular values are plotted in Fig. 7.2. In the second scenario, the effect of the distribution of the mode-n singular values on CCRB, OB, and MSE is analyzed. Therefore, we fixed the multilinear ranks and the order of the tensor, and changed the distance between each consecutive mode-n singular values using $d \in \{2, 1, 0.5, 0.25\}$.

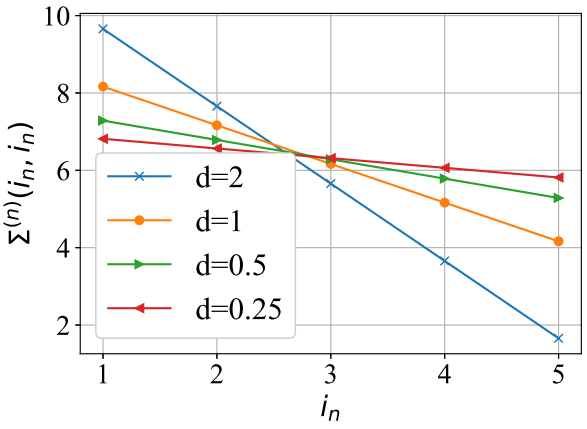


Figure 7.2: Mode-n singular values of the 4th-order tensors with multilinear rank (5,5,5,5) generated in Scenario 2 as described in Section 7.6.

The CCRB and OB are calculated using (7.13) and (7.8) for an SNR range of $\{0, 5, 10, 15, 20, 25, 30\}$ dB. The following steps are taken to achieve a given SNR. Each element of the noise tensor $\underline{\mathbf{M}}$ is sampled from a normal distribution. The scaling factor is then obtained by the ratio of the standard deviation of $\underline{\mathbf{L}}$ to the standard deviation $\underline{\mathbf{M}}$ scaled by an SNR-related parameter, that is,

$$\sigma = \frac{1}{\sqrt{10^{\mathrm{SNR}/10}}} \cdot \frac{\left\|\underline{\mathbf{L}} - \bar{\underline{\mathbf{L}}}\right\|_F}{\left\|\underline{\mathbf{M}} - \bar{\underline{\mathbf{M}}}\right\|_F}, \tag{7.17}$$

which is consecutively multiplied with each entry of $\underline{\mathbf{M}}$.

The CCRB and OB are compared with the estimates from HOOI [19] and tr-MLSVD [13], which are averaged over 10^4 Monte Carlo simulations. The tolerance of HOOI is set to 10^{-6}, the maximum iteration is set to 2000, and it is initialized with the tr-MLSVD. The true ranks are used to truncate in tr-MLSVD. Define the estimate of the true vector at the tth Monte Carlo iteration as $\hat{\boldsymbol{\theta}}^{(t)}$. The mean squared error is calculated by

$$\frac{1}{10000} \sum_{t=1}^{10000} \left\|\hat{\boldsymbol{\theta}}^{(t)} - \boldsymbol{\theta}\right\|_F^2 \tag{7.18}$$

for both estimators after fixing the sign ambiguity. We fix the sign ambiguity and permutation ambiguity by using the Hungarian algorithm. The sign of the estimated singular vector and the corresponding slice of the estimated core tensor are flipped accordingly. We enforce all-orthogonality of the estimated core tensor by applying an additional MLSVD decomposition and absorbing the rotation into the estimated factor matrices.

7.7. RESULTS

The results are plotted for the two scenarios in Fig. 7.3 and Fig. 7.4, respectively. In Fig. 7.3, we investigated the effect of order on the MSE between the true parameter $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ on four subplots. In addition, the OB and CCRB are shown in each subplot. Starting from the left subplot, multilinear ranks $(3,3,3,3,3,3,3)$, $(3,3,3,3,3,3)$, $(3,3,3,3,3)$, and $(3,3,3,3)$ are investigated, consecutively. In each case, we fixed the gap between the modal singular values to $d = 2$. We can observe that as the orders increase, the distance between the OB and the CCRB decreases, and a tight connection between the estimates and the CCRB is observed for all SNRs. As the order decreases, a deviation from the CCRB is observed for both estimators, which can be seen in the right-most subplot of Fig. 7.3. The effect of the gap between the consecutive mode-n singular values on the CCRB is shown using gaps from the set $d \in \{2, 1, 0.5, 0.25\}$ in Fig. 7.4. Starting from the left, the gaps $d \in \{2, 1, 0.5, 0.25\}$ are represented, consecutively. We fixed the multilinear ranks to $(5,5,5,5)$ in all subplots. Similarly, the MSE between the true parameter $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ is plotted along with OB and CCRB. We see that as the gaps decrease, a deviation of the MSE of estimates from the CCRB is observed, and the distance between OB and the CCRB increases. Additionally, when the gap between the modal singular values decreases, the estimators perform significantly better than CCRB for low SNRs. In high SNRs, the estimators converge to CCRB.

We observed the rank of the constrained FIM to be equal to the dimension of the manifold of Tucker tensors as provided in Theorem 3.6 in [163]. The rank of the constrained FIM was always

$$N_\theta - N_c = \sum_{n=1}^{N} I_n R_n + \prod_{n=1}^{N} R_n - \sum_{n=1}^{N} R_n^2, \tag{7.19}$$

which confirms that $(\mathbf{V}^T \boldsymbol{\Omega}(\boldsymbol{\theta}) \mathbf{V})$ is of full rank and the inverse exists.

We tabularized the summary of the two scenarios in Table 7.1 along with the condition number of the constrained FIM (7.13) for the tensors that are generated. Increasing the tensor order leads to a reduction in the condition number, indicating that the estimation problem becomes better conditioned. Similarly, greater separation between consecutive mode-n singular values also results in a lower condition number. In well-conditioned cases, such as those shown in the leftmost subplots of Fig. 7.3 and Fig. 7.4, the CCRB approaches the oracle bound (OB) closely.

7.8. DISCUSSION AND CONCLUSION

In this chapter, we introduced a lower bound on the mean-squared error of the unbiased estimates of the components of MLSVD under additive WGN using CCRB. Asymptotically, for high SNRs and high Monte Carlo simulations, we showed through simulation that the HOOI and tr-MLSVD converge fairly close to the CCRB. In the non-asymptotic case, a deviation is observed from the CCRB. The SNR in which the convergence occurs is dependent on the condition number of $(\mathbf{V}^T \boldsymbol{\Omega}(\boldsymbol{\theta}) \mathbf{V})$, which is shown in Table 7.1. If the condition number is high, both estimators get close to the CCRB in higher SNRs. Such a case occurs when the singular values of the true tensor are close to each other, and the order of the

Table 7.1: Summary of the scenarios and the condition number of constrained FIM introduced in Section 7.6.

	Multilinear Ranks	Mode-n Singular Value Separation	Condition Number of $(\mathbf{V}^T \boldsymbol{\Omega}(\boldsymbol{\theta}) \mathbf{V})$
Scenario 1	(3,3,3,3,3,3,3)	$d = 2$	$2 \cdot 10^3$
	(3,3,3,3,3,3)	$d = 2$	$3 \cdot 10^3$
	(3,3,3,3,3)	$d = 2$	$4 \cdot 10^3$
	(3,3,3,3)	$d = 2$	$6 \cdot 10^3$
Scenario 2	(5,5,5,5)	$d = 2$	$2 \cdot 10^3$
	(5,5,5,5)	$d = 1$	$8 \cdot 10^3$
	(5,5,5,5)	$d = 0.5$	$20 \cdot 10^3$
	(5,5,5,5)	$d = 0.25$	$72 \cdot 10^3$

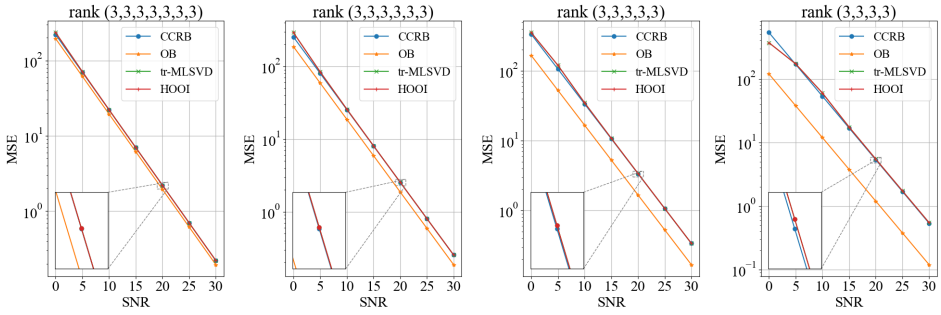


Figure 7.3: The CCRB, OB, and the MSE between the true parameters of interest $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ obtained by HOOI and tr-MLSVD for tensors in Scenario 1, where the distribution between the mode-n singular values is fixed and the order is changed. The CCRB, OB, and MSE are defined in (7.13), (7.7), and (7.18), respectively. The setup of Scenario 1 is described in Table 7.1.

true tensor is low. For well-conditioned scenarios, the OB and CCRB get close to each other, which suggests that the theoretical gain provided by any constraints beyond those presented in (7.9) and (7.10) will be marginal.

We have observed that the MSE of the estimates can lie between CCRB and OB in low SNRs if the condition number is high. This can be seen in Fig. 7.4. Both tr-MLSVD and HOOI fail to hold the constraint (7.10) in the non-asymptotic region. In the noisy setting, the off-diagonals of the truncated core tensor are not zero. The lack of this restriction can explain their better performance compared to CCRB in Fig. 7.4. The authors in [164] introduce the CCRB using Lehman-unbiasedness, which is a weaker restriction than the unbiasedness definition in [155], which might further improve the bound in the non-asymptotic region.

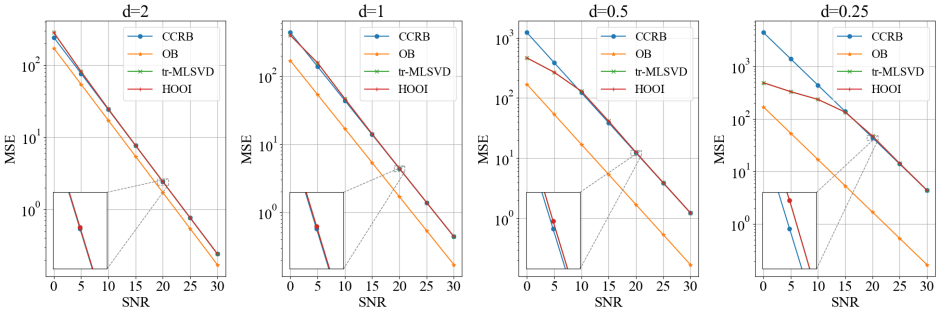


Figure 7.4: The CCRB, OB, and the MSE between the true parameters of interest $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ obtained by HOOI and tr-MLSVD for tensors in Scenario 2, where the order is fixed but the gap d between the consecutive mode-n singular values is changed. The CCRB, OB, and MSE are defined in (7.13), (7.7), and (7.18), respectively. The setup of Scenario 2 is described in Table 7.1.

The two estimators that are used in this chapter do not enforce all-orthogonality into the low-rank approximation. Note that although the MLSVD core is all-orthogonal, the truncation step, which removes columns, rows, and fibers of the core, may render the new tr-MLSVD core not all-orthogonal. Although we apply an additional MLSVD operation such that the tr-MLSVD core is all-orthogonal, inherently, the estimators do not incorporate such a constraint. The HOOI estimates $\underline{\mathbf{L}}$ by minimizing $||\underline{\mathbf{Y}}-\underline{\mathbf{L}}||^2$, with the constraint that $\underline{\mathbf{L}}$ has a low multilinear rank, which is the mean squared estimate of a low-rank approximation. Similar to the tr-MLSVD, the all-orthogonality of the core tensor is not enforced during estimation.

Additionally, the unbiasedness assumption of the CCRB is violated for both HOOI and tr-MLSVD. When a matrix is perturbed, its eigenvectors and eigenvalues change, and the magnitude of this change depends on the noise power and the spectral gap between the eigenvalues [165]. While classical perturbation theory, such as in [166] and references therein, shows that subspace estimates (e.g., eigenvectors) may be asymptotically unbiased under high SNR and well-separated eigenvalues, this assumption does not always hold in finite-sample scenarios. In particular, when eigenvalues are closely spaced or repeated, the SVD becomes non-unique due to rotational ambiguity among the corresponding eigenvectors.

In the multilinear case, although the mode-n unfoldings are processed independently, the resulting components (factor matrices and the core tensor) are assembled jointly. Therefore, noise-induced perturbations in each unfolding can influence the overall decomposition and estimation accuracy. This effect is more difficult to characterize analytically than in the matrix case and has been noted in multilinear perturbation studies [152]. The right-most subplots of Fig. 7.3 and Fig. 7.4 confirm this, showing that the MSE of the estimates can fall below the

CCRB, which is indicative of estimator bias. For well-separated mode-n singular values and high SNR scenarios, the estimates from tr-MLSVD and HOOI lie fairly close to the CCRB, indicating an unbiased estimation regime. Thus, while asymptotic unbiasedness is theoretically valid under ideal conditions, in general, tensor decomposition can often deviate from this assumption.

We propose additional directions for future work. The constrained and the unconstrained FIM given in (7.13) and (7.6) can get fairly large with increasing ranks and orders. We have not conducted a study to invert the matrices efficiently. Matrix inversion lemma such as the one used in [146] can be used to invert the FIM in a less expensive way. The CCRB is calculated assuming that the noise is WGN. Only the log-likelihood in (7.20) depends on the noise type. Given that the unbiasedness and the regularity conditions hold, the constrained Cramér-Rao bound can be calculated for other additive noise types. The subspace spanned by the factor matrices is commonly used in signal processing applications [167][168]. If the subspaces are of interest, the intrinsic Cramér-Rao bound [169] can be readily extended to a higher-order case.

Many low-rank tensor decomposition methods, such as HOOI [19] tr-MLSVD [13], or algorithms based on optimization on manifolds [18] [163], provide theoretical upper bounds to the low-rank approximation problem. We introduced the CCRB as an information-theoretic lower bound for the low multilinear rank approximation problem. More significantly, we identify the condition number of $\mathbf{V}^T \boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{V}$ as a metric that explains how the locally unbiased estimators closely approach the CCRB, particularly in non-asymptotic and high condition number regimes. In such regions, the OB and CCRB are separated, suggesting that there are constraints other than (7.9) and (7.10), which could yield a lower MSE. We leave the investigation of more efficient or optimal estimation techniques as future work.

7.9. CHAPTER SUMMARY

In high SNR scenarios, adding sparsity constraints did not significantly improve the results of despeckling performance. Therefore, in this chapter, we introduced a lower bound on the mean squared error of the unbiased estimates of the components of MLSVD under additive WGN using CCRB. We found out that for well-conditioned scenarios defined by the condition number of the constrained FIM (7.13), the mean squared errors of tr-MLSVD and HOOI were explained by the proposed lower bound. The problem became well conditioned if the mode-n singular values of the original tensor were significantly different from each other.

In the next chapter, we collect all the results from the previous chapters and conclude the dissertation by answering the research questions. Finally, we provide future research directions.

7.10. APPENDIX

Let $\mathbf{e}_{i_n} \in \mathbb{R}^{I_n}$, $\mathbf{e}_{r_n} \in \mathbb{R}^{R_n}$, and $\mathbf{e}_{\overline{r_1 \dots r_N}} \in \mathbb{R}^{\prod_{n=1}^{N} R_n}$ be the basis vectors that have 1 located at i_n, r_n and $r_1 + \cdots + (r_N - 1)R_1 \dots R_{N-1}$, respectively, and 0s elsewhere. We can then write the partial derivative of $\ln f(\underline{\mathbf{Y}}; \boldsymbol{\theta})$ with respect to the elements

of $\boldsymbol{\theta}$ as

$$\begin{aligned}\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \mathbf{U}^{(n)}(i_n,r_n)} &= \frac{1}{\sigma^2}\mathbf{e}_{i_n}^T(\mathbf{Y}_{(n)}-\mathbf{L}_{(n)})(\mathbf{U}^{(N)}\otimes\cdots\otimes\mathbf{U}^{(n+1)}\otimes\mathbf{U}^{(n-1)}\otimes\cdots\otimes\mathbf{U}^{(1)})\mathbf{S}_{(n)}^T\mathbf{e}_{r_n},\\ \frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \underline{\mathbf{S}}(r_1,\ldots,r_N)} &= \frac{1}{\sigma^2}\mathbf{e}_{\overline{r_1\ldots r_N}}^T(\mathbf{U}^{(N)}\otimes\cdots\otimes\mathbf{U}^{(1)})^T(\mathrm{vec}(\underline{\mathbf{Y}})-\mathrm{vec}(\underline{\mathbf{L}})).\end{aligned} \tag{7.20}$$

The submatrices of the FIM in (7.6) are defined for $n\in\{1,\ldots,N\}$ as

$$\begin{aligned}\boldsymbol{\Omega}_{\mathbf{U}^{(n)}\mathbf{U}^{(n)}}\left(\overline{i_nr_n},\overline{i_nr_n}\right) &= \mathbb{E}\Big[\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \mathbf{U}^{(n)}(i_n,r_n)}\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \mathbf{U}^{(n)}(i_n,r_n)}\Big],\\ \boldsymbol{\Omega}_{\mathbf{U}^{(n)}\underline{\mathbf{S}}}\left(\overline{i_nr_n},\overline{r_1\ldots r_N}\right) &= \mathbb{E}\Big[\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \mathbf{U}^{(n)}(i_n,r_n)}\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \underline{\mathbf{S}}(r_1,\ldots,r_N)}\Big],\\ \boldsymbol{\Omega}_{\underline{\mathbf{S}}\underline{\mathbf{S}}}\left(\overline{r_1\ldots r_N},\overline{r_1\ldots r_N}\right) &= \mathbb{E}\Big[\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \underline{\mathbf{S}}(r_1,\ldots,r_N)}\frac{\partial \ln f(\underline{\mathbf{Y}};\boldsymbol{\theta})}{\partial \underline{\mathbf{S}}(r_1,\ldots,r_N)}\Big].\end{aligned} \tag{7.21}$$

For the submatrices in the diagonal of the FIM in (7.6), we have the following expressions after simplifications for $n\in\{1,\ldots,N\}$,

$$\begin{aligned}\boldsymbol{\Omega}_{\mathbf{U}^{(n)}\mathbf{U}^{(n)}}\left(\overline{i_nr_n},\overline{i_nr_n}\right) &= \frac{1}{\sigma^2}\mathbf{e}_{r_n}^T\mathbf{S}_{(n)}\mathbf{S}_{(n)}^T\mathbf{e}_{r_n}\delta_{i_ni_n},\\ \boldsymbol{\Omega}_{\underline{\mathbf{S}}\underline{\mathbf{S}}}\left(\overline{r_1\ldots r_N},\overline{r_1\ldots r_N}\right) &= \frac{1}{\sigma^2}\delta_{r_1r_1}\ldots\delta_{r_nr_n}.\end{aligned} \tag{7.22}$$

For $n\in\{1,\ldots,N\}$, $k\in\{1,\ldots,N\}$, and $k>n$, we have the cross partial derivatives of the factor matrices in the sub-diagonals of the FIM in (7.6) with the elements

$$\begin{aligned}\boldsymbol{\Omega}_{\mathbf{U}^{(n)}\mathbf{U}^{(k)}}\left(\overline{i_nr_n},\overline{i_kr_k}\right) = \frac{1}{\sigma^2}\mathbf{e}_{r_n}^\top\mathbf{S}_{(n)}\left(\mathbf{U}^{(N)}\otimes\cdots\otimes\mathbf{U}^{(n+1)}\otimes\mathbf{U}^{(n-1)}\otimes\cdots\otimes\mathbf{U}^{(1)}\right)^\top\\ \mathbf{G}^{(n;i_n,k:i_k)}\left(\mathbf{U}^{(N)}\otimes\cdots\otimes\mathbf{U}^{(k+1)}\otimes\mathbf{U}^{(k-1)}\otimes\cdots\otimes\mathbf{U}^{(1)}\right)\mathbf{S}_{(k)}^\top\mathbf{e}_{r_k}.\end{aligned} \tag{7.23}$$

Additionally, we have the cross partial derivatives of factor matrices and the core tensor of the FIM in (7.6) for $n\in\{1,\ldots,N\}$ as

$$\begin{aligned}\boldsymbol{\Omega}_{\mathbf{U}^{(n)}\underline{\mathbf{S}}}\left(\overline{i_nr_n},\overline{r_1\ldots r_N}\right) = \frac{1}{\sigma^2}\mathbf{e}_{r_n}^\top\mathbf{S}_{(n)}\left(\mathbf{U}^{(N)}\otimes\cdots\otimes\mathbf{U}^{(n+1)}\otimes\mathbf{U}^{(n-1)}\otimes\cdots\otimes\mathbf{U}^{(1)}\right)^\top\\ \mathbf{G}^{(n;i_n)}\left(\mathbf{U}^{(N)}\otimes\cdots\otimes\mathbf{U}^{(1)}\right)\mathbf{e}_{\overline{r_1\ldots r_N}}.\end{aligned} \tag{7.24}$$

Finally, the partial derivatives of the constraints given in (7.12) with respect to the factor matrices have the columns described with

$$\mathbf{C}_{\mathbf{U}^{(n)}}\left(:,\overline{i_nr_n}\right) = \mathbf{E}^{(n)}\mathrm{vec}(\mathbf{U}^{(n)^T}\mathbf{e}_{i_n}\mathbf{e}_{r_n}^T+\mathbf{e}_{r_n}\mathbf{e}_{i_n}^T\mathbf{U}^{(n)}), \tag{7.25}$$

and the partial derivatives of the constraints given in (7.12) with respect to the mode-n unfoldings of the core tensor have the columns described with

$$\begin{aligned}\mathbf{C}_{\mathbf{S}_{(n)}}\left(:,\overline{r_1\dots r_{n-1}r_{n+1}\dots r_N}\right) = \mathbf{K}^{(n)}\mathrm{vec}(\mathbf{S}_{(n)}\,\mathbf{e}_{\overline{r_1\dots r_{n-1}r_{n+1}\dots r_N}}\mathbf{e}_{r_n}^T \\ + \mathbf{e}_{r_n}\mathbf{e}_{\overline{r_1\dots r_{n-1}r_{n+1}\dots r_N}}^T\mathbf{S}_{(n)}^T),\end{aligned} \tag{7.26}$$

for $n \in \{1,\dots,N\}$. The matrices $\mathbf{E}^{(n)}$, $\mathbf{K}^{(n)}$, $\mathbf{P}^{(\cdot)}$, and $\mathbf{G}^{(\cdot,\cdot)}$ are described in Section 7.2.

8

CONCLUSION AND FUTURE WORK

8.1. CONCLUSION

In this dissertation, we investigated denoising contrast-enhanced ultrasound sequences using low-rank tensor decomposition approaches, specifically those that incorporate the orthonormality of the factor matrices. We modeled the 4th-order DCEUS acquisitions as a low multilinear rank tensor or a low tubal rank tensor and we denoised speckles. We investigated two frameworks, namely the multilinear singular value decomposition and the t-SVD framework. Both of these low-rank definitions, along with the tensor-related mathematical relations, are described in Chapter 2. We continued the background information in Chapter 3, in which we gave a brief introduction to the formation of the dynamic contrast-enhanced ultrasound acquisitions and explained the markers of prostate cancer estimated by quantitative analysis of the acquisitions. These markers are used further in the chapter as a way to assess the performance of the denoising algorithms, with a special focus on the model fitting error before and after denoising. Furthermore, we used a simple linear classifier per feature (prostate cancer marker) to assess prostate cancer classification performance. The use of a simple classifier was intentionally chosen to obtain a better assessment of the denoising effects, which was independent of the classifier's behavior. To this end, we also kept the extraction of the features and preprocessing of DCEUS acquisitions the same as the previous contrast ultrasound dispersion imaging studies [55, 61, 69, 72, 84], and used a simple classifier to assess the distinguishability of the malignant and benign voxels.

This work was motivated by the SVD-based decomposition of the DCEUS acquisitions [93], where the data was flattened into a spatio-temporal matrix called the Casorati matrix. This is, in fact, a mode-4 unfolding of DCEUS data, assuming that the first three are the Cartesian domain and the last is the time axis. In previous work for the extraction of DCEUS features, the authors employed a standard SVD filter, where they kept the highest few singular values and the singular vectors and formed the denoised tensor. This was used as a filtering method that led to improved TIC fitting. We postulated that spatiotemporal low-rankness is a problem that must be tackled by considering all mode-n unfoldings of the tensor.

Therefore, we modeled the DCEUS as a low multilinear rank tensor. Additionally, DCEUS signals exhibit low-frequency content in space and time. We used the t-SVD framework to utilize this information and modeled DCEUS signals as a low tubal rank tensor. The success of the tensor despeckling algorithms lies in the match between how well the modeled problem fits the reality. For that reason, we started our research with a simulation-based chapter that investigated the speckle noise characteristics, along with the low-rank definitions that we specified.

The log-transformed Rayleigh distribution becomes a Fisher-Tippett distribution. In [92], the authors approximated this distribution by modeling it as WGN and sparse outliers. In Chapter 4, we investigated several optimization algorithms that incorporate tensor multilinear rankness and tubal rankness for denoising additive WGN with sparse outliers. We relaxed the constraint regarding the sparse outliers with the L1 norm, which is minimized through a proximity operator. The tensor's low multilinear rank and low tubal rank can be relaxed in several different ways. We explained these convex relaxation methods and the proximity operators in Chapter 2. In essence, it can be summarized into two different methods. The first one is the sum of the nuclear norm (SNN), and the other is the tensor nuclear norm (TNN). The former considers the low-rankness in the mode-n unfoldings. Iteratively, we can apply soft thresholding to the singular values of the SVD of the mode-n unfoldings. This proximity operator can be used to minimize the objective function related to the SNN. The second is the tensor nuclear norm, TNN, which is related to the t-SVD framework. Similarly, we can apply the t-SVD operation and apply soft-thresholding to the singular values tensor. Such a proximity operator can be used to minimize the objective function related to the TNN. In this framework, low tubal rankness can be described through the tensor's low-rank frontal slices. This results in an orientation-dependent low-rankness. We can permute the modes and reach a different tensor decomposition. In this fashion, we can consider various permutations and apply the proximity operator related to the TNN. We have investigated these optimization functions in Chapter 4 and found out that the orientation invariant nuclear norm method performs the best in low SNR scenarios. This method considers four different permutations of the modes of the original 4th-order DCEUS acquisitions. As the SNR increased, methods that were suitable for WGN performed well. Although a statistically significant improvement was observed by adding the sparsity constraint, the mean squared error difference was found to be minimal, especially with the added time cost. We only investigated these techniques through simulations. In the following Chapters 5 and 6, we continued with an in-vivo study.

We started with multilinear singular value decomposition that flattens the DCEUS acquisition in each mode and tackles each unfolding in Chapter 5. We explored an information-theoretic rank estimation method and found that the robust rank estimation techniques work better at estimating the ranks in a simulation study. According to the estimated ranks, we truncated the MLSVD of DCEUS acquisitions and recovered the denoised low multilinear rank signal that explains the TICs. The tr-MLSVD is unsuitable for sparse signal denoising, i.e., when the noise is additive and corrupted by outliers. MLSVD has a least squares cost function and is better suited for denoising WGN. Although we have incorporated a robust rank estimator, we have not incorporated such a technique into the optimization problem. In the following chapters, we investigated this.

There are a few ways to model the speckle noise. The most common model is a Rayleigh-distributed random variable that is multiplicative. This assumption holds well when there is a significant number of scatterers in one resolution cell. In cases where there is not a significant number of scatterers, different distributions, such as the Rician or Nakagami distribution, have been proposed for a better description of the speckle noise. In this dissertation, we assumed that there are enough scatterers, and the Rayleigh assumption of the speckle noise holds. We introduced two algorithms that incorporate the prior distribution into the tensor decomposition. This question initiated several research directions. The first one was to incorporate the negative log-likelihood into the optimization problem. If we add constraints regarding the estimated tensor being of low multilinear rank, we reach the general tensor estimation framework, which is explained in Chapter 6. Such an approach incorporates the prior distribution of the noise type into the low multilinear rank description. The problem is non-convex, and we found several ways to reach a converging result. One technique was to subtract the gain of the DCEUS acquisition, and the other was to use a varying step size, which was reduced whenever the optimization problem diverged. Here, we use the step size to describe the step that we take toward the negative of the gradient of the loss function in GTE introduced in Section 6.2.2. The other research direction was relaxing the optimization problem into a convex problem.

In Chapter 6, we compared two approaches of incorporating the prior distribution of the speckle noise into the low multilinear rank and low tubal rank tensor decompositions and extended the analysis to a large patient cohort. These approaches were GTE and OITNN, introduced in Section 6.2.2 and 6.2.3, respectively. We first ran a simulation in a tensor recovery problem with varying multilinear ranks that imitate the situations that can occur in DCEUS acquisitions. Namely, all modes were of low rank, modes had varying ranks, and all modes had a slightly higher rank. In the simulation, we found out that most of the time, OITNN resulted in the best denoising performance, followed by GTE and, lastly, tr-MLSVD. Adding a prior distribution or its convex relaxation to the tensor recovery problem had a statistically significant benefit. We followed this analysis with a phantom study and reported the model fitting root mean squared error after denoising by various algorithms in the literature and the algorithms that we introduced. We found out that both GTE and OITNN caused a better suppression of speckle noise around the background of the phantom. This was visualized through a figure that included the early appearance time, the peak time, and the wash-out time of a phantom recording. In all the cases, only GTE and OITNN were able to suppress the speckle artifacts around the phantom. Furthermore, the one that gave the best model-fitting performance was OITNN. No statistically significant result was observed between tr-MLSVD and GTE. Finally, we used an in-vivo study introduced in Chapter 3. Dynamic contrast-enhanced ultrasound recordings with varying degrees of prostate cancer were included. Most of the features benefited from using a tensor-based denoising method. Three out of nine features performed the best for OITNN, five out of nine performed the best for tr-MLSVD, and one performed the best for a non-tensor-based technique. In conclusion, tr-MLSVD with the highest ROC-AUC metric among 94 patients is the algorithm that improved the prostate cancer diagnosis in the most efficient way in terms of fastest calculation time and maximum compression power. This is shown in Table 8.1. In this table, we summarize the results with regard to three

algorithms that we proposed, namely tr-MLSVD, GTE, and OITNN. We include the NME from the simulation section, MFR-RMSE from the in-vitro section, the calculation time, memory after compression, computational complexity of the algorithms, and the average ROC-AUC metric from the in-vivo section of Chapter 6. The average ROC-AUC metric is calculated by taking the mean of the ROC-AUC of nine CUDI features from Section 6.4.3 for each method. Average NME from the simulation section 6.4.1 is calculated similarly, where the mean over three tensors is calculated. Tr-MLSVD compressed the approximately 80 GB patient data into 840 MB, while increasing the average ROC-AUC compared to the methods proposed in the literature, and GTE and OITNN. Furthermore, tr-MLSVD required the least amount of computational time and complexity compared to GTE and OITNN. On the other hand, we observed that OITNN had the least NME in the simulation with a mean value of 0.0287, compared to 0.0376 of GTE and 0.0414 of tr-MLSVD. In addition, a small improvement in model fitting performance is observed for OITNN with an MFR-RMSE value of 4.5 compared to 4.8 of tr-MLSVD and GTE.

In Chapter 7, we examined the denoising performance of low multilinear rank tensor decomposition methods for DCEUS data through the lens of statistical efficiency. By deriving the constrained Cramér-Rao bound (CCRB) under additive white Gaussian noise, we established an asymptotic lower bound on the mean squared error (MSE) of unbiased estimates of the MLSVD components. The obtained theoretical insights are applicable to despeckling, although we did not incorporate the Fisher-Tippett distribution explained in Chapter 3. Simulations demonstrated that in high-SNR regimes and well-conditioned scenarios, characterized by distinct mode-n singular values, high multilinear ranks and orders, the MSE of practical algorithms such as HOOI and tr-MLSVD approached the CCRB, confirming their asymptotic efficiency. However, in low-SNR or ill-conditioned cases, these estimators deviated from the CCRB due to violations of unbiasedness and all-orthogonality constraints, as well as the non-uniqueness of the modal singular vectors. The tr-MLSVD and HOOI performed better than CCRB in these cases. Our findings indicate that the asymptotic lower bound on the variance of the components that make up a multilinear rank tensor can only be used in ideal conditions. In such cases, the potential improvement in MSE that an unbiased estimator can offer for estimating the low multilinear rank tensor components is minimal.

Table 8.1: Summary of quantitative and computational results.

Method	Simulation NME (avg $\pm$ std)	In-Vitro MFR-RMSE	Time	Memory	In-Vivo Computational Complexity	ROC-AUC
tr-MLSVD	0.0414 ± 0.0008	4.8 ± 2.7	8.3 h	840 MB	$\mathcal{O}\left(D\sum_{n=1}^{4} I_n\right)$	0.8117
GTE	0.0376 ± 0.0008	4.8 ± 2.8	17 h	840 MB	$\mathcal{O}\left(KD\max\{I_1,\ldots,I_4\}\right)$	0.7911
OITNN	0.0287 ± 0.0032	4.5 ± 2.8	42 h	24.94 GB	$\mathcal{O}\left(K(D\log D + D\sum_{n=1}^{4} I_n)\right)$	0.7946

8.2. ANSWERS TO THE RESEARCH QUESTIONS

Let us revisit and answer the research questions that we posed in Chapter 1.

Q1: Under which signal-to-noise ratios can the speckle noise be modeled as white Gaussian noise with sparse outliers?

The speckle noise is modeled as Rayleigh distributed only when there are a *high* number of scatterers in one resolution cell. This statement in itself requires a fair amount of research, i.e., for example, what does *high* mean? Is it the same when there are strong reflectors in one resolution cell and a high number of scatterers? What happens in such a case? There is, in fact, a vast amount of research that answers these questions. In [170], the conditions in which the speckle noise is multiplicative are investigated. The authors apply a fundamental mode imaging scenario and claim that the multiplicative model fails when the object contains fine details that cannot be resolved by the imaging system. Chapter 7 of the PhD dissertation [86] describes various cases regarding speckle noise. The authors analyze probability distributions in various cases: sums of speckle patterns, speckle patterns plus a non-random phasor, and non-uniform phase distributions. In all three cases, we have different probability distributions. In this dissertation, we assumed a fully developed speckle pattern and assumed that the speckle noise follows a Rayleigh distribution. We aim to answer this research question by further constraining the domain of interest. Since this dissertation aimed to apply tensor decomposition-based denoising techniques, we answer the research question within this scope.

We simulated speckle denoising scenarios in Chapter 4 for varying SNRs and applied multiple optimization algorithms that incorporate the tubal low-rankness and the multilinear low-rankness. A range of values was swept for both regularization parameters that control the low-rankness and the sparsity. We found out that the addition of a sparsity constraint is significantly more critical for low SNR scenarios, namely for SNRs less than 20 dB. OITNN, which utilizes various permutations of the noisy tensor for denoising, gave the best-performing results in simulations for low SNRs. In addition, in the phantom study, OITNN gave the best model-fitting performance. In-vivo results show similarity when it comes to model fitting results. OITNN performed the best for fitting the LDRW model as shown in Chapter 6. Furthermore, three cancer markers out of nine showed an improved performance for OITNN, while five out of nine cancer markers showed a better performance for tr-MLSVD. This suggests that the addition of sparsity into the optimization problem helped some of the cancer markers, while it degraded the performance of others. Given that tr-MLSVD gave the best performance for the majority of the cancer markers and that OITNN converges in 42 hours compared to 8 hours for calculating the tr-MLSVD, we can conclude that using tr-MLSVD is a better approach for denoising DCEUS acquisitions.

Another important aspect is the rank estimation. In Chapter 5, we estimated the ranks using the SCORE [109] algorithm, which is tailored towards WGN with sparse outliers. Although the tr-MLSVD does not incorporate a sparsity-related element during optimization, the estimation of the ranks does. Once Akaike's information criteria or minimum description length were used instead of SCORE, higher ranks were estimated, which caused degraded denoising performance. Therefore, the addition of sparsity is important for the rank estimation of DCEUS acquisitions.

Q2: Can we improve the prostate cancer classification performance using tensor-based denoising techniques applied to DCEUS?

The t-SVD framework has an Eckhart-Young-like result using the new t-product definition that we explained in Chapter 2. The truncation of the t-SVD decomposition yields the best low-rank 3rd-order tensor on the Frobenius norm of the

residual error. This is proven in [8]. However, the DCEUS acquisitions are 4th-order. The extension of the t-SVD framework to 4th-order, as proposed in [9], does not have the same proof. Therefore, to denoise DCEUS acquisitions, we came up with several questions. What are the ranks that will be used for truncation? Which orientation yields the best performance? We defined an optimization problem that included all the possible permutations that gave different t-SVD decompositions and soft thresholded the tubal singular values. In this fashion, we created a result that was independent of the orientation of the modes. Additionally, we did not estimate the ranks. We created a regularization parameter that soft-thresholded the singular values until the convergence was reached according to the stopping criteria. With these techniques, we found in simulations and in-vitro that the best performance with respect to a mean squared error was t-SVD-related low-rankness, i.e., low tubal rankness. For the simulation, we had a ground truth. Therefore, we used normalized error criteria for comparison. For the in-vitro experiments, we did not have a ground truth. For that reason, we used the model fitting performance as a criterion.

OITNN, which is based on the t-SVD framework, gave the best model fitting performance in terms of root mean squared error. For prostate cancer classification, the improved performance of OITNN was not seen for six out of the nine prostate markers. Mostly, the tr-MLSVD gave the best-performing result. In such a setting, we can claim that the low multilinear rank methods give the best classification performance.

We want to discuss a few limitations of the study. Firstly, we used LOGIQ devices in all our experiments. Due to the proprietary restrictions, we do not have full control and knowledge over what happens before the image formation chain. There are several preprocessing techniques, such as pulse inversion and coded imaging, that are explained in Chapter 3. All the in-vivo results are tailored towards this specific equipment and the firmware that is being used. We did not conduct a study that compared different equipment. In such a case, the results might change. Additionally, the rank estimation that we have analyzed is sensitive to the number of data points. This can be seen in the difference between the rank estimation-related hyperparameter difference between Chapter 5 and 6. In Chapter 5, the temporal frequency is 0.25 Hz, whereas in Chapter 6, the temporal frequency is 1 Hz. The data size is increased by a factor of 4, and the rank estimation-related hyperparameter is changed by a factor of 10. Such a tuning needs to be done on a subset of data, which causes a problem in answering this question. Within the scope of this dissertation, we claim that low multilinear rankness is better suited for denoising DCEUS acquisitions for a better prostate cancer classification.

Q3: Can a lower bound on the variance of locally unbiased estimators be derived in the context of low multilinear rank tensor estimation?

In Chapter 7, we calculated a lower bound for estimating the components of the MLSVD using a constrained Cramér-Rao bound for denoising WGN. We incorporated the constraints regarding the all-orthogonality of the core tensor and the orthonormality of the factor matrices. We assumed the mode-n singular values were distinct to achieve a unique MLSVD decomposition. As the mode-n singular values get close to each other, the constrained Cramér-Rao bound becomes degenerate. This property was observed through the condition number of the constrained FIM given in (7.13). The estimators, such as tr-MLSVD, did not suf-

fer when the modal singular values were not unique. The MSE of the estimates was found to be lower than the CCRB, suggesting a bias in the estimation. For high multilinear ranks, high orders, and well-separated mode-n singular values, the CCRB is a tight bound for common low multilinear rank approximation techniques such as tr-MLSVD and HOOI. The performance of tr-MLSVD is satisfactory, and the additional benefit of methods such as [18] is expected to be minimal. In DCEUS acquisitions, the mode-n singular values can be arbitrarily close. The CCRB is limited to cases where the mode-n singular values are well separated, which is often not guaranteed in real data. We concluded that the CCRB should only be used to compare the low multilinear rank approximation of DCEUS when the mode-n singular values were distinct. On the contrary, the performance of low multilinear rank tensor approximation algorithms did not degrade when the mode-n singular values were close. For close modal singular values, the singular vectors were not unique, although the corresponding subspaces were still well defined (the factor matrix $\mathbf{U}^{(n)}$ contains an arbitrary basis for the subspace in mode-n). The FIM was based on a unique parametrization hence CCRB became problematic even if the algorithms were fine. Therefore, the proposed low multilinear rank and low tubal rank tensor decompositions could be used for denoising DCEUS independent of how the mode-n singular values of the DCEUS tensor were distributed.

The answer to the third research question is that an asymptotic lower bound on the variance of the estimates that make up a low multilinear rank tensor can be calculated through CCRB only if the modal singular values are well-separated and the tensor has high multilinear ranks and orders. Since we do not have control over the modal singular values of a DCEUS acquisition, CCRB can not be used in general to model its denoising performance.

8.3. Future Work

To conclude the dissertation, we propose several directions for future work.

8.3.1. Low-rankness for 4th-order tensors

In an informal way, the 3rd-order t-SVD framework considers low-rankness in the frequency content of the signals in the 3rd mode, while incorporating correlations in frontal slices. In such a fashion, the t-product that is defined in Chapter 2 is the matrix multiplication between the respective frontal slices of two 3rd-order tensors, where a 1D Fourier transform is applied to the 3rd mode of the tensors. The reasoning behind choosing the Fourier transform is explained in [8]. In [10], it is shown that any invertible transform can be applied to the 3rd mode. For example, if the invertible transform is an identity matrix, we will have the matrix multiplication of frontal slices. However, the selection of such a matrix does not produce a sparse core tensor, which, in the end, does not produce a low-rank description. Therefore, the authors in [8] select the Fourier transform matrix, whereas the authors in [10] select a data-dependent unitary matrix as $\mathbf{F}^{(3)}$ that is obtained by the application of SVD to the mode-3 unfolding of the tensor and selecting the left singular vectors. The selection of this matrix, again, results in a sparse core tensor that can be further processed for a low-rank description. The benefit of selecting such a matrix as the transform that is applied to the

3rd mode is the data dependency. Once the Fourier transform is selected, we assume that a sparse core tensor will be generated. In images and videos, this is a valid assumption. Many real datasets are band-limited, i.e., they will have more power in a range of frequencies, which can explain the majority of the data. On the other hand, the selection of the unitary matrix forces the sparsity of the core tensor due to the property of the SVD. The application of this approach to DCEUS acquisitions is left for future work.

The t-SVD framework has optimal low-tubal rank approximation guarantees for a 3rd-order tensor. This works well in applications such as video, where the first two modes represent the 2nd-order images, and the 3rd mode is time. However, a question remains regarding the DCEUS case. What happens if the first three modes are the Cartesian domain, and the fourth mode is time? This is an open question that has not been answered yet. It is an interesting research direction and would be the proper way to approach the low-rank approximation of DCEUS acquisitions. The strong suit of the tensor decompositions is that they are able to keep the original format of the data intact. However, this statement is not true in most cases. In multilinear singular value decomposition, the tensor is unfolded into matrices, and SVD is applied. The variance in each mode is considered through flattening the tensor. For this reason, the t-SVD framework provides an interesting direction that aims to create multilinear algebraic descriptions that define operations on tensors. The definition of an algebraic group for 4th-order tensors is an interesting future work. We believe such an approach would improve the denoising of DCEUS acquisitions.

8.3.2. TRUNCATION BASED ON TUBAL LOW-RANK

In this dissertation, we have observed that the truncated multilinear singular value decomposition performs the best in denoising the DCEUS acquisitions. Since tr-MLSVD has a closed-form solution, denoising can be performed quickly. Four SVD operations of a 4th-order tensor are required. In this dissertation, we have not considered the low tubal rank approximation based on truncated t-SVD defined in Chapter 2 because in Chapter 4, we have found out that OITNN performed better. However, OITNN is an iterative algorithm and requires more computation time compared to calculating the closed-form truncated t-SVD. We believe this is an interesting research direction for denoising DCEUS acquisitions. Two questions need to be answered. The first one is which ranks should be used. Similar to the tr-MLSVD case, where the ranks are estimated by the SCORE algorithm, we can extend the analysis to t-SVD. We can order the singular values and define the ranks by rank estimation algorithms such as information-theoretic methods like Akaike's information criteria [171] or Minimum description length [127], or heuristic methods such as the Elbow method [172]. It might be the case that, similar to tr-MLSVD, the estimated ranks might be high, or the data is not suitable for such methods. Then, a rank estimation method that extends the aforementioned methods to t-SVD can be considered. An example might be the extension of the SCORE algorithm to t-SVD singular values.

To achieve the low tubal rank approximation, an orientation needs to be fixed. Since t-SVD results in different decompositions for different orientations, we need to consider which permutation of the modes works the best. The distribution of the singular values, along with the denoising capabilities, should be tested. The

resulting denoised signals can be transformed back to the original domain by reversing the permutation.

In preparing Chapter 6, we investigated different tensorizations of the 4th-order DCEUS acquisitions. Given that the t-SVD is always applied to the 3rd-order, we can define a spatially invariant t-SVD framework. A nuclear norm with the tensorizations, $\{i_1, \overline{i_2 i_3}, i_4\}$, $\{i_2, \overline{i_1 i_3}, i_4\}$, $\{i_3, \overline{i_1 i_2}, i_4\}$ can be used. Here, $\{i_1, \overline{i_2 i_3}, i_4\}$ is used to denote the grouping of the 2nd-order and the 3rd-order of the DCEUS acquisitions, and keeping the 1st-order and the 4th-order. Similarly, the other groupings, $\{i_2, \overline{i_1 i_3}, i_4\}$ and $\{i_3, \overline{i_1 i_2}, i_4\}$, can be described. The orientation invariant tensor nuclear norm algorithm given in Chapter 4 can be applied to the aforementioned tensorizations. This motivates the separation of time and space, and consistently applies a 1D Fourier transform to the time mode. We expect the frequency content of the UCA movement to be present in low frequencies [74]. Therefore, the 1D Fourier transform applied in the time mode is expected to sparsify the core tensor, and possibly provide a good low tubal rank approximation. In our analysis, we observed smoother time evolutions. However, the prostate cancer classification performance did not improve. The reason might be due to the limited grid search that we applied with the sparsity and low tubal rankness parameter γ_L and γ_S, and the limited number of patients. We leave the low tubal rank approximation that applies the Fourier transform to only time, and the investigation of various tensorizations of the Cartesian domain, as future work.

8.3.3. MODELLING THE RECIRCULATION

In DCEUS applications, commonly, the first pass is fitted [74][58][55] with a pharmacokinetic model, such as the LDRW model. The rest of the information regarding the TICs that correspond to the recirculation is then discarded. This is done because the first pass of the bubbles gives a better representation of perfusion dynamics. The recirculation information is problematic due to the different passages the bubbles might take before reaching the prostate and because of the lower SNR. We propose that the recirculation of the bubbles can be formulated in a harmonic retrieval problem.

The local density random walk model assumes an infinitely long tube where the bubbles pass the detection site only once. One direction to model the task would be to model the circulatory system as a straight tube that is extended from a circular tube. Consider a scenario where we intend to model from the first pass until the fourth pass of a detection site. We can model this by imagining a tube with the detection sites layered at three different distances that represent the first, second, and third passes. We can model each of these with three local density random walks, with different times and distances, but the same initial bubble concentration. The summation of these three models might result in a grouping of expressions that involve a harmonic. This can be decomposed by the multilinear extension of the ESPRIT method [173] that is proposed in [174].

In another direction, the circular Brownian motion models [175] can be incorporated into the model-fitting of TICs. Previously, only the first pass of the microbubbles was modeled. If the intention is to use harmonic retrieval modeling, the circulation should be incorporated to obtain a periodic description. In another direction, a Kronecker structure such as the one proposed in [176] could

be used, where the UCA evolution is modeled as a multiple-input multiple-output system. Instead of modeling the path UCAs take as an infinitely long tube, we can model it as the summation of multiple systems that can possibly interact with each other.

There has been some previous research [177][178][179] that investigated the idea of multiple passes of TICs. The second wave is not necessarily correlated with the recirculation of the UCAs; it is also correlated with bubbles that arrive from another feeding artery or a different loop through the systemic circulation. In addition, UCAs tend to disappear after multiple passes because of the cardiac pressure and the transpulmonary circulation. It is shown in [174] [180] that separating different sources in low SNR scenarios is possible. The added benefit is the separation of the sources and the extraction of meaningful information regarding malignant and benign TICs.

ACRONYMS

CCRB Constrained Cramer-Rao Bound
CD Convective dispersion
CEv Conditional entropy
CRB Cramer-Rao Bound
CUDI Contrast ultrasound dispersion imaging
DCEUS Dynamic Contrast-Enhanced Ultrasound
GTE General tensor estimation
HOOI Higher order orthogonal iterations
LDRW local density random walk
MFR Model fitter
mLDRW Modified local density random walk
MLSVD Multilinear singular value decomposition
OITNN Orientation invariant tubal nuclear norm
PCa Prostate cancer
SA Similarity analysis
SNN Sum of nuclear norm
SNR Signal to noise ratio
sPCa Significant prostate cancer
SVD Singular value decomposition
t-SVD Tensor singular value decomposition
TIC Time intensity curve
TNN Tubal nuclear norm
tr-MLSVD Truncated multilinear singular value decomposition
UCA Ultrasound contrast agent
VE Velocity entropy

BIBLIOGRAPHY

[1] M. Calis, M. Mischi, A. van der Veen, and B. Hunyadi. “Denoising of dynamic contrast-enhanced ultrasound sequences: a multilinear approach.” In: *BIOSIGNALS*. 2022, pp. 192–199.

[2] P. Chen, M. Calis, H. Wijkstra, P. Huang, B. Hunyadi, and M. Mischi. “Multiparametric ultrasound and machine learning for prostate cancer localization”. In: *European Signal Processing Conference*. 2022, pp. 907–911.

[3] M. Calis and B. Hunyadi. “Denoising of the speckle noise by robust low-rank tensor decomposition”. In: *2024 32nd European Signal Processing Conference (EUSIPCO)*. 2024.

[4] P. Comon. “Tensors: a brief introduction”. In: *IEEE Signal Processing Magazine* 31.3 (2014), pp. 44–53.

[5] I. Domanov and L. De Lathauwer. “On the uniqueness of the canonical polyadic decomposition of third-order tensors. Part I: basic results and uniqueness of one factor matrix”. In: *SIAM Journal on Matrix Analysis and Applications* 34.3 (2013), pp. 855–875.

[6] A. Cichocki, N. Lee, I. V. Oseledets, A.-H. Phan, Q. Zhao, and D. Mandic. “Low-rank tensor networks for dimensionality reduction and large-scale optimization problems: Perspectives and challenges part 1”. In: *arXiv preprint arXiv:1609.00893* (2016).

[7] A. Cichocki, A.-H. Phan, Q. Zhao, N. Lee, I. Oseledets, M. Sugiyama, D. P. Mandic, *et al.* “Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives”. In: *Foundations and Trends® in Machine Learning* 9.6 (2017), pp. 431–673.

[8] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover. “Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging”. In: *SIAM Journal on Matrix Analysis and Applications* 34.1 (2013), pp. 148–172.

[9] C. D. Martin, R. Shafer, and B. LaRue. “An order-p tensor factorization with applications in imaging”. In: *SIAM Journal on Scientific Computing* 35.1 (2013), A474–A490.

[10] C. Zeng and M. K. Ng. “Decompositions of third-order tensors: HOSVD, T-SVD, and beyond”. In: *Numerical Linear Algebra with Applications* 27.3 (2020), e2290.

[11] A. K. Jain. *Fundamentals of digital image processing*. USA: Prentice-Hall, Inc., 1989. isbn: 0133361659.

[12] M. E. Kilmer and C. D. Martin. “Factorization strategies for third-order tensors”. In: *Linear Algebra and its Applications* 435.3 (2011), pp. 641–658.

[13] L. De Lathauwer, B. De Moor, and J. Vandewalle. “A multilinear singular value decomposition”. In: *SIAM journal on Matrix Analysis and Applications* 21.4 (2000), pp. 1253–1278.

[14] G. H. Golub and C. F. Van Loan. *Matrix computations*. JHU press, 2013.

[15] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan. “Tensor decompositions for signal processing applications: from two-way to multiway component analysis”. In: *IEEE signal processing magazine* 32.2 (2015), pp. 145–163.

[16] F. Van Eeghem, O. Debals, and L. De Lathauwer. “Tensor similarity in two modes”. In: *IEEE Transactions on Signal Processing* 66.5 (2017), pp. 1273–1285.

[17] W. Hackbusch. *Tensor spaces and numerical tensor calculus*. Vol. 42. Springer, 2012.

[18] M. Ishteva, P.-A. Absil, S. Van Huffel, and L. De Lathauwer. “Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme”. In: *SIAM Journal on Matrix Analysis and Applications* 32.1 (2011), pp. 115–135.

[19] L. De Lathauwer, B. De Moor, and J. Vandewalle. “On the best rank-1 and rank-($r_1, r_2, \ldots, r_n$) approximation of higher-order tensors”. In: *SIAM journal on Matrix Analysis and Applications* 21.4 (2000), pp. 1324–1342.

[20] M. Ishteva. “Numerical methods for the best low multilinear rank approximation of higher-order tensors”. PhD thesis. Kasteelpark Arenberg 1/2200, B-3001 Leuven (Belgium): Katholieke Universiteit Leuven, Dec. 2009.

[21] C. Eckart and G. Young. “The approximation of one matrix by another of lower rank”. In: *Psychometrika* 1.3 (1936), pp. 211–218.

[22] M. Kilmer, L. Horesh, H. Avron, and E. Newman. “Tensor-tensor products for optimal representation and compression”. In: *arXiv preprint arXiv:2001.00046* (2019).

[23] G. Ely, S. Aeron, N. Hao, and M. E. Kilmer. “5D seismic data completion and denoising using a novel class of tensor decompositions”. In: *Geophysics* 80.4 (2015), pp. V83–V95.

[24] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. “Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization”. In: *Advances in neural information processing systems* 22 (2009).

[25] N. Parikh, S. Boyd, *et al.* “Proximal algorithms”. In: *Foundations and trends® in Optimization* 1.3 (2014), pp. 127–239.

[26] D. Goldfarb and Z. Qin. “Robust low-rank tensor recovery: models and algorithms”. In: *SIAM Journal on Matrix Analysis and Applications* 35.1 (Jan. 2014), pp. 225–253.

[27] S. Gandy, B. Recht, and I. Yamada. “Tensor completion and low-n-rank tensor recovery via convex optimization”. In: *Inverse problems* 27.2 (2011), p. 025010.

[28] J. Eckstein and D. P. Bertsekas. “On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators”. In: *Mathematical programming* 55 (1992), pp. 293–318.

[29] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[30] A. Wang, Q. Zhao, Z. Jin, C. Li, and G. Zhou. “Robust tensor decomposition via orientation invariant tubal nuclear norms”. In: *Science China Technological Sciences* 65.6 (2022), pp. 1300–1317.

[31] A. Wang, Z. Jin, and G. Tang. “Robust tensor decomposition via t-SVD: near-optimal statistical guarantee and scalable algorithms”. In: *Signal Processing* 167 (2020), p. 107319. issn: 0165-1684.

[32] J. Curie and P. Curie. “Développement par compression de l’électricité polaire dans les cristaux hémièdres à faces inclinées”. In: *Bulletin de Minéralogie* 3.4 (1880), pp. 90–93.

[33] J. W. Strutt and J. W. S. Rayleigh. *The theory of sound*. Vol. 1. Macmillan, 1877.

[34] P. G. Newman and G. S. Rozycki. “The history of ultrasound”. In: *Surgical Clinics of North America* 78.2 (1998), pp. 179–195. issn: 0039-6109.

[35] J. J. Wild and J. M. Reid. “Further pilot echographic studies on the histologic structure of tumors of the living intact human breast”. In: *The American journal of pathology* 28.5 (1952), p. 839.

[36] J. Lohrke, T. Frenzel, J. Endrikat, F. C. Alves, T. M. Grist, M. Law, J. M. Lee, T. Leiner, K.-C. Li, K. Nikolaou, *et al.* “25 years of contrast-enhanced MRI: developments, current challenges and future perspectives”. In: *Advances in therapy* 33 (2016), pp. 1–28.

[37] M. A. Averkiou, M. F. Bruce, J. E. Powers, P. S. Sheeran, and P. N. Burns. “Imaging methods for ultrasound contrast agents”. In: *Ultrasound in medicine & biology* 46.3 (2020), pp. 498–517.

[38] B. A. Schrope and V. L. Newhouse. “Second harmonic ultrasonic blood perfusion measurement”. In: *Ultrasound in Medicine Biology* 19.7 (1993), pp. 567–579. issn: 0301-5629.

[39] F. Forsberg, W. T. Shi, and B. Goldberg. “Subharmonic imaging of contrast agents”. In: *Ultrasonics* 38.1-8 (2000), pp. 93–98.

[40] P. Frinking, A. Bouakaz, J. Kirkhorn, F. Cate, and N. de Jong. “Ultrasound contrast imaging: current and new potential methods”. In: *Ultrasound in Medicine & Biology* 26.6 (2000), pp. 965–975. issn: 0301-5629.

[41] M. A. Averkiou, C. Mannaris, M. Bruce, and J. Powers. “Nonlinear pulsing schemes for the detection of ultrasound contrast agents”. In: *Journal of the Acoustical Society of America* 123.5 (2008), p. 3110.

[42] D. H. Simpson, C. T. Chin, and P. N. Burns. “Pulse inversion Doppler: a new method for detecting nonlinear echoes from microbubble contrast agents”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 46.2 (1999), pp. 372–382.

[43] G. A. Brock-Fisher, M. D. Poland, and P. G. Rafter. *Means for increasing sensitivity in non-linear ultrasound imaging systems*. US Patent 5,577,505. Nov. 1996.

[44] P. Phillips. “Contrast pulse sequences (CPS): imaging nonlinear microbubbles”. In: *2001 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No. 01CH37263)*. Vol. 2. IEEE. 2001, pp. 1739–1745.

[45] A. G. J. Besson. “Imaging from echoes: on inverse problems in ultrasound”. In: (2019), p. 234.

[46] A. R. Sureshkumar. “Investigation into the origin and nature of variability in quantitative measurements of tumour blood flow with contrast-enhanced ultrasound”. PhD thesis. University of Toronto, 2012.

[47] R. L. Siegel, A. N. Giaquinto, and A. Jemal. “Cancer statistics, 2024.” In: *CA: a cancer journal for clinicians* 74.1 (2024).

[48] J. Folkman, K. Watson, D. Ingber, and D. Hanahan. “Induction of angiogenesis during the transition from hyperplasia to neoplasia”. In: *Nature* 339.6219 (1989), pp. 58–61.

[49] M. K. Brawer. “Quantitative microvessel density: a staging and prognostic marker for human prostatic carcinoma”. In: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 78.2 (1996), pp. 345–349.

[50] M. Borre, B. V. Offersen, B. Nerstrøm, and J. Overgaard. “Microvessel density predicts survival in prostate cancer patients subjected to watchful waiting”. In: *British journal of cancer* 78.7 (1998), pp. 940–944.

[51] N. Weidner, P. Carroll, J. Flax, W. Blumenfeld, and J. Folkman. “Tumor angiogenesis correlates with metastasis in invasive prostate carcinoma.” In: *The American journal of pathology* 143.2 (1993), p. 401.

[52] J. Folkman. “Angiogenesis”. In: *Biology of endothelial cells* (1984), pp. 412–428.

[53] R. J. Eckersley, J. M. Sedelaar, M. J. Blomley, H. Wijkstra, N. M. deSouza, D. O. Cosgrove, and J. J. de la Rosette. “Quantitative microbubble enhanced transrectal ultrasound as a tool for monitoring hormonal treatment of prostate carcinoma”. In: *The Prostate* 51.4 (2002), pp. 256–267.

[54] N. Elie, A. Kaliski, P. Péronneau, P. Opolon, A. Roche, and N. Lassau. “Methodology for quantifying interactions between perfusion evaluated by DCE-US and hypoxia throughout tumor growth”. In: *Ultrasound in medicine & biology* 33.4 (2007), pp. 549–560.

[55] M. Kuenen, M. Mischi, and H. Wijkstra. “Contrast-ultrasound diffusion imaging for localization of prostate cancer”. In: *IEEE Transactions on Medical Imaging* 30.8 (2011), pp. 1493–1502. issn: 0278-0062.

[56] M. P. J. Kuenen, A. T. Saidov, H. Wijkstra, J. M. C. H. J. de la Rosette, and M. Mischi. “Correspondence-spatiotemporal correlation of ultrasound contrast agent dilution curves for angiogenesis localization by dispersion imaging”. In: *IEEE Trans. on Ultrasonics, Ferroelectrics, and Freq. Control* 60.12 (2013), pp. 2665–2669.

[57] M. Kuenen, T. Saidov, H. Wijkstra, and M. Mischi. “Contrast-ultrasound dispersion imaging for prostate cancer localization by improved spatiotemporal similarity analysis”. In: *Ultrasound in medicine & biology* 39.9 (2013), pp. 1631–1641.

[58] M. P. J. Kuenen, H. F. I. Herold, H. M. H. Korsten, J. M. C. H. J. de la Rosette, H. Wijkstra, and M. Mischi. “Maximum-likelihood estimation for indicator dilution analysis”. In: *IEEE Transactions on Biomedical Engineering* 61.3 (2013), pp. 821–831.

[59] M. Kuenen. “Contrast-ultrasound dispersion imaging for prostate cancer localization”. pp. 35. PhD thesis. Electrical Engineering, 2014. isbn: 978-90-386-3568-2.

[60] M. Kuenen. “Contrast-ultrasound dispersion imaging for prostate cancer localization”. PhD thesis. Technische Universiteit Eindhoven, 2014, pp. 30–31.

[61] M. Mischi, M. P. Kuenen, and H. Wijkstra. “Angiogenesis imaging by spatiotemporal analysis of ultrasound contrast agent dispersion kinetics”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 59.4 (2012), pp. 621–629.

[62] M. Mischi, N. Bouhouch, L. Demi, M. P. Kuenen, A. Postema, J. J. De La Rosette, T. J. Tjalkens, and H. Wijkstra. “Contrast-ultrasound dispersion imaging of cancer neovascularization by mutual-information analysis”. In: *2014 IEEE International Ultrasonics Symposium*. IEEE. 2014, pp. 1148–1151.

[63] S. G. Schalk, L. Demi, M. Smeenge, J. J. M. C. H. de la Rosette, P. Huang, H. Wijkstra, and M. Mischi. “3D contrast ultrasound dispersion imaging by mutual information for prostate cancer localization”. In: *2015 IEEE International Ultrasonics Symposium (IUS)*. 2015, pp. 1–4.

[64] S. G. Schalk, L. Demi, M. Smeenge, D. M. Mills, K. D. Wallace, J. J. M. C. H. de la Rosette, H. Wijkstra, and M. Mischi. “4-D spatiotemporal analysis of ultrasound contrast agent dispersion for prostate cancer localization: a feasibility study”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 62.5 (2015), pp. 839–851.

[65] S. Schalk et al. “Contrast-enhanced ultrasound angiogenesis imaging by mutual information analysis for prostate cancer localization”. In: *IEEE Trans. on Biomedical Eng.* 64.3 (2016), pp. 661–670.

[66] S. Schalk. “Towards 3D prostate cancer localization by contrast-ultrasound dispersion imaging”. PhD thesis. Technical University of Eindhoven, 2017.

[67] R. R. Wildeboer, R. J. van Sloun, P. Huang, H. Wijkstra, and M. Mischi. “3D multi-parametric contrast-enhanced ultrasound for the prediction of prostate cancer”. In: *Ultrasound in medicine & biology* 45.10 (2019), pp. 2713–2724.

[68] R. Wildeboer, R. Van Sloun, C. Mannaerts, J. Van der Linden, P. Huang, H. Wijkstra, and M. Mischi. “Probabilistic 3D contrast-ultrasound tractography based on a convective-dispersion finite-element scheme”. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. 2018, pp. 1–9.

[69] R. R. Wildeboer, R. J. G. Van Sloun, S. G. Schalk, C. K. Mannaerts, J. C. Van Der Linden, P. Huang, H. Wijkstra, and M. Mischi. “Convective-Dispersion Modeling in 3D Contrast-Ultrasound Imaging for the Localization of Prostate Cancer”. In: *IEEE Transactions on Medical Imaging* 37.12 (2018), pp. 2593–2602.

[70] C. W. Sheppard. “Basic principles of the tracer method: introduction to mathematical tracer kinetics”. In: *(No Title)* (1962).

[71] K. Norwich and S. Zelin. “The dispersion of indicator in the cardiopulmonary system”. In: *The bulletin of mathematical biophysics* 32 (1970), pp. 25–43.

[72] R. J. van Sloun, L. Demi, A. W. Postema, J. J. de la Rosette, H. Wijkstra, and M. Mischi. “Ultrasound-contrast-agent dispersion and velocity imaging for prostate cancer localization”. In: *Medical Image Analysis* 35 (2017), pp. 610–619. issn: 1361-8415.

[73] J. Bogaard, S. Smith, A. Versprille, M. Wise, and F. Hagemeijer. “Physiological interpretation of the skewness of indicatordilution curves; theoretical considerations and a practical application”. In: *Basic research in cardiology* 79.4 (1984), pp. 479–493.

[74] M. Mischi, A. Kalker, and H. Korsten. “Videodensitometric methods for cardiac output measurements”. In: *EURASIP Journal on Applied Signal Processing* 2003.5 (2003), pp. 479–489. issn: 1110-8657.

[75] R. W. Stow and P. S. Hetzel. “An empirical formula for indicator-dilution curves as obtained in human beings”. In: *Journal of Applied Physiology* 7.2 (1954). PMID: 13211491, pp. 161–167.

[76] T. Benner, S. Heiland, G. Erb, M. Forsting, and K. Sartor. “Accuracy of gamma-variate fits to concentration-time curves from dynamic susceptibility-contrast enhanced MRI: Influence of time resolution, maximal signal drop and signal-to-noise”. In: *Magnetic Resonance Imaging* 15.3 (1997), pp. 307–317. issn: 0730-725X.

[77] S. Turco, P. Frinking, R. Wildeboer, M. Arditi, H. Wijkstra, J. R. Lindner, and M. Mischi. “Contrast-enhanced ultrasound quantification: from kinetic modeling to machine learning”. In: *Ultrasound in medicine & biology* 46.3 (2020), pp. 518–543.

[78] G. I. Taylor. “Dispersion of soluble matter in solvent flowing slowly through a tube”. In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 219.1137 (1953), pp. 186–203.

[79] A. Postema et al. “Dynamic contrast-enhanced ultrasound parametric imaging for the detection of prostate cancer”. In: *BJU Int.* 117.4 (2016), pp. 598–603.

[80] J. Tang, J. Yang, Y. Luo, J. Li, Y. Li, and H. Shi. “Enhancement characteristics of benign and malignant focal peripheral nodules in the peripheral zone of the prostate gland studied using contrast-enhanced transrectal ultrasound”. In: *Clinical Radiology* 63.10 (2008), pp. 1086–1091.

[81] J. G. R. van Sloun, L. Demi, A. Postema, J. M. C. H. J. de la Rosette, H. Wijkstra, and M. Mischi. “Entropy of ultrasound-contrast-agent velocity fields for angiogenesis imaging in prostate cancer”. In: *IEEE Transactions on Medical Imaging* 36.3 (2016), pp. 826–837.

[82] A. Jager, A. W. Postema, M. Mischi, H. Wijkstra, H. P. Beerlage, and J. R. Oddens. “Clinical trial protocol: developing an image classification algorithm for prostate cancer diagnosis on three-dimensional multiparametric transrectal ultrasound”. In: *European Urology Open Science* 49 (2023), pp. 32–43.

[83] R. Wildeboer, S. Schalk, L. Demi, H. Wijkstra, and M. Mischi. “Three-dimensional histopathological reconstruction as a reliable ground truth for prostate cancer studies”. In: *Biomedical Physics & Engineering Express* 3.3 (2017), p. 035014.

[84] S. G. Schalk, A. Postema, T. A. Saidov, L. Demi, M. Smeenge, J. J. de la Rosette, H. Wijkstra, and M. Mischi. “3D surface-based registration of ultrasound and histology in prostate cancer imaging”. In: *Computerized Medical Imaging and Graphics* 47 (2016), pp. 29–39.

[85] A. Jager, A. W. Postema, H. van der Linden, P. T. Nooijen, E. Bekers, C. F. Kweldam, G. Daures, W. Zwart, M. Mischi, H. P. Beerlage, *et al.* “Reliability of whole mount radical prostatectomy histopathology as the ground truth for artificial intelligence assisted prostate imaging”. In: *Virchows Archiv* 483.2 (2023), pp. 197–206.

[86] J. Janjic. “Looking Forward with Minimally Invasive Ultrasound”. PhD thesis. Erasmus University Rotterdam, 2018.

[87] A. D. Bar-Zion, C. Tremblay-Darveau, M. Yin, D. Adam, and F. S. Foster. “Denoising of Contrast-Enhanced Ultrasound Cine Sequences Based on a Multiplicative Model”. In: *IEEE Transactions on Biomedical Engineering* 62.8 (2015), pp. 1969–1980.

[88] P. Mohana Shankar. “A general statistical model for ultrasonic backscattering from tissues”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 47.3 (2000), pp. 727–736.

[89] B. Raju and M. Srinivasan. “Statistics of envelope of high-frequency ultrasonic backscatter from human skin in vivo”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 49.7 (2002), pp. 871–882.

[90] J.-M. Nicolas and F. Tupin. “A new parameterization for the Rician distribution”. In: *IEEE Geoscience and Remote Sensing Letters* 17.11 (2019), pp. 2011–2015.

[91] R. Wagner, S. Smith, J. Sandrik, and H. Lopez. “Statistics of speckle in ultrasound B-scans”. In: *IEEE Transactions on Sonics and Ultrasonics* 30.3 (1983), pp. 156–163.

[92] O. V. Michailovich and A. Tannenbaum. “Despeckling of medical ultrasound images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 53.1 (2006), pp. 64–78.

[93] R. R. Wildeboer, F. Sammali, R. J. G. van Sloun, Y. Huang, P. Chen, M. Bruce, C. Rabotti, S. Shulepov, G. Salomon, B. C. Schoot, H. Wijkstra, and M. Mischi. “Blind source separation for clutter and noise suppression in ultrasound imaging: review for different applications”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 67.8 (2020), pp. 1497–1512.

[94] O. Solomon, R. J. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar. “Exploiting flow dynamics for superresolution in contrast-enhanced ultrasound”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 66.10 (2019), pp. 1573–1586.

[95] N. Yahya, N. S. Kamel, and A. S. Malik. “Subspace-based technique for speckle noise reduction in SAR images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 52.10 (2014), pp. 6257–6271.

[96] J. Grabek and B. Cyganek. “Speckle noise filtering in side-scan sonar Images based on the Tucker tensor decomposition”. In: *Sensors* 19.13 (2019). issn: 1424-8220.

[97] J. Bioucas-Dias and M. Figueiredo. “Multiplicative noise removal using variable splitting and constrained optimization”. In: *IEEE Transactions on Image Processing* 19 (Aug. 2010), pp. 1720–1730.

[98] Y. Dong and T. Zeng. “A convex variational model for restoring blurred images with multiplicative noise”. In: *SIAM Journal on Imaging Sciences* 6.3 (2013), pp. 1598–1625.

[99] J. Fang, T. Mao, F. Bo, B. Hao, N. Zhang, S. Hu, W. Lu, and X. Wang. “An SAR image-despeckling method based on HOSVD using tensor patches”. In: *Remote Sensing* 15.12 (2023), p. 3118.

[100] S. V. M. Sagheer, S. N. George, and S. K. Kurien. “Despeckling of 3D ultrasound image using tensor low rank approximation”. In: *Biomedical Signal Processing and Control* 54 (2019), p. 101595.

[101] Q. Gu, H. Gui, and J. Han. “Robust tensor decomposition with gross corruption”. In: *Advances in Neural Information Processing Systems* 27 (2014).

[102] R. Tomioka and T. Suzuki. “Convex tensor decomposition via structured schatten norm regularization”. In: *Advances in neural information processing systems* 26 (2013).

[103] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan. “Tensor robust principal component analysis with a new tensor nuclear norm”. In: *IEEE transactions on pattern analysis and machine intelligence* 42.4 (2019), pp. 925–938.

[104] W. Fang, D. Wei, and R. Zhang. “Stable tensor principal component pursuit: Error bounds and efficient algorithms”. In: *Sensors* 19.23 (2019), p. 5335.

[105] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.* "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122.

[106] S. Gu, L. Zhang, W. Zuo, and X. Feng. "Weighted nuclear norm minimization with application to image denoising". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 2862–2869.

[107] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer. "Novel methods for multilinear data completion and denoising based on tensor-SVD". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 3842–3849.

[108] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S.-I. Amari. "Bayesian robust tensor factorization for incomplete multiway data". In: *IEEE transactions on neural networks and learning systems* 27.4 (2015), pp. 736–748.

[109] T. Yokota, N. Lee, and A. Cichocki. "Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria". In: *IEEE Transactions on Signal Processing* 65.5 (2016), pp. 1196–1206.

[110] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal. "Cancer statistics, 2021". In: *CA: A Cancer Journal for Clinicians* 71.1 (2021), pp. 7–33.

[111] N. et al. "EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent". In: *European Urology* 71.4 (2017), pp. 618–629. issn: 0302-2838.

[112] O. Ukimura, J. A. Coleman, A. de la Taille, M. Emberton, J. I. Epstein, S. J. Freedland, G. Giannarini, A. S. Kibel, R. Montironi, G. Ploussard, M. J. Roobol, V. Scattoni, and J. S. Jones. "Contemporary role of systematic prostate biopsies: indications, techniques, and implications for patient care". In: *European Urology* 63.2 (2013), pp. 214–230. issn: 0302-2838.

[113] E. W. Stacy. "A generalization of the Gamma distribution". In: *The Annals of Mathematical Statistics* 33.3 (1962), pp. 1187–1192.

[114] M. M. C. Elwenspoek, A. L. Sheppard, M. D. F. McInnes, S. W. D. Merriel, E. W. J. Rowe, R. J. Bryant, J. L. Donovan, and P. Whiting. "Comparison of multiparametric magnetic resonance imaging and targeted biopsy with systematic biopsy alone for the diagnosis of prostate cancer: a systematic review and meta-analysis". In: *JAMA Network Open* 2.8 (Aug. 2019), e198427–e198427.

[115] M. Seitz, A. Shukla-Dave, A. Bjartell, K. Touijer, A. Sciarra, P. J. Bastian, C. Stief, H. Hricak, and A. Graser. “Functional magnetic resonance imaging in prostate cancer”. In: *European Urology* 55.4 (2009), pp. 801–814. issn: 0302-2838.

[116] G. Russo, M. Mischi, W. Scheepens, J. J. De la Rosette, and H. Wijkstra. “Angiogenesis in prostate cancer: onset, progression and imaging”. In: *BJU International* 110.11c (2012), E794–E808.

[117] G. Liu, S. Wu, and L. Huang. “Contrast-enhanced ultrasound evaluation of the prostate before transrectal ultrasound-guided biopsy can improve diagnostic sensitivity: A STARD-compliant article”. In: *Medicine* 99.19 (2020).

[118] C. Strouthos, M. Lampaskis, V. Sboros, A. McNeilly, and M. Averkiou. “Indicator dilution models for the quantification of microvascular blood flow with bolus administration of ultrasound contrast agents”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 57.6 (2010), pp. 1296–1310.

[119] R. J. Gillies, P. A. Schomack, T. W. Secomb, and N. Raghunand. “Causes and effects of heterogeneous perfusion in tumors”. In: *Neoplasia* 1.3 (1999), pp. 197–207.

[120] S. Delorme and M. Knopp. “Non-invasive vascular imaging: assessing tumour vascularity”. In: *European radiology* 8.4 (1998), pp. 517–527.

[121] M. Mischi, M. Kuenen, H. Wijkstra, A. Hendrikx, and H. Korsten. “Prostate cancer localization by contrast-ultrasound diffusion imaging”. In: *International Ultrasonics Symposium*. 2009, pp. 283–286.

[122] T. Joel and R. Sivakumar. “An extensive review on despeckling of medical ultrasound images using various transformation techniques”. In: *Applied Acoustics* 138 (2018), pp. 18–27.

[123] S. Bjaerum, H. Torp, and K. Kristoffersen. “Clutter filter design for ultrasound color flow imaging”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 49.2 (2002), pp. 204–216.

[124] Y. Zhu, M. Kim, C. Hoerig, and M. F. Insana. “Experimental validation of perfusion imaging with HOSVD clutter filters”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 67.9 (2020), pp. 1830–1838.

[125] K. Ozgun and B. Byram. “A channel domain higher-order SVD clutter rejection filter for small vessel ultrasound imaging”. In: *2020 IEEE International Ultrasonics Symposium (IUS)*. 2020, pp. 1–4.

[126] G. Barrois, A. Coron, T. Payen, A. Dizeux, and L. Bridal. “A multiplicative model for improving microvascular flow estimation in dynamic contrast-enhanced ultrasound (DCE-US): theory and experimental validation”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 60.11 (2013), pp. 2284–2294.

[127] J. Rissanen. “A universal prior for integers and estimation by minimum description length”. In: *The Annals of statistics* 11.2 (1983), pp. 416–431.

[128] J. Folkman. “Role of angiogenesis in tumor growth and metastasis”. In: *Seminars in Oncology*. Vol. 29. 6. Elsevier. 2002, pp. 15–18.

[129] P. J. Frinking, A. Bouakaz, J. Kirkhorn, F. J. Ten Cate, and N. De Jong. “Ultrasound contrast imaging: current and new potential methods”. In: *Ultrasound in medicine & biology* 26.6 (2000), pp. 965–975.

[130] M. Mischi et al. “Contrast-ultrasound dispersion imaging of cancer neovascularization by mutual-information analysis”. In: *2014 IEEE Int. Ultrasonics Symposium*. 2014, pp. 1148–1151.

[131] A. G. Bruce, D. L. Donoho, H. Gao, and R. D. Martin. “Denoising and robust nonlinear wavelet analysis”. In: *Wavelet Applications*. Vol. 2242. SPIE. 1994, pp. 325–336.

[132] R. Han, R. Willett, and A. R. Zhang. “An optimal statistical and computational framework for generalized tensor estimation”. In: *The Annals of Statistics* 50.1 (2022), pp. 1–29.

[133] L. Zhu, C.-W. Fu, M. S. Brown, and P.-A. Heng. “A non-local low-rank framework for ultrasound speckle reduction”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 5650–5658.

[134] S. V. M. Sagheer and S. N. George. “Denoising of medical ultrasound images based on non-local similarity: a low-rank approach”. In: *TENCON 2017-2017 IEEE Region 10 Conference*. IEEE. 2017, pp. 176–181.

[135] F. Mezzadri. “How to generate random matrices from the classical compact groups”. In: *arXiv preprint math-ph/0609050* (2006).

[136] X. Hua, L. E. Pierce, and F. T. Ulaby. “Statistical properties of logarithmically transformed speckle”. In: *IEEE Trans. on Geoscience and Remote Sensing* 40.3 (2002), pp. 721–727.

[137] B. Delahunt, J. Miller, R. Srigley, J. A. Evans, and H. Samaratunga. “Gleason grading: past, present and future”. In: *Histopathology* 60.1 (2012), pp. 75–86.

[138] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[139] G. H. Golub and C. Reinsch. “Singular value decomposition and least squares solutions”. In: *Handbook for Automatic Computation: Volume II: Linear Algebra*. Springer, 1971, pp. 134–151.

[140] C. Lanczos. “An iteration method for the solution of the eigenvalue problem of linear differential and integral operators”. In: *Journal of Research of the National Bureau of Standards* 45.4 (1950).

[141] D. T. Fetzer, V. Rafailidis, C. Peterson, E. G. Grant, P. Sidhu, and R. G. Barr. “Artifacts in contrast-enhanced ultrasound: a pictorial essay”. In: *Abdominal Radiology* 43 (2018), pp. 977–997.

[142] M. R. Lowerison, M. N. Hague, A. F. Chambers, and J. C. Lacefield. “Improved linear contrast-enhanced ultrasound imaging via analysis of first-order speckle statistics”. In: *IEEE Trans. on Ultrasonics, Ferroelectrics, and Freq. Control* 63.9 (2016), pp. 1409–1421.

[143] C. F. Dietrich, M. Averkiou, M. B. Nielsen, R. G. Barr, P. N. Burns, F. Calliada, V. Cantisani, B. Choi, M. C. Chammas, D.-A. Clevert, *et al.* “How to perform contrast-enhanced ultrasound (CEUS)”. In: *Ultrasound international open* 4.01 (2018), E2–E15.

[144] A. Matoso and J. I. Epstein. “Defining clinically significant prostate cancer on the basis of pathological findings”. In: *Histopathology* 74.1 (2019), pp. 135–145.

[145] J. Liu, P. Musialski, P. Wonka, and J. Ye. “Tensor completion for estimating missing values in visual data”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.1 (2013), pp. 208–220.

[146] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. “Tensor decomposition for signal processing and machine learning”. In: *IEEE Transactions on signal processing* 65.13 (2017), pp. 3551–3582.

[147] B. Hunyadi, S. Van Huffel, and M. De Vos. “The power of tensor decompositions in biomedical applications”. In: *Machine Learning in Healthcare Technologies, IET* (2016).

[148] J. Zuiddam. “A note on the gap between rank and border rank”. In: *Linear Algebra and its Applications* 525 (2017), pp. 33–44.

[149] P. K. Kothari, A. Moitra, and A. S. Wein. “Overcomplete Tensor Decomposition via Koszul-Young Flattenings”. In: *arXiv preprint arXiv:2411.14344* (2024).

[150] L. R. Tucker *et al.* “The extension of factor analysis to three-dimensional matrices”. In: *Contributions to mathematical psychology* 110119 (1964), pp. 110–182.

[151] G. W. Stewart. *Perturbation theory for the singular value decomposition*. Citeseer, 1998.

[152] W. Hackbusch, D. Kressner, and A. Uschmajew. “Perturbation of higher-order singular values”. In: *SIAM Journal on applied algebra and geometry* 1.1 (2017), pp. 374–387.

[153] C. R. Rao. “Information and the accuracy attainable in the estimation of statistical parameters”. In: *Breakthroughs in Statistics: Foundations and basic theory*. Springer, 1992, pp. 235–247.

[154] J. D. Gorman and A. O. Hero. “Lower bounds for parametric estimation with constraints”. In: *IEEE Transactions on Information Theory* 36.6 (1990), pp. 1285–1301.

[155] P. Stoica and B. C. Ng. “On the Cramér-Rao bound under parametric constraints”. In: *IEEE Signal Processing Letters* 5.7 (1998), pp. 177–179.

[156] T. J. Moore, B. M. Sadler, and R. J. Kozick. “Maximum-likelihood estimation, the Cramér–Rao bound, and the method of scoring with parameter constraints”. In: *IEEE Transactions on Signal Processing* 56.3 (2008), pp. 895–908.

[157] X. Liu and N. D. Sidiropoulos. “Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays”. In: *IEEE Transactions on Signal Processing* 49.9 (2001), pp. 2074–2086.

[158] C. Prévost, K. Usevich, M. Haardt, P. Comon, and D. Brie. “Constrained Cramér–Rao bounds for reconstruction problems formulated as coupled canonical polyadic decompositions”. In: *Signal Processing* 198 (2022), p. 108573. issn: 0165-1684.

[159] R. Boyer, L. De Lathauwer, and K. Abed-Meraim. “Higher order tensor-based method for delayed exponential fitting”. In: *IEEE Transactions on Signal Processing* 55.6 (2007), pp. 2795–2809.

[160] R. Rajan. “Relative Space-Time Kinematics of an Anchorless Network”. Dissertation. Delft University of Technology, 2016.

[161] Y.-H. Li and P.-C. Yeh. “An interpretation of the Moore-Penrose generalized inverse of a singular Fisher Information Matrix”. In: *IEEE Transactions on Signal Processing* 60.10 (2012), pp. 5532–5536.

[162] M. Calis and B. Hunyadi. “Generating a tensor with prescribed mode-n singular values”. In: *TechRxiv* (2025).

[163] M. M. Steinlechner. “Riemannian optimization for solving high-dimensional problems with low-rank tensor structure”. PhD thesis. EPFL, 2016.

[164] E. Nitzan, T. Routtenberg, and J. Tabrikian. “Cramér–Rao bound for constrained parameter estimation using Lehmann-unbiasedness”. In: *IEEE Transactions on Signal Processing* 67.3 (2019), pp. 753–768.

[165] C. Davis and W. M. Kahan. “The rotation of eigenvectors by a perturbation. III”. In: *SIAM Journal on Numerical Analysis* 7.1 (1970), pp. 1–46.

[166] M. Viberg, B. Ottersten, and T. Kailath. “Detection and estimation in sensor arrays using weighted subspace fitting”. In: *IEEE transactions on Signal Processing* 39.11 (1991), pp. 2436–2449.

[167] R. Schmidt. “Multiple emitter location and signal parameter estimation”. In: *IEEE Transactions on Antennas and Propagation* 34.3 (1986), pp. 276–280.

[168] A.-J. Van Der Veen, E. F. Deprettere, and A. L. Swindlehurst. “Subspace-based signal analysis using singular value decomposition”. In: *Proceedings of the IEEE* 81.9 (1993), pp. 1277–1308.

[169] S. T. Smith. “Covariance, subspace, and intrinsic Cramér-Rao bounds”. In: *IEEE Transactions on Signal Processing* 53.5 (2005), pp. 1610–1630.

[170] M. Tur, K. C. Chin, and J. W. Goodman. “When is speckle noise multiplicative?” In: *Appl. Opt.* 21.7 (Apr. 1982), pp. 1157–1159.

[171] H. Bozdogan. “Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions”. In: *Psychometrika* 52.3 (1987), pp. 345–370.

[172] P. Bholowalia and A. Kumar. “EBK-means: a clustering technique based on elbow method and k-means in WSN”. In: *International Journal of Computer Applications* 105.9 (2014).

[173] R. Roy and T. Kailath. “ESPRIT-estimation of signal parameters via rotational invariance techniques”. In: *IEEE Transactions on acoustics, speech, and signal processing* 37.7 (1989), pp. 984–995.

[174] J.-M. Papy, L. De Lathauwer, and S. Van Huffel. “Exponential data fitting using multilinear algebra: the single-channel and multi-channel case”. In: *Numerical linear algebra with applications* 12.8 (2005), pp. 809–826.

[175] W. Buijsman. “Efficient circular Dyson Brownian motion algorithm”. In: *Physical Review Research* 6.2 (2024), p. 023264.

[176] Y. He and G. Joseph. “Bayesian algorithms for Kronecker-structured sparse vector recovery with application to IRS-MIMO channel estimation”. In: *IEEE Transactions on Signal Processing* (2024).

[177] M. Mischi, T. A. Kalker, and E. H. Korsten. “Contrast echocardiography for pulmonary blood volume quantification”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 51.9 (2004), pp. 1137–1147.

[178] M. Mischi, A. H. Jansen, and H. H. Korsten. “Identification of cardiovascular dilution systems by contrast ultrasound”. In: *Ultrasound in medicine & biology* 33.3 (2007), pp. 439–451.

[179] C. Chen, R. Perera, M. C. Kolios, H. Wijkstra, A. A. Exner, M. Mischi, and S. Turco. “The unique second wave phenomenon in contrast enhanced ultrasound imaging with nanobubbles”. In: *Scientific reports* 12.1 (2022), p. 13619.

[180] J.-M. Papy, L. De Lathauwer, and S. Van Huffel. “Exponential data fitting using multilinear algebra: the decimative case”. In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 23.7-8 (2009), pp. 341–351.

ACKNOWLEDGEMENTS

Every story has a beginning and an end. This journey would not have been possible without the help of many people who have accompanied me. I would like to start with acknowledging the shoulders of the giants upon which I stand. They have helped me form an academic identity and have provided consistent reassurance that has fostered my confidence. I would like to thank my promotor, Prof. Dr. Ir. Alle-Jan van der Veen; you have been invaluable in laying the theoretical foundations of this work. Your insightful understanding of what is essential helped initiate some of the research directions in this dissertation. I would like to thank my co-promotor, Prof. Massimo Mischi, for helping with all the ultrasound and prostate cancer-related questions. It was a pleasure to embark on the journey of making a product from the research. Thank you very much for providing me with a foundation that I was able to build on. I would like to thank my daily supervisor and promotor, Dr. Ir. Borbála Hunyadi, for all the discussions and the feedback that created this dissertation. Your network in the tensor community allowed me to connect with many researchers and gave way to multi-faceted discussions. Thank you for your patience, feedback, and encouragement.

A special thanks to Prof. Dr. Ir. Richard Heusdens and Dr. Ir. Richard Hendriks for initiating my academic journey. When I first entered your office in the first year of my Master's, you encouraged me to pursue research opportunities, helping me develop confidence in my academic abilities. Your presence made me feel supported and cared for throughout my time at TU Delft. Furthermore, I would like to thank Dr. Raj Rajan for the discussions that we had about information theory. You were always there when we encountered a problem, and I ended up at your office to "Let's talk about that." Finally, I would like to thank all the members of the staff in SPS and all the researchers for keeping up with the ever-changing landscape of the PhD world and making our lives smoother.

This research project has been in collaboration with Angiogenesis Analytics and TU Eindhoven. Without their guidance and help, we would not be able to conduct an analysis on real prostate recordings. I would like to thank the Angiogenesis Analytics staff, Arman, Wim, Ben, Giuseppe, Rishith, and Marije, for their contributions. Furthermore, I would like to thank Mark Bloemendaal, who defined the project and provided part of the funding. His enthusiasm and management prowess grew the company from three people to thirty. The clinical trials were led and conducted by Hessel Wijkstra, Auke Jager, Arnoud Postema, and Pintong Huang. Thank you for providing me with the datasets. Finally, my colleagues from TU Eindhoven, Florian Delberghe, and Peiran Chen. I would like to thank you for your collaboration and help.

I would like to thank my family, Esin, Vural, Melih, Hare Nil Genç Çalış, and Nesrin Göktuğ, for the absolute trust and care they have provided throughout my studies. Furthermore, I would like to thank my selected family, Denizhan Taşkınsu, for the endless support that he provided in my final years. Thanks for listening to me and comforting me.

I have shared living spaces with many people, and the bond that we created stood firm for me. I want to thank Leo, Carlotta, Petra, Lorenzo, and Amey. You showed me how great it feels to have a home to return to. Of course, thank you, Chatu. You were always there for me when I needed you the most. I was lucky to meet my college friends anew in Delft. Ege Çakır and Irem Gürdal, you became my anchors and safe shores. I appreciate how you created a safe space to be myself and grow to become the person I am. Finally, I would like to thank my friends from the Master's, Idil, Hande, and Pierre, who stayed with me until the end of my PhD. I cherish all the memories we have. Thank you, Sueda, for your companionship on a similar journey. Dear Vicente, at first light on the fifth day, I always look to the west. Thanks, Begüm, for helping me design several illustrations in the dissertation.

Every research starts with a question, and the first help comes from the coffee breaks. I would like to thank my colleagues Yanbo, Zhonggang, Ids, Ruben, Ellen, Changheng, Didem, Aybüke, Cristian, Sinian, Shuoyan, and Peiyuan. It was a pleasure to work with you. After joining a project group during a course, I met a long-haired Dutch colleague who shared my interest in video games. Costas, thanks for the game talks and the beers in Oktoberfest. Yanbin, thank you for the great hotpots and for following through on the secret handshake we created. Giovanni, my oldest colleague and friend in the Netherlands, you are a gem. A special thanks to my office mates, Seline and Alberto, for leading by example. When you work, I work. Sofia, you are the cherry on top. Beware, you might still find my biometric photos somewhere. Anurodh, thanks for the dissertation writing escapades and the snooker and table tennis challenges we had. I will cherish all the nice chats we had after work. What am I going to do without you all?

Finally, I would like to thank my Rotterdam community. You showed me that growth comes from the parts you avoid looking at, the pursuit of happiness comes with many colors, and discourses are an inevitable product of enmeshment. Thanks for challenging me and making me a better human.

CURRICULUM VITæ

Metin Çalış

Metin Calis received his B.Sc. degree from Bilkent University, Ankara, Turkey, in 2016. In 2019, he received his M.Sc. degree in a topic related to privacy-preserving distributed wireless networks from the Delft University of Technology, Delft, the Netherlands. In the Signal Processing Systems (SPS) group at TU Delft, he worked as a research assistant for the NWO-funded project about improving sound localization for hearing-impaired people until 2020. Starting in 2020, he worked as a Ph.D. student working on tensor techniques to improve the analysis of contrast-enhanced ultrasound sequences. His main research interests are audio and acoustic signal processing, biomedical imaging, and tensor decomposition methods. In his free time, his goal is to improve himself in every possible direction, like an RPG character with equal points in all attributes. You may find him sketching, making music, dancing, doing sports, working on a microcontroller, or something completely different.