Rayleigh–quotient optimization on tensor products of Grassmannians

Dissertationsschrift zur Erlangung des naturwissenschaftlichen Doktorgrades der Julius-Maximilians-Universität Würzburg

vorgelegt von

Oana Curtef

aus

Bistrita, Rumänien

Würzburg 2012

Contents

List of Figures

to Luca and Vali

“The truth is rarely pure and never simple.”

Oscar Wilde

Acknowledgements

There are many people who have contributed to this moment and I wish to use this occasion to assure them of my appreciation.

First of all, I would like to thank Prof. Dr. Uwe Helmke, my supervisor, for his constant care and support, his guidance, insightful comments and encouragement.

I also want to thank the entire Mathematical Institute, for the unconventional and friendly atmosphere. To all my present and former colleagues, I thank for the pleasant working climate, for the serious discussions and for the cheerful ones. Special thanks go to Dr. Gunther Dirr for his valuable mathematical advices, stimulative words and his infinite patience.

I would like to express my gratitude to the organizations that have founded my research: Bayerische Eliteförderungsgesetz (BAYEFG), Federal Ministry of Education and Research (BMBF). In particular, I want to thank the Frauenbüro of University Würzburg for the extraordinary support offered to manage research and family during my final PhD phase.

Further, I thank the Group of Foreign Scientists, for many of the friends that I have made and more important, for making me feel like home in Würzburg.

I am thanking to my good friends for their understanding and bearing. To my dear parents and brother, I wish to thank for their unconditional care and trust in me. Exceptional thanks are addressed to all those people who took care of my son when I was busy writing the thesis.

Finally, I thank Vali and Luca, for bringing the Sun on rainy days.

Chapter 1

Introduction

The last decades have introduced a new approach into the theory of constrained optimization, which exploits the geometrical structure of the problem and develops convenient numerical strategies. Explicitly, classical constrained optimization problems can be recast as unconstrained ones by endowing the set of restrictions with a differentiable structure. The new approach is referred to as Riemannian optimization. Applications of Riemannian optimization abound in engineering and areas such as numerical linear algebra, statistics, signal processing, data compression, quantum computing and computer vision. We refer, e.g., to [3, 34, 72] and the references therein. By replacing the traditional tools from numerical optimization (directional derivatives, line search, etc.) with their Riemannian counterparts (geodesics, Levi-Civita connection, parallel transport, etc.), one develops intrinsic methods, which evolve on a parameter space of a smaller dimension than the one of the environment space. Take as an example the eigenvalue computation of a symmetric matrix $A \in \mathbb{R}^{n\times n}$. It is well-known that this can be achieved by determining the critical points of the Rayleigh–quotient function $x^\top Ax$ with $\|x\| = 1$, see e.g. [22, 26]. The classical Lagrange multiplier rule works with a parameter space of dimension $n + 1$, whereas, the set of constraints is the unit sphere and is a $n - 1$ dimensional manifold.

Let $V_1, \dots, V_r$ be finite dimensional vector spaces and let $V_1 \otimes \cdots \otimes V_r$ denote their tensor product space. We call as *r-fold tensor product of manifolds* the set of all *simple tensors* $\mathbf{X} := X_1 \otimes \cdots \otimes X_r$, where X_j is from a submanifold $\mathcal{M}_j$ of V_j, for all $j = 1, \dots, r$. In this work, we propose the task of optimizing a generalization of the Rayleigh–quotient map on a r-fold tensor product of manifolds. Special attention is payed to the optimization task on the r-fold tensor product of Grassmannians and on the r-fold tensor product of Lagrange–Grassmannians. The Grassmannian is the manifold of all rank m self–adjoint projectors of $\mathbb{C}^n$:

$$\mathrm{Gr}_{m,n} = \{P \in \mathbb{C}^{n\times n} \mid P^2 = P^\dagger = P,\ \mathrm{tr}(P) = m\}.$$

The Lagrange–Grassmannian LG_n is the submanifold of $\mathrm{Gr}_{n,2n}$ consisting of all self-adjoint projectors that correspond to Lagrangian subspaces of $\mathbb{C}^{2n}$. By identifying the abstract vector spaces V_j with spaces of Hermitian matrices and the tensor product

$\otimes$ with the matrix Kronecker product, the r-fold tensor product of Grassmannians $\mathrm{Gr}^{\otimes}(\mathbf{m}, \mathbf{n})$ is the set of all Kronecker products $\mathbf{P} := P_1 \otimes \cdots \otimes P_r$ with $P_j \in \mathrm{Gr}_{m_j,n_j}$, for $j = 1, \ldots, r$. Here, $(\mathbf{m}, \mathbf{n})$ is a shortcut for the multi–index $\big((m_1, n_1), \ldots, (m_r, n_r)\big)$. Similarly, the r-fold tensor product of Lagrange–Grassmannians $\mathrm{LG}^{\otimes}(\mathbf{n})$ is the set of all Kronecker products $P_1 \otimes \cdots \otimes P_r$ of self-adjoint projectors $P_j \in \mathrm{LG}_{n_j}$, for $j = 1, \ldots, r$.

The optimization task that we are interested in is given as follows:

$$\max_{\mathbf{P} \in \mathcal{M}} \operatorname{tr}(A\mathbf{P}), \tag{1.1}$$

where $\mathcal{M}$ is either $\mathrm{Gr}^{\otimes}(\mathbf{m}, \mathbf{n})$ or $\mathrm{LG}^{\otimes}(\mathbf{n})$ and $A \in \mathbb{C}^{N \times N}$ is Hermitian, $N := n_1 n_2 \cdots n_r$. We call the map $\rho_A(\mathbf{P}) := \operatorname{tr}(A\mathbf{P})$, the *generalized Rayleigh–quotient* (GRQ) of A with respect to the partitioning $(\mathbf{m}, \mathbf{n})$. Depending on the structure of A and on the partitioning $(\mathbf{m}, \mathbf{n})$, optimization problem (1.1) relates to well–known optimization tasks in numerical linear algebra such as:

(i) **the best approximation of a tensor with a tensor of lower rank**,

(ii) **geometric measures of pure state entanglement**,

(iii) **the problem of subspace reconstruction**.

In the sequel, we give a short description of the applications mentioned above and also their historical background.

(i) For Hermitian matrices of rank–one, i.e. $A = vv^{\dagger}$, problem (1.1) becomes an application from areas such as statistics, signal processing and data compression, the best low-rank approximation of a tensor $\mathcal{T} \in \mathbb{C}^{n_1 \times \cdots \times n_r}$, [12, 30, 46, 79]. A detailed explanation on this relation is given in Section 4.2.1. Here we just mention that the matrix A and the tensor $\mathcal{T}$ are related by the fact that v is obtained by arranging the elements of $\mathcal{T}$ in a lexicographical order. To tackle the problem, one has to define first an appropriate notion for the tensor rank. This can be done by noting that any tensor $\mathcal{T} \in \mathbb{C}^{n_1 \times \cdots \times n_r}$ can be decomposed in a finite sum of rank–one tensors $x^1 \otimes \cdots \otimes x^r$ with $x^j \in \mathbb{C}^{n_j}$ and $j = 1, \ldots, r$. Thus, one can define the rank of a tensor as the minimal number of rank–one summands. In contrast to the rank of a matrix, for which it is known that the column rank and the row rank are equal, for a tensor this is not necessarily true. There are several possible rank definitions for a tensor and also corresponding singular value decompositions (SVD), each satisfying only partly the properties of the matrix rank and matrix SVD, see [51, 53, 66]. One of these definition which is of interest to us is the so-called multilinear rank-$(m_1, \ldots, m_r)$ of a tensor $\mathcal{T} \in \mathbb{C}^{n_1 \times \cdots \times n_r}$, which is the analog of the row–column rank for matrices, [50]. A classical result from linear algebra, the Eckart–Young theorem [18], asserts that the best low rank approximation to a matrix is given by a truncated singular value decomposition. With the known definitions of rank and SVD for tensors, there is still no available analog of the Eckart–Young theorem for higher–order tensors. The first question that arises is if the problem of best low-rank tensor approximation is well–posed. In [66], it is proven that this is not always the case since the set of all tensors of rank (minimal number of rank-one summands) less than k with $k > 1$ is not always closed. But, the set of all tensors with a certain multilinear rank is closed and hence, the problem of best approximating a tensor with a tensor of lower multilinear rank is well-defined. We make the following remark: the set of rank–one tensors and the set of tensors with multilinear rank-$(1, \ldots, 1)$ are identical. From now on, whenever we discuss about the rank of a tensor we have in mind the multilinear rank. An already classical numerical approach used to tackle the best low-rank approximation problem is the higher-order orthogonal iteration (HOOI) [51], i.e. an alternating least-squares algorithm that generalizes the well–known Power Method [27]. Recently, there have been developed several new methods which exploit the geometric structure of the problem: Newton algorithms have been proposed in [20, 41], quasi-Newton methods in [61], conjugate gradient and trust region methods in [42].

(ii) A key application in quantum computing characterizes and quantifies pure state entanglement. The geometric measure of entanglement provides a measure from a pure state to the set of all product states, [15, 56, 76]. Recalling that a pure state is an element in a tensor product space and a product state is just a simple tensor, the quantum entanglement problem is equivalent to a best rank–1 tensor approximation problem.

(iii) The so-called "chicken–and–egg" problem in computer vision refers to the task of recovering subspaces of possibly different dimensions from noisy data, known also as subspace detection or subspace clustering problem [37, 73]. For a particular class of matrices defined in (4.37), the subspace clustering task can be characterized by problem (1.1). The subspace clustering problem of estimating a mixture of linear subspaces from sampled data points has numerous applications in computer vision (image segmentation [78], motion segmentation [75], face clustering [37]), image processing (image representation and image compression [38]) and system theory (hybrid system identification [74]). Classical iterative approaches used for the subspace clustering task, are generalizations of the K–means algorithm [17] such as K–planes [9] and K-subspaces algorithms [1, 70], or probabilistic methods such as Maximum Likelihood estimation [69]. A new approach, that exploits the algebraic structure of the clustering problem, was proposed in [73]. The method proposed in [73] gives a good starting point for iterative methods and in the case of unperturbed data (data lying exactly in the union of some subspaces), it computes the exact subspaces.

In this thesis we perform a thorough analysis on the critical points of the GRQ and develop two Riemannian methods to solve problem (1.1): a Newton-like and a conjugate gradient method. Since the convergence properties of Newton algorithms depend on the nondegeneracy of the critical points, we present a careful analysis of the genericity properties of generalized Rayleigh-quotient functions. As a consequence of the Parametric Transversality Theorem [36], we conclude that for a generic Hermitian matrix $A \in \mathbb{C}^{N\times N}$, the critical points of ρ_A are nondegenerate. A similar result can be formulated also for the GRQ on $\mathrm{LG}^{\otimes}(\mathbf{m})$ for a thin subspace of the space of Hermitian matrices, that we have called the space of decomposable Hermitian-Hamiltonian matrices. Of particular interest from the application point of view is the situation when A is semi–positive definite and even more interesting, when A is of rank–one. For the last situation, we obtain that the critical points satisfying a certain condition are generically nondegenerate and in particular, the global maximizers satisfy this property. In the case of the best rank–one tensor approximation problem, all critical points of the generalized Rayleigh–quotient except the global minimizers, are generically nondegenerate. All these results are detailed in Section 4.4.

On a Riemannian manifold, the intrinsic Newton method is usually described by means of the Levi–Civita connection thus, the iterations are performed along the geodesics and we refer the reader to the works of Gabay [23] and Smith [68]. However, a closed form for the geodesics is not always possible and even when there is such a closed form, the computation of the exponential map can be time expensive. Thus, a more general approach by Shub in [65] uses local coordinates to replace geodesics, Levi-Civita connections and parallel transport by suitable approximations without losing convergence properties of the algorithms. This technique has gained a lot of interest through the works of Helmke [34] in the 1990s and later on by the work of Absil, Sepulchre and Mahony [3]. The idea in the work of Helmke, Hüper and Trumpf [32] is to employ efficient local parametrizations which preserve local convergence. Here, we adapt the ideas in [32] and use a pair of local parametrizations —normal coordinates for the push-forward and QR–type coordinates or Cayley coordinates for the pull-back— satisfying an additional compatibility condition to preserve quadratic convergence. In this way, the resulting intrinsic Newton–like method has the advantage of computational flexibility. However, for high–dimensional problems, the computation of the Hessian and of the solution of the Newton equation remains an expensive task, both in terms of computational complexity and memory requirements. Hence, as an alternative, we propose a conjugate gradient method, which has the advantage of algorithmic simplicity. We replace the global line-search of the classical conjugate gradient method by a one–dimensional Newton–step. This yields a better convergence behavior near stationary points than the commonly used Armijo–rule. The two methods, tailored to the applications described previously, are compared with other algorithms in the literature.

This work is structured as follows: In Chapter 2 we introduce the basic ingredients of Riemannian optimization, i.e. Levi–Civita connection, geodesics, parallel transport and the computation of the intrinsic gradient and Hessian for smooth objective functions defined on a Riemannian manifold. Starting with Section 2.2 we address the question when is the $r-$fold tensor product $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_r$ of submanifolds $\mathcal{M}_j$ a Riemannian submanifold of the tensor product vector space $V_1 \otimes \cdots \otimes V_r$. Identifying an equivalence relation $\sim$ on the $r-$fold direct product of manifolds $\mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ such that $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_r$ is in a one-to-one correspondence with the space of all equivalence classes of $\sim$, we give sufficient conditions for $(\mathcal{M}_1 \times \cdots \times \mathcal{M}_r)/_\sim$ to be a quotient manifold. In particular, when the equivalence relation $\sim$ is induced by the action of a Lie-subgroup of

$$G = \{(\alpha_1, \dots, \alpha_r) \in \mathbb{C}^r \mid \alpha_1\alpha_2 \cdots \alpha_r = 1\},$$

then the quotient space is a Riemannian manifold. We mention that in the thesis we have defined Lie-subgroups as closed subgroups of Lie-groups, that can be equipped with a Lie-group structure. The one-to-one correspondence between $(\mathcal{M}_1 \times \cdots \times \mathcal{M}_r)/_\sim$ and $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_r$ induces a Riemannian manifold structure on $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_r$, but the topology is not consistent with the subspace topology of $V_1 \otimes \cdots \otimes V_r$. Some transversality conditions are still necessary to guarantee the submanifold structure of $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_r$.

The geometry of the Grassmannian and the Lagrange-Grassmannian are well studied in the literature, [4, 8, 34]. In Chapter 3, we recall the basics and generalize the geometric structure of Grassmannian and Lagrange–Grassmannian to the r-fold tensor product of Grassmannians and Lagrange–Grassmannians, respectively. Moreover, in Proposition 3.2.3 and Proposition 3.4.2 we show that $\mathrm{Gr}^{\otimes}(\mathbf{m}, \mathbf{n})$ and $\mathrm{LG}^{\otimes}(\mathbf{n})$ are isometric to the direct product of Grassmannians and Lagrange–Grassmannians, respectively. Beside the classical Lagrangian subspaces of $\mathbb{R}^{2n}$ or $\mathbb{C}^{2n}$, we tackle also the case of complex Lagrangian subspaces of $\mathbb{C}^{2n}$, i.e. Lagrangian subspaces of $\mathbb{C}^{2n}$ defined with respect to a sesquilinear map. It was proven in [5] that the set of all the complex Lagrangian subspaces of $\mathbb{C}^{2n}$ have a manifold structure. We give a diffeomorphism to the set of all self-adjoint projectors $P \in \mathrm{Gr}(n, 2n)$ that satisfy $PJP = 0$, where J is the standard symplectic form

$$J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}. \tag{1.2}$$

Moreover, we show that the r-fold tensor product of complex Lagrange–Grassmannians is a submanifold of $\mathrm{Gr}^{\otimes}(\mathbf{n}, \mathbf{2n})$ and is isometric to the direct product of complex Lagrange–Grassmannians.

Chapter 4 is dedicated to the optimization of the generalized Rayleigh–quotient on the r-fold tensor product of Grassmannians. We start with a comparison between the optimization of the GRQ and the optimization of the classical Rayleigh–quotient. The main point is that the classical Rayleigh–quotient has only global maximizers and minimizers, while GRQ has also local ones, as we show in Example 4.1.2. We give an explicit form of the Riemannian gradient and the Hessian of ρ_A and investigate in detail the critical points of the GRQ. In Section 4.2 we explicitly describe the connection between the optimization of the GRQ and problems from various areas such as the multilinear low–rank tensor approximation, the geometric measure of entanglement, subspace clustering problem and combinatorial problems. An important question regarding the local extrema of GRQ is if they are nondegenerate. Using tools from the transversality theory, we prove in Theorem 4.3.3, that the critical points of GRQ on $\mathrm{Gr}^{\otimes}(\mathbf{m}, \mathbf{n})$ are generically nondegenerate. Explicitly, the set of all Hermitian matrices A such that all the critical points of ρ_A are nondegenerate is open and dense. For the proof, we make use of a more general result stated as Theorem 4.3.2. To avoid confusion, we specify that a property holds *generically* if it holds on a *residual set*, which is a subset of a topological space that contains the intersection of a countable family of dense and open sets. It is obvious that if a set is open and dense, then it is also residual. In Section 4.4, we investigate the nondegeneracy of the critical points of ρ_A when the parameter is a low-rank semi-positive definite matrix, i.e. $A = XX^\dagger,\ X \in \mathrm{St}_{K,N} = \{Y \in \mathbb{C}^{N\times K} \mid N > K,\ \text{rank } Y = K\}$. In Theorem 4.4.2 we have emphasized an important property of ρ_A (A low–rank semi–positive definite), which gives a necessary condition for the nondegeneracy of the critical points of GRQ. Actually, Theorem 4.4.2 is a key tool in proving the generic nondegeneracy of the critical points of GRQ for the restricted parameter set. In Theorem 4.4.5 we give a lower bound for K such that the critical points of $\rho_{XX^\dagger}$ are nondegenerate for X in an open and dense subset of $\mathrm{St}_{K,N}$. As previously stated, the optimization of $\rho_{xx^\dagger},\ x \in \mathbb{C}^N$ on $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ corresponds to the best approximation of a tensor with a tensor of lower rank. In Theorem 4.4.10 we prove that the set of all $x \in \mathbb{C}^N$ for which all the critical points of $\rho_{xx^\dagger}$ that satisfy some conditions, is residual. In particular, Corollary 4.4.11 states that the global maximizers of the GRQ for the best low–rank tensor approximation problem are generically nondegenerate.

The optimization of the generalized Rayleigh-quotient on the r-fold tensor product of Lagrange–Grassmannians is the subject of Chapter 5. We motivate our optimization task by specifying a relation between the optimization of the classical Rayleigh–quotient of a Hamiltonian matrix and solutions of a matrix Riccati equation. For a Hermitian matrix A, we show that the optimization of the classical Rayleigh–quotient $\mathrm{LG}(n)$ is equivalent to the optimization of the classical Rayleigh-quotient of the Hamiltonian part of A. By introducing the notions of decomposable Hermitian Hamiltonian $A^{\mathfrak{h}}$ and decomposable skew–Hermitian Hamiltonian $A^{\mathfrak{s}}$ matrices for $A \in \mathfrak{her}_N$ we show that when A is of the form $A = A^{\mathfrak{h}} + A^{\mathfrak{s}}$, then the optimization of ρ_A on $\mathrm{LG}^{\otimes}(\mathbf{n})$ is equivalent to the optimization of $\rho_{A^{\mathfrak{h}}}$ on $\mathrm{LG}^{\otimes}(\mathbf{n})$. Further, we analyze the critical points of the GRQ on $\mathrm{LG}^{\otimes}(\mathbf{n})$ and derive explicit formulas for the gradient and the Hessian of GRQ. With a similar argumentation as in the case of GRQ on $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$, we prove that the critical points of ρ_A on $\mathrm{LG}^{\otimes}(\mathbf{n})$ are nondegenerate for A from an open and dense subset of the space of Hermitian Hamiltonian matrices. In Corollary 5.3.8 we show that this result still holds even if the space of parameters is reduced to the space of decomposable Hermitian Hamiltonian matrices.

In Chapter 6 we develop two numerical approaches for the optimization problem (1.1): a Riemannian Newton–like and a Riemannian conjugate gradient algorithm. As mentioned, we follow the strategy initiated in [32] and use a QR–decomposition for the first order approximation of the matrix exponential to compute the update. The costs are considerably cheaper than the costs needed for the matrix exponential. For the conjugate gradient method we compute a one-dimensional Newton–step instead of the standard Armijo step-size. This procedure has the advantage of a fast convergence rate, as the experimental results at the end of the chapter will show. As a consequence of a result from [32], we specify that the sequence generated by the Newton-like algorithm converges quadratically to a stationary point of the generalized Rayleigh–quotient, when one starts in the neighborhood of that stationary point. Unfortunately, there is no convergence result available for our conjugate gradient method. However, the numerical experiments tailored to the applications show a very good behavior for the conjugate gradient method. At the end of the chapter, we compare our algorithms with the state of the art.

We mention that part of the results in Chapter 4 and Chapter 6 are published in the paper [13]
"Riemannian optimization on tensor products of Grassmann manifolds: Applications to generalized Rayleigh-quotients manifolds."

Chapter 2

Tensor products of Riemannian manifolds

Applications in various research areas such as signal processing [11, 50, 71], quantum computing [56], and computer vision [73], can be described as constrained optimization tasks on certain subsets of tensor products of vector spaces. Here we want to make use of techniques from Riemannian geometry and tackle constrained optimization tasks as unconstrained ones. In this chapter, we investigate when the subsets of so-called simple tensors can be equipped with a differentiable structure. We start with some preliminary notions from differential geometry and refer to the broad literature for more thorough investigation [1, 2, 3, 34, 36, 44].

By spotting an equivalence relation on the direct product of manifolds $\mathcal{M}_1 \times \cdots \times \mathcal{M}_r$, which is consistent with the tensor product operation, we give sufficient conditions for the quotient space to be a quotient manifold. In fact, we prove that if the equivalence relation is induced by the action of a certain Lie-group on the direct product manifold, then the quotient space is a manifold. And hence, the set of simple tensors has a manifold structure induced by the one-to-one correspondence with the quotient space. Unfortunately, this does not imply that the set of simple tensors is a submanifold of the tensor product vector space, as Example 2.2.11 shows. In Theorem 2.2.12 we give sufficient conditions for the set of simple tensors to have a submanifold structure.

2.1 Preliminaries

The object of this section is to familiarize the reader with the basic language of and fundamental theorems in differential geometry, which are essential for the entire thesis. For a detailed discussion on this topic, the reader is referred to the rich literature [2, 3, 36].

We recall that an *n-dimensional manifold* $\mathcal{M}$ is a second countable Hausdorff space which is locally homeomorphic to $\mathbb{R}^n$. A formal definition of a manifold uses the concepts of charts and atlases. A *chart* (φ, U) is a couple consisting of an open set $U \subset \mathcal{M}$ and a homeomorphism $\varphi : U \to \varphi(U) \subset \mathbb{R}^n$. Thus, for a set $\mathcal{M}$ to be a manifold or a *topological manifold* it is required that each point must be at least in one chart. The collection of charts $\{(U_\alpha, \varphi_\alpha)\}$ that cover $\mathcal{M}$ forms an *atlas*. The *transition maps* of the atlas

$$\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$$

define the differentiable structure of the manifold, which allows one to do calculus on manifolds. Hence, if the transition maps are differentiable, then we have a *differentiable manifold*. If the transition maps are smooth, then we have *smooth manifold*, and if the transition maps are analytic, then we have an *analytic manifold*. The inverse of a chart is called a *local parametrization*. Every n-dimensional vector space is an n-dimensional manifold. Manifolds naturally arise as solutions of systems of equations and as graphs of functions. The circle and the line are one-dimensional manifolds, the n-sphere is an n-dimensional manifold.

An important feature of a manifold $\mathcal{M}$ is the *tangent space* $\mathrm{T}_p\mathcal{M}$ at a point $p \in \mathcal{M}$. There are various ways to define tangent spaces to a manifold, and the most intuitive one describes tangent vectors via equivalence classes of velocity vectors of curves. See the literature also for other approaches [2, 3]. The disjoint union of all tangent spaces of a manifold is called the *tangent bundle* of the manifold, that is

$$\mathrm{T}\mathcal{M} = \bigcup_{p \in \mathcal{M}} \{p\} \times \mathrm{T}_p\mathcal{M}.$$

A smooth section in the tangent bundle of a manifold $\mathcal{M}$ is called *vector field*, i.e. $\mathcal{X} : \mathcal{M} \to \mathrm{T}\mathcal{M},\ p \mapsto \mathcal{X}(p) \in \mathrm{T}_p\mathcal{M}$. The space of all vector fields on $\mathcal{M}$ is denoted with $C^\infty(\mathcal{M}, \mathrm{T}\mathcal{M})$.

Using charts, one introduces the concept of differentiability of maps between manifolds. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be smooth manifolds of dimensions n_1 and n_2 respectively and let F be a map from $\mathcal{M}_1$ to $\mathcal{M}_2$. If (U_1, φ_1) and (U_2, φ_2) are charts around x and $F(x)$ on $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively, then F is *differentiable at* $x \in \mathcal{M}_1$ if its *coordinate representation*

$$\varphi_2 \circ F \circ \varphi_1^{-1} : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$$

is smooth at $\varphi_1^{-1}(x)$. And F is differentiable if it is differentiable at every $x \in \mathcal{M}_1$. Moreover, F is *smooth* if it has derivatives of all orders. The concept of differentiability of a map between two smooth manifolds does not depend on the choice of the charts around x. The *differential* or the *tangent map* of F at $x \in \mathcal{M}_1$ is the linear map

$$\mathrm{T}_x F : \mathrm{T}_x\mathcal{M}_1 \to \mathrm{T}_{F(x)}\mathcal{M}_2$$

and the *rank* of F at $x \in \mathcal{M}_1$ is the dimension of the image of $\mathrm{T}_x F$. If the rank of F is everywhere equal to n_1, then F is an *immersion* and if the rank of F is everywhere equal to n_2, then it is a *submersion*. If F is an immersion and a homeomorphism onto its image, then it is called *embedding*.

A *submanifold* of a manifold $\mathcal{M}$ is a subset $\mathcal{N} \subset \mathcal{M}$ which is a manifold with respect to the subspace topology. Next, we recall two important results that are often used to define submanifolds, the *regular value theorem* and the *embedding theorem*.

Theorem 2.1.1 *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be manifolds of dimension n_1 and n_2, respectively, with $n_1 \geq n_2$ and $F : \mathcal{M}_1 \to \mathcal{M}_2$ a smooth function. If $y \in \mathcal{M}_2$ is a regular value for F, i.e. F has rank n_2 at every point in $F^{-1}(y)$, then $F^{-1}(y)$ is either empty or a submanifold of $\mathcal{M}_1$ of dimension $n_1 - n_2$.*

Theorem 2.1.2 *A subset $\mathcal{N}$ of a manifold $\mathcal{M}$ is a submanifold if and only if there exists an embedding F such that $\mathcal{N}$ is the image of F.*

Quotient manifolds are other well studied objects in differential geometry. Some general aspects related to quotient manifolds are given as follows. If R is an equivalence relation on a set $\mathcal{M}$, then the set

$$[x] = \{y \in \mathcal{M} \mid xRy\}$$

is called an *equivalence class* for R and the set of all equivalence classes is called the *quotient space* of $\mathcal{M}$ with respect to R and denoted $\mathcal{M}/R$. The quotient topology on $\mathcal{M}/R$ is defined to be the finest topology such that the *canonical projection*

$$\pi : \mathcal{M} \to \mathcal{M}/R,\ x \mapsto [x]$$

is continuous. If $\mathcal{M}$ is a manifold and $\mathcal{M}/R$ is a quotient space of $\mathcal{M}$ with respect to an equivalence relation R, then we say that $\mathcal{M}/R$ is a *quotient manifold* if it carries a unique manifold structure such that the canonical projection is a submersion. In this case we say that the equivalence relation R is *regular*. The next result gives a necessary and sufficient condition for an equivalence relation on a manifold to be regular (see [2], page 209).

Theorem 2.1.3 *Let R be an equivalence relation on a manifold $\mathcal{M}$ with graph*

$$\Gamma_R := \{(x, y) \in \mathcal{M} \times \mathcal{M} \mid xRy\} \subset \mathcal{M} \times \mathcal{M}.$$

Then R is regular if and only if

(i) *Γ_R is a closed submanifold of $\mathcal{M} \times \mathcal{M}$,*

(ii) *$\pi_1 : \Gamma_R \to \mathcal{M}$, $\pi_1(x, y) = x$ is a submersion, and*

Moreover, one has

$$\dim(\mathcal{M}/R) = 2 \dim \mathcal{M} - \dim(\Gamma_R).$$

Since the canonical projection π is a submersion, it follows that every $[x] \in \mathcal{M}/R$ is a regular value for π and hence, each equivalence class $[x] \subset \mathcal{M}$ is a submanifold of $\mathcal{M}$ of dimension $\dim \mathcal{M} - \dim(\mathcal{M}/R)$. The tangent space $\mathcal{V}_y := \mathrm{T}_y[x]$ of $[x]$ at $y \in [x]$ is a subspace of $\mathrm{T}_y\mathcal{M}$, called the *vertical space*. For every $y \in \mathcal{M}$ there exists a subspace $\mathcal{H}_y$ of $\mathrm{T}_y\mathcal{M}$ such that $\mathcal{V}_y \oplus \mathcal{H}_y = \mathrm{T}_y\mathcal{M}$. We refer to $\mathcal{H}_y$ as a *horizontal space* at y. The vectors in a horizontal space provide a convenient representation for the tangent vectors of the quotient manifold $\mathcal{M}/R$ as will be explained next.

If $[x] \in \mathcal{M}/R$ and $y \in [x]$, then a representation of $\xi_R \in \mathrm{T}_{[x]}\mathcal{M}/R$ is a tangent vector $\xi \in \mathrm{T}_y\mathcal{M}$ such that

$$\mathrm{T}_y\pi(\xi) = \xi_R. \tag{2.1}$$

However, there are infinitely many such representations since $\xi + \eta$ satisfies (2.1) for all $\eta \in \mathrm{T}_y[x]$. From computational considerations one would like to have a unique representation for ξ_R and this is possible in terms of elements in the horizontal space. Particularly, there exists a unique $\xi_y^h \in \mathcal{H}_y$ (depending on y) that satisfies (2.1) and this will be taken as a representation for the tangent vector ξ_R and named the *horizontal lift* of ξ_R.

2.1.1 Riemannian manifolds and the Levi-Civita connection

In differential geometry, a *Riemannian metric* g on a manifold $\mathcal{M}$ is a smoothly varying family of inner products $g_p : \mathrm{T}_p\mathcal{M} \times \mathrm{T}_p\mathcal{M} \to \mathbb{R}$ on the tangent spaces $\mathrm{T}_p\mathcal{M}$, $p \in \mathcal{M}$. A *Riemannian manifold* is a pair $(\mathcal{M}, g)$ of a smooth real manifold $\mathcal{M}$ and a Riemannian metric g. When the explicit Riemannian metric is not important, we simply say that $\mathcal{M}$ is a Riemannian manifold.
If $\mathcal{N}$ is a submanifold of a Riemannian manifold $(\mathcal{M}, g)$, then $(\mathcal{N}, \widetilde{g})$ is a *Riemannian submanifold* of $\mathcal{M}$, where $\widetilde{g}$ is the induced Riemannian metric, i.e.

$$\widetilde{g}_p(\xi, \eta) = g_p(\xi, \eta), \tag{2.2}$$

for all $\xi,\ \eta \in \mathrm{T}_p\mathcal{N}$, $p \in \mathcal{N}$.

An *affine connection* on a manifold $\mathcal{M}$ is a bilinear map

$$\boldsymbol{\nabla} : C^\infty(\mathcal{M}, \mathrm{T}\mathcal{M}) \times C^\infty(\mathcal{M}, \mathrm{T}\mathcal{M}) \to C^\infty(\mathcal{M}, \mathrm{T}\mathcal{M}),$$

with the following properties:

(1) $\boldsymbol{\nabla}_{f\mathcal{X}}\mathcal{Y} = f\boldsymbol{\nabla}_{\mathcal{X}}\mathcal{Y}$,

(2) $\boldsymbol{\nabla}_{\mathcal{X}}(f\mathcal{Y}) = df(\mathcal{X})\mathcal{Y} + f\boldsymbol{\nabla}_{\mathcal{X}}\mathcal{Y}$,

for all $f : \mathcal{M} \to \mathbb{R}$ smooth and $\mathcal{X}, \mathcal{Y} \in C^\infty(\mathcal{M}, \mathrm{T}\mathcal{M})$.
On a Riemannian manifold $(\mathcal{M}, g)$, one usually chooses an affine connection $\boldsymbol{\nabla}$ that is symmetric and compatible with the metric, i.e.

(i) symmetry: $\boldsymbol{\nabla}_{\mathcal{X}}\mathcal{Y} - \boldsymbol{\nabla}_{\mathcal{Y}}\mathcal{X} = [\mathcal{X}, \mathcal{Y}]$, for any vector fields $\mathcal{X},\ \mathcal{Y}$ on $\mathcal{M}$,

(ii) compatibility with the metric: $\boldsymbol{\nabla} g = 0$,

called the *Riemannian* or *Levi-Civita connection.* The notion of symmetry is defined by the concept of *Lie-bracket* $[\mathcal{X}, \mathcal{Y}]$ of two vector fields $\mathcal{X}$ and $\mathcal{Y}$. If $\mathcal{X}$ and $\mathcal{Y}$ are vector fields on a manifold $\mathcal{M}$, then $[\mathcal{X}, \mathcal{Y}]$ is the differential operator, which assigns to $\mathcal{X}$ and $\mathcal{Y}$ another vector field defined as

$$[\mathcal{X}, \mathcal{Y}](f) := \mathcal{X}(\mathcal{Y}(f)) - \mathcal{Y}(\mathcal{X}(f)),$$

for all smooth real-valued functions f on $\mathcal{M}$. It is well-known that every Riemannian manifold carries a unique *Levi-Civita connection*.

Let $\mathcal{N}$ be a Riemannian submanifold of a Riemannian manifold $\mathcal{M}$ and $\widetilde{\mathcal{X}} \in C^\infty(\mathcal{M}, \mathrm{T}\mathcal{M})$ be a smooth extension of a vector field $\mathcal{X}$ on $\mathcal{N}$. If $\widetilde{\boldsymbol{\nabla}}$ is the Levi-Civita connection on $\mathcal{M}$, then the Levi-Civita connection on $\mathcal{N}$ is given by

$$\boldsymbol{\nabla}_\xi \mathcal{X} := \widetilde{\boldsymbol{\nabla}}_\xi \widetilde{\mathcal{X}},$$

for all $\xi \in \mathrm{T}_p\mathcal{N}$. If moreover, $\mathcal{M}$ is an Euclidean vector space, then the Levi-Civita connection on $\mathcal{N}$ is obtained by projecting the usual directional derivative of $\widetilde{\mathcal{X}} \in C^\infty(\mathcal{M}, \mathcal{M})$ onto the tangent space $\mathrm{T}_p\mathcal{N}$, $p \in \mathcal{N}$, i.e.

$$\boldsymbol{\nabla}_\xi \mathcal{X} = \mathrm{proj}_p(D\widetilde{\mathcal{X}}(p)(\xi)), \tag{2.3}$$

for all $\xi \in \mathrm{T}_p\mathcal{N}$, where proj_p is the orthogonal projection onto $\mathrm{T}_p\mathcal{N}$.

Let $\widetilde{\boldsymbol{\nabla}}$ be the Levi-Civita connection on a Riemannian manifold $\mathcal{M}$. Then, for the Riemannian quotient manifold $\mathcal{M}/R$, the Levi-Civita connection $\boldsymbol{\nabla}^h$ is given by

$$\boldsymbol{\nabla}^h_{\xi_R} \mathcal{X}_R := \mathrm{proj}^h_q(\widetilde{\boldsymbol{\nabla}}_{\xi^h} \mathcal{X}^h), \tag{2.4}$$

for all $\xi_R \in \mathrm{T}_{[p]}\mathcal{M}/R$ and all vector fields $\mathcal{X}_R$ on $\mathcal{M}/R$, where ξ^h is the horizontal lift of ξ, $\widetilde{\mathcal{X}} \in C^\infty(\mathcal{M}, \mathrm{T}\mathcal{M})$ is a smooth extension of X and proj^h_q is the orthogonal projection onto the horizontal space $\mathcal{H}_q$, $q \in [p]$.

2.1.2 Parallel transport and geodesics

By means of the Levi-Civita connection $\boldsymbol{\nabla}$, one defines parallel transport and geodesics on a Riemannian manifold $\mathcal{M}$ as follows. Let $t \mapsto \mathcal{X}(t)$ be a vector field along a smooth curve $\gamma : I \subset \mathbb{R} \to \mathcal{M}$, i.e. $\mathcal{X}(t) \in \mathrm{T}_{\gamma(t)}\mathcal{M}$ for all $t \in I$. Then, $\mathcal{X}$ is *parallel along* γ, if and only if

$$\boldsymbol{\nabla}_{\frac{d}{dt}\gamma(t)} \mathcal{X}(t) = 0 \tag{2.5}$$

for all $t \in I$. Conversely, given $\xi \in \mathrm{T}_{\gamma(0)}\mathcal{M}$, there exists a unique parallel vector field $\mathcal{X}$ along γ such that $\mathcal{X}(0) = \xi$, and the vector

$$\mathcal{X}(t) \in \mathrm{T}_{\gamma(t)}\mathcal{M}$$

is called the *parallel transport* of ξ to $\mathrm{T}_{\gamma(t)}\mathcal{M}$ along γ. In particular, γ is called a *geodesic* on $\mathcal{M}$, if $\frac{d}{dt}\gamma$ is parallel along γ, i.e. if

$$\boldsymbol{\nabla}_{\frac{d}{dt}\gamma(t)} \left(\frac{d}{dt}\gamma(t) \right) = 0 \tag{2.6}$$

for all $t \in I$. From the theory of ODEs one knows that for any $p \in \mathcal{M}$ and $\xi \in \mathrm{T}_p\mathcal{M}$ there exists a unique geodesic γ with $\gamma(0) = p$ and $\frac{d}{dt}\gamma(0) = \xi$. These abstract concepts simplify considerably if $\mathcal{M}$ inherits its Riemannian structure from an embedding Euclidean space V. Thus, (2.5) and (2.6) take the explicit form

$$\boldsymbol{\nabla}_{\dot\gamma(t)} \mathcal{X}(t) = \mathrm{proj}_{\gamma(t)}\big(\tfrac{d}{dt}\mathcal{X}(t)\big), \quad \boldsymbol{\nabla}_{\dot\gamma(t)} \dot\gamma(t) = \mathrm{proj}_{\gamma(t)}(\ddot\gamma(t)),$$

where proj_P and $\mathrm{proj}_{\gamma(t)}$ stand for the orthogonal projection onto $\mathrm{T}_P\mathcal{M}$ and $\mathrm{T}_{\gamma(t)}\mathcal{M}$, respectively.

2.1.3 The Riemannian gradient and Hessian

If $(\mathcal{M}, g)$ is a Riemannian manifold and $f : \mathcal{M} \to \mathbb{R}$ a smooth objective function on $\mathcal{M}$, then, the *Riemannian gradient* at $p \in \mathcal{M}$ is defined as the unique tangent vector $\mathrm{grad}f(p) \in \mathrm{T}_p\mathcal{M}$ satisfying

$$\mathrm{d}_p f(\xi) = g_p(\mathrm{grad}f(p), \xi) \tag{2.7}$$

for all $\xi \in \mathrm{T}_p\mathcal{M}$, where $\mathrm{d}_p f$ denotes the differential (tangent map) of f at p.

The *Riemannian Hessian* of f at p is the self-adjoint operator $\mathbf{H}_f(p) : \mathrm{T}_p\mathcal{M} \to \mathrm{T}_p\mathcal{M}$ defined by

$$\mathbf{H}_f(p)\xi = \boldsymbol{\nabla}_\xi \mathrm{grad}f(p), \tag{2.8}$$

for all $\xi \in \mathrm{T}_p\mathcal{M}$.

When $\mathcal{M}$ is a Riemannian submanifold of a vector space, then the gradient and the Hessian are given by the following:

$$\mathrm{grad}f(p) = \mathrm{proj}_p(\nabla \widetilde{f}(p)), \quad \mathbf{H}_f(p)\xi = \mathrm{proj}_p(D\widetilde{\mathcal{X}}(p)\xi), \tag{2.9}$$

where proj_p is the orthogonal projection onto $\mathrm{T}_p\mathcal{M}$ and $\nabla \widetilde{f}$ denotes the standard gradient of the smooth extension $\widetilde{f}$ of f on V.

2.2 Tensor products of manifolds

In this section we study the structure of a special subset of the tensor product of vector spaces, i.e. the set of simple tensors which satisfy some properties. In special, we are interested when can such a set be endowed with a manifold structure and more important with a submanifold structure of the tensor product space. We establish a bijection between our set of simple tensors and a quotient space. From the theory of Lie-group actions we obtain sufficient conditions such that the quotient space can have a differentiable structure. Furthermore, by additionally imposing some transversality conditions, we conclude that the set of simple tensors can be equipped with a submanifold structure of the tensor product vector space.

2.2.1 Tensor product of vector spaces

In this section, we recall fundamental objects and concepts from multilinear algebra (see [28, 49] for more details). Let $V_1, \ldots, V_r$ be vector spaces over $\mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$). The *tensor product of vector spaces* $V_1, \ldots, V_r$ is defined as the pair $(V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r, \otimes_{\mathbb{K}})$ of a $\mathbb{K}$-vector space $V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ and a multilinear map $\otimes_{\mathbb{K}} : V_1 \times \cdots \times V_r \to V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ that satisfies the universal property:

For any $\mathbb{K}$-vector space V and any multilinear map $h : V_1 \times \cdots \times V_r \to V$, there exists a unique linear map $\widetilde{h} : V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r \to V$ such that the diagram

$$\begin{tikzcd} V_1 \times \cdots \times V_r \arrow[r, "\otimes_{\mathbb{K}}"] \arrow[dr, dashed, "h"'] & V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r \arrow[d, "\widetilde{h}"] \\ & V \end{tikzcd} \tag{2.10}$$

commutes, i.e., $h(X_1, \ldots, X_r) = \widetilde{h}(X_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} X_r)$, for all $(X_1, \ldots, X_r) \in V_1 \times \cdots \times V_r$.

Given bases $B_1, \ldots, B_r$ of $V_1, \ldots, V_r$ respectively, the set

$$\{b_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} b_r \mid b_j \in B_j,\ j = 1, \ldots, r\}$$

is a basis for $V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ and hence the tensor product space has dimension equal to the product of dimensions of vector spaces $V_1, \ldots, V_r$. The elements of $V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ are called *tensors* and the elements of the form $X_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} X_r$ are called *pure tensors* or *simple tensors*. Any tensor X is a finite linear combination of simple ones and the smallest number of simple tensors required to express it is called the *tensor rank* of X. The *tensor order* refers to the number of spaces involved in the tensor product, i.e. in our notation the order of a tensor $X \in V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ is always r.

For simple tensors we give the following result, that we are going to use in the next sections to define an equivalence relation on a direct product of manifolds.

Lemma 2.2.1 *Let $V_1 \otimes_{\mathbb{K}} V_2$ be the $\mathbb{K}$-tensor product of the $\mathbb{K}$-vector spaces V_1 and V_2. For $A_1, B_1 \in V_1 \setminus \{0\}$ and $A_2, B_2 \in V_2 \setminus \{0\}$ the following holds: $A_1 \otimes_{\mathbb{K}} A_2 = B_1 \otimes_{\mathbb{K}} B_2$ if and only if there exists $\alpha_1, \alpha_2 \in \mathbb{K}$ such that*

$$B_1 = \alpha_1 A_1, \quad B_2 = \alpha_2 A_2, \quad \alpha_1 \cdot \alpha_2 = 1.$$

Proof. If there exist $\alpha_1,\ \alpha_2 \in \mathbb{K}$ with $\alpha_1\alpha_2 = 1$ such that $B_1 = \alpha_1 A_1$ and $B_2 = \alpha_2 A_2$, then we get immediately by the bilinearity of $\otimes_{\mathbb{K}}$ that

$$B_1 \otimes_{\mathbb{K}} B_2 = \alpha_1 A_1 \otimes_{\mathbb{K}} \alpha_2 A_2 = \alpha_1\alpha_2 A_1 \otimes_{\mathbb{K}} A_2 = A_1 \otimes_{\mathbb{K}} A_2.$$

To prove the other direction, assume that there exist $A_1, B_1 \in V_1$ and $A_2, B_2 \in V_2$ with $A_1 \otimes A_2 = B_1 \otimes B_2$ and

$$(B_1, B_2) \neq (\alpha_1 A_1, \alpha_2 A_2),$$

for all $\alpha_1,\ \alpha_2 \in \mathbb{K}$ with $\alpha_1\alpha_2 = 1$. We prove that there exists a $\mathbb{K}$-vector space V and a bilinear map $h : V_1 \times V_2 \to V$ such that

$$h(B_1, B_2) \neq h(\alpha_1 A_1, \alpha_2 A_2), \tag{2.11}$$

for all $\alpha_1,\ \alpha_2 \in \mathbb{K}$ with $\alpha_1\alpha_2 = 1$. We distinguish the following situations.

Case 1. Let $B_1 \neq \alpha_1 A_1$, for all $\alpha_1 \in \mathbb{K}$. If $h : V_1 \times V_2 \to V$ is the bilinear map defined by

$$h(X_1, X_2) = \mu_1(X_1)\mu_2(X_2), \tag{2.12}$$

where $\mu_1 : V_1 \to V$ and $\mu_2 : V_2 \to V$ are linear maps such that $\mu_1(B_1) \neq 0$ and $\mu_1(A_1) = 0$, it follows that

$$h(B_1, B_2) = \mu_1(B_1)\mu_2(B_2) \neq 0 = \mu_1(A_1)\mu_2(A_2) = h(\alpha_1 A_1, \alpha_2 A_2).$$

Case 2. Similarly, it can be proved that if $B_2 \neq \alpha_2 A_2$, for all $\alpha_2 \in \mathbb{K}$, then the bilinear map h defined by (2.12) with $\mu_2(A_2) = 0$ and $\mu_2(B_2) \neq 0$, satisfies (2.11).

From the universal property of $(V_1 \otimes_{\mathbb{K}} V_2, \otimes_{\mathbb{K}})$ we conclude that there exists a unique linear map $\widetilde{h} : V_1 \otimes_{\mathbb{K}} V_2 \to V$ such that the diagram (2.10) commutes. Hence,

$$h(B_1, B_2) \neq h(\alpha_1 A_1, \alpha_2 A_2) \Longrightarrow \widetilde{h}(B_1 \otimes_{\mathbb{K}} B_2) \neq \widetilde{h}(A_1 \otimes_{\mathbb{K}} A_2)$$

and $A_1 \otimes_{\mathbb{K}} A_2 \neq B_1 \otimes_{\mathbb{K}} B_2$, which contradicts the hypothesis. $\square$

Remark 2.2.2 *If V_1 and V_2 are $\mathbb{C}$-vector spaces and we construct their $\mathbb{R}$-tensor product $V_1 \otimes_{\mathbb{R}} V_2$, then for any $A_1, B_1 \in V_1 \setminus \{0\}$ and $A_2, B_2 \in V_2 \setminus \{0\}$ we have that $A_1 \otimes A_2 = B_1 \otimes B_2$ if and only if there exists $\alpha_1, \alpha_2 \in \mathbb{R}$ such that*

$$B_1 = \alpha_1 A_1, \quad B_2 = \alpha_2 A_2, \quad \alpha_1 \cdot \alpha_2 = 1.$$

The above lemma can be generalized to any number of components in the $\mathbb{K}$-tensor product: for $X = (X_1, \ldots, X_r)$ and $Y = (Y_1, \ldots, Y_r)$ from $(V_1 \setminus \{0\}) \times \cdots \times (V_r \setminus \{0\})$, $X_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} X_r = Y_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} Y_r$ if and only if

$$\exists\, \alpha_1, \ldots, \alpha_r \in \mathbb{K} \text{ with } \alpha_1\alpha_2\cdots\alpha_r = 1 \text{ such that } Y_1 = \alpha_1 X_1, \ldots, Y_r = \alpha_r Y_r. \tag{2.13}$$

We point out some examples of tensor products over $\mathbb{K}$.

1. A classical example of a tensor product over $\mathbb{K}$ is obtained from the $\mathbb{K}$-vector space $\mathbb{K}^{mp\times nq}$ with the bilinear map

$$\otimes_{\mathbb{K}} : \mathbb{K}^{m\times n} \times \mathbb{K}^{p\times q} \to \mathbb{K}^{mp\times nq},$$

defined as

$$(X, Y) \mapsto X \otimes_{\mathbb{K}} Y := [x_{ij}Y]_{i,j=1}^{m,n} = \begin{bmatrix} x_{11}Y & \cdots & x_{1n}Y \\ \vdots & & \\ x_{m1}Y & \cdots & x_{mn}Y \end{bmatrix},$$

where

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}.$$

This tensor product is called the *matrix Kronecker product.* Here we recall some of its important properties which will be used later on. Let $A \in K^{m\times n}$, $B \in \mathbb{K}^{p\times q}$, $C \in \mathbb{K}^{n\times s}$, $D \in \mathbb{K}^{q\times t}$, then

$$(A \otimes_{\mathbb{K}} B)(C \otimes_{\mathbb{K}} D) = AC \otimes_{\mathbb{K}} BD, \quad \mathrm{tr}(A \otimes_{\mathbb{K}} B) = \mathrm{tr}(A)\,\mathrm{tr}(B).$$

2. In the case when $\mathbb{K} = \mathbb{C}$, By considering $\mathbb{C}^{m\times n}$ and $\mathbb{C}^{p\times q}$ as real vector spaces, then

$$\begin{aligned} \otimes_{\mathbb{R}} : \mathbb{C}^{m\times n} \times \mathbb{C}^{p\times q} &\to \mathbb{C}^{mp\times nq} \times \mathbb{C}^{mp\times nq}, \\ (X, Y) &\mapsto X \otimes_{\mathbb{R}} Y := ([x_{ij}Y], [x_{ij}\bar{Y}]), \end{aligned} \tag{2.14}$$

defines an $\mathbb{R}$-tensor product. In fact, for any $\mathbb{C}$-vector spaces V_1 and V_2, the $\mathbb{R}$-bilinear map (2.14) defines a tensor product on the $\mathbb{R}$-vector space $V_1 \otimes_{\mathbb{R}} V_2$.

2.2.2 Simple tensors and quotient manifolds

Let $\mathcal{M}_1, \ldots, \mathcal{M}_r$ real submanifolds of the finite dimensional $\mathbb{K}$-inner product spaces $(V_1, \langle\cdot,\cdot\rangle_1) \ldots, (V_r, \langle\cdot,\cdot\rangle_r)$ with $0 \notin \mathcal{M}_j,\ j = 1, \ldots, r$. The *$r$-fold tensor product of the manifolds* $\mathcal{M}_1, \ldots, \mathcal{M}_r$ is defined as the set of simple tensors

$$\mathcal{M}^{\otimes_{\mathbb{K}}} := \mathcal{M}_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} \mathcal{M}_r := \{X_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} X_r \mid X_j \in \mathcal{M}_j,\ j = 1, \ldots, r\}. \tag{2.15}$$

In the sequel we construct a quotient space which is in a one-to-one correspondence with $\mathcal{M}^{\otimes_{\mathbb{K}}}$ and use basic concepts from the theory of quotient manifolds to give sufficient conditions for the quotient space to have a differentiable structure.

On the manifold

$$\mathcal{M}^{\times} := \mathcal{M}_1 \times \cdots \times \mathcal{M}_r, \tag{2.16}$$

we naturally consider the following equivalence relation

$$X \sim Y :\Longleftrightarrow X_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} X_r = Y_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} Y_r, \tag{2.17}$$

for all $X := (X_1, \ldots, X_r),\ Y := (Y_1, \ldots, Y_r) \in \mathcal{M}^{\times}$, and denote the set of all equivalence classes

$$[X] := \{(Y_1, \ldots, Y_r) \in \mathcal{M}^{\times} \mid X_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} X_r = Y_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} Y_r\}$$

with $\mathcal{M}^{\times}/_{\sim}$. According to Theorem 2.1.3, the quotient space $\mathcal{M}^{\times}/_{\sim}$ carries the structure of a quotient manifold if the graph of the equivalence relation

$$\Gamma_{\sim} = \{(X, Y) \mid X \sim Y\} \subset \mathcal{M}^{\times} \times \mathcal{M}^{\times} \tag{2.18}$$

is a closed submanifold of $\mathcal{M}^{\times} \times \mathcal{M}^{\times}$ and the projection onto the first component

$$\pi_1 : \Gamma_{\sim} \to \mathcal{M}^{\times}, \quad (X, Y) \mapsto X \tag{2.19}$$

is a submersion. Since $\Gamma_{\sim}$ is the preimage of $\{0\} \subset V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ under the continuous map

$$\mathcal{M}^{\times} \times \mathcal{M}^{\times} \to V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r, \quad (X, Y) \mapsto X_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} X_r - Y_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} Y_r,$$

it is clearly a closed subset of $\mathcal{M}^{\times} \times \mathcal{M}^{\times}$. However, it does not always have to be a submanifold of $\mathcal{M}^{\times} \times \mathcal{M}^{\times}$, as the following counterexample shows.

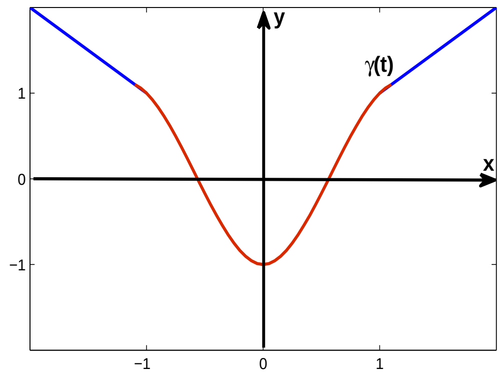


Figure 2.1: The graph of the function γ as a submanifold of $\mathbb{R}^2$.

Example 2.2.3 Let $\mathcal{M}_1 := \mathbb{R} \setminus \{0\}$ and $\mathcal{M}_2$ be the one-dimensional submanifold of $\mathbb{R}^2$ defined by the graph of the curve

$$\gamma : \mathbb{R} \to \mathbb{R},\ \gamma(t) = \begin{cases} t, & |t| \geq 1 \\ f(t), & |t| < 1 \end{cases}, \tag{2.20}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a smooth function with $f(0) = -1$ such that γ is smooth, see Figure 2.1. For $X_2 := (2,\ 2)^\top \in \mathcal{M}_2$, one notices that $\alpha X_2 \in \mathcal{M}_2$ for all $\alpha \in \mathbb{R}$ with $\alpha \geq 1/2$. On the other hand, by taking $\widetilde{X}_2 := (0,\ -1)^\top \in \mathcal{M}_2$, one has $\alpha \widetilde{X}_2 \in \mathcal{M}_2$ if and only if $\alpha = 1$. Consequently, for fixed $X_1 \in \mathcal{M}_1$ the following hold:

$$((X_1, X_2), (\frac{1}{\alpha} X_1, \alpha X_2)) \in \Gamma_\sim \text{ for all } \alpha \geq 1/2$$

and

$$((X_1, \widetilde{X_2}), (\frac{1}{\beta} X_1, \beta \widetilde{X_2})) \in \Gamma_\sim \text{ if and only if } \beta = 1.$$

Let $\varphi_1 : U_1 \subset \mathcal{M}_1 \to \mathbb{R}$ and $\varphi_2 : U_2 \subset \mathcal{M}_2 \to \mathbb{R}$ be the charts of $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively. Then, in a neighborhood of $((X_1, X_2), (X_1, X_2)) \in \Gamma_\sim$ the map

$$((Y_1, Y_2), (Z_1, Z_2)) \mapsto \left(\varphi_1(Y_1), \varphi_2(Y_2), \frac{\langle Z_1, Y_1 \rangle}{\|Y_1\|^2}\right)$$

defines a homeomorphism to an open subset in $\mathbb{R}^3$, and in a neighborhood of $((X_1, \widetilde{X_2}), (X_1, \widetilde{X_2})) \in \Gamma_\sim$ the map

$$((Y_1, Y_2), (Z_1, Z_2)) \mapsto \left(\varphi_1(Y_1), \varphi_2(Y_2)\right)$$

defines a homeomorphism to an open subset in $\mathbb{R}^2$.

Our goal is to impose some conditions on the manifolds $\mathcal{M}_1, \dots, \mathcal{M}_r$ such that $\Gamma_\sim$ is a submanifold of $\mathcal{M}^\times \times \mathcal{M}^\times$. To this end, we reformulate the equivalence relation $\sim$ in terms of a Lie-group action on the vector space $V_1 \times \cdots \times V_r$ and use a well-known result from differential geometry due to Dieudonne ([14], page 60), stated as Theorem 2.2.4. First recall that a *Lie-group* G is a group which has a differentiable structure such that the group operation and inversion are compatible with the smooth structure. A *Lie-subgroup* A of a Lie-group G is a closed subgroup of G, equipped with the structure of a Lie-group. A Lie-group G acts on a manifold $\mathcal{M}$ if there exists a smooth map

$$\sigma : G \times \mathcal{M} \to \mathcal{M}, \quad (\alpha, X) \mapsto \alpha \cdot x \tag{2.21}$$

that satisfies

$$\alpha \cdot (\beta \cdot X) = (\alpha\beta) \cdot X, \quad e \cdot X = X, \tag{2.22}$$

where e is the identity element in G. The *graph map* associated to σ is the map

$$\widehat{\sigma} : G \times \mathcal{M} \to \mathcal{M} \times \mathcal{M}, \quad (\alpha, X) \mapsto (X, \alpha \cdot X). \tag{2.23}$$

For a thorough discussion on Lie-groups and their actions we refer to the literature [35, 45]. In the case of Lie-groups acting on a manifold, the statement of Theorem 2.1.3 simplifies as follows.

Theorem 2.2.4 *Let $\mathcal{M}$ be a differentiable manifold and G a Lie-group which acts on $\mathcal{M}$. Then, there exists a unique manifold structure on $\mathcal{M}/G$ such that the canonical projection π is a submersion if and only if the image of the graph map $\widehat{\sigma}$ is a closed submanifold of* $\mathcal{M} \times \mathcal{M}$.

Using Lemma 2.2.1, in what follows, we rewrite the equivalence relation $\sim$ on $\mathcal{M}^\times$ defined by (2.17), as the restriction of a Lie-group action on the $\mathbb{K}$-vector space $V_1 \times \cdots \times V_r$. For this, we define the set

$$G := \{\alpha = (\alpha_1, \dots, \alpha_r) \in \mathbb{K}^r \mid \alpha_1 \cdot \alpha_2 \cdots \alpha_r = 1\}, \tag{2.24}$$

which is a Lie-group with respect to the subspace topology of $\mathbb{K}^r$ and the group operation

$$\alpha\beta = (\alpha_1\beta_1, \dots, \alpha_r\beta_r),$$

for all $\alpha = (\alpha_1, \dots, \alpha_r)$ and $\beta = (\beta_1, \dots, \beta_r)$ in G. The Lie-algebra of G is given by

$$\mathfrak{g} = \{\omega \in \mathbb{K}^r \mid \omega_1 + \cdots + \omega_r = 0\}. \tag{2.25}$$

Furthermore, the action of G on the vector space

$$V^\times := V_1 \times \cdots \times V_r \tag{2.26}$$

is given by

$$\sigma_G : G \times V^\times \to V^\times, \quad (\alpha, X) \mapsto \alpha \cdot X = (\alpha_1 X_1, \dots, \alpha_r X_r) \tag{2.27}$$

and it induces an equivalence relation on $V^\times$ as

$$X \sim_G Y \Longleftrightarrow \exists\, \alpha \in G \text{ such that } Y = \alpha \cdot X, \tag{2.28}$$

for $X, Y \in V_1 \times \cdots \times V_r$. The equivalence classes of $\sim_G$ are called *orbits*, i.e.

$$\mathcal{O}_G(X) = \{\alpha \cdot X \mid \alpha \in G\},$$

and the set of all equivalence classes $V^\times/G$ is referred to as the *orbit space*. For any $X \in (V_1 \setminus \{0\}) \times \cdots \times (V_r \setminus \{0\})$, the *stabilizer group* of X is the trivial subgroup

$$\operatorname{Stab}_G(X) = \{\alpha \in G \mid \alpha \cdot X = X\} = \{(1, \ldots, 1)\} \tag{2.29}$$

of G. In this case, we say that the group action σ_G is *free*.

With the above specifications, the graph $\Gamma_\sim$ of the equivalence relation $\sim$ defined in (2.17) can be written in terms of the graph of the equivalence relation $\sim_G$

$$\Gamma_G = \{(X, \alpha \cdot X) \mid X \in V^\times,\ \alpha \in G\} \subset V^\times \times V^\times$$

as

$$\Gamma_\sim = \Gamma_G \cap (\mathcal{M}^\times \times \mathcal{M}^\times). \tag{2.30}$$

There are several possible approaches one could follow to obtain conditions to ensure that $\Gamma_\sim$ is a submanifold of $\mathcal{M}^\times \times \mathcal{M}^\times$. One of these approaches involves the transversality theory. From differential topology it is known that if two submanifolds S_1 and S_2 of a manifold $\mathcal{M}$ *intersect tranversally*, i.e.

$$\mathrm{T}_p S_1 + \mathrm{T}_p S_2 = \mathrm{T}_p \mathcal{M}, \tag{2.31}$$

for all $p \in S_1 \cap S_2$, then their intersection is also a submanifold (see [29, 36, 55]). However, it is clear from Example 2.2.3 that this is not the case in our situation, i.e. (2.30), hence we will not follow this path. Instead, we will concentrate on the situation when the equivalence relation $\sim$ is induced by the action of a Lie-subgroup of G.

Next, we show that for any Lie-subgroup $\mathcal{A}$ of G which acts on the manifold $\mathcal{M}^\times$ according to the group action

$$\sigma_{\mathcal{A}} : \mathcal{A} \times \mathcal{M}^\times \to \mathcal{M}^\times, \quad (\alpha, X) \mapsto \alpha \cdot X = (\alpha_1 X_1, \ldots, \alpha_r X_r), \tag{2.32}$$

the image of the graph map $\widehat{\sigma}_{\mathcal{A}}$ is a closed submanifold of $\mathcal{M}^\times \times \mathcal{M}^\times$.

Theorem 2.2.5 *Let $\mathcal{M}_j$ be real submanifolds of the $\mathbb{K}$-inner product spaces $(V_j, \langle \cdot, \cdot \rangle_j)$, for $j = 1, \ldots, r$ and G be the Lie-group defined by* (2.24). *If $\mathcal{A}$ is a Lie-subgroup of G which acts on $\mathcal{M}^\times := \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ according to $\sigma_{\mathcal{A}}$ defined by* (2.32), *then the graph $\Gamma_{\mathcal{A}}$ of the equivalence relation induced by $\sigma_{\mathcal{A}}$ is a closed submanifold of $\mathcal{M}^\times \times \mathcal{M}^\times$. Moreover, $\mathcal{M}^\times/\mathcal{A}$ is a manifold of dimension*

$$\dim(\mathcal{M}^\times/\mathcal{A}) = \dim \mathcal{M}^\times - \dim \mathcal{A}.$$

Proof. Since $\mathrm{Stab}_G(X) = \{(1,\dots,1)\}$ for any $X \in \mathcal{M}^\times$ and $\mathcal{A}$ is a Lie-subgroup of G, it results that the group action $\sigma_\mathcal{A}$ is free. To apply Theorem 2.2.4 we need to prove that the graph map is proper. The graph map

$$\widehat{\sigma}_\mathcal{A} : \mathcal{A} \times \mathcal{M}^\times \to \mathcal{M}^\times \times \mathcal{M}^\times, \quad (\alpha, X) = (X, \alpha \cdot X)$$

is injective, since the group action is free. Moreover, $\widehat{\sigma}_\mathcal{A}$ is a homeomorphism onto its image, with the inverse map

$$\varphi : \mathcal{M}^\times \times \mathcal{M}^\times \to G \times \mathcal{M}^\times, \quad \varphi(X,Y) = \left(\tfrac{\langle X_1, Y_1 \rangle_1}{\|X_1\|_1^2}, \dots, \tfrac{\langle X_r, Y_r \rangle_r}{\|X_r\|_r^2}, X \right),$$

is continuous and has the property that

$$\varphi|_{\mathrm{Im}\,\widehat{\sigma}_\mathcal{A}} = \widehat{\sigma}_\mathcal{A}^{-1}.$$

Hence, $\widehat{\sigma}_\mathcal{A}^{-1}$ is continuous, and moreover $\widehat{\sigma}_\mathcal{A}$ is proper. From Theorem 2.2.4 it follows that $\sigma_\mathcal{A}$ is a regular group action. □

From the above theorem, we can conclude that the set Γ_G is a submanifold of $V^\times \times V^\times$. Thus, if the group G acts on $\mathcal{M}^\times$, then $\Gamma_G \subset \mathcal{M}^\times \times \mathcal{M}^\times$ and thus, $\mathcal{M}^\times/_\sim = \mathcal{M}^\times/G$ is a manifold. More generally, we have the following sufficient condition for the quotient space $\mathcal{M}^\times/_\sim$ to be a manifold.

Corollary 2.2.6 *Let $\mathcal{M}_j$ be real submanifolds of the $\mathbb{K}$-inner product spaces $(V_j, \langle \cdot, \cdot \rangle_j)$, for $j = 1, \dots, r$ and G be the Lie-group defined by (2.24). If there exists a Lie-subgroup $\mathcal{A}$ of G which acts on $\mathcal{M}^\times := \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ according to $\sigma_\mathcal{A}$ defined by (2.32) and*

$$\mathcal{M}^\times/_\sim = \mathcal{M}^\times/\mathcal{A},$$

then, the quotient space $\mathcal{M}^\times/_\sim$ has a unique manifold structure such that the canonical projection is a submersion.

We enclose this section with some examples of manifolds $\mathcal{M}_1, \dots, \mathcal{M}_r$ for which $\mathcal{M}^\times/_\sim$ is a quotient manifold.

Example 2.2.7 If $\mathcal{M}_j = V_j \setminus \{0\}$, for all $j = 1, \dots, r$, then (2.27) defines a group action of the Lie-group G on $\mathcal{M}^\times$. According to Theorem 2.2.5 it follows that $(\mathcal{M}_1 \times \cdots \times \mathcal{M}_r)/G$ is a manifold.

Example 2.2.8 A similar situation is encountered when the product manifold $\mathcal{M}^\times$ is $\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_r}$, where GL_{n_j} is the Lie-group of all invertible $n_j \times n_j$ matrices with entries in $\mathbb{K}$. Here, the restriction of the map σ_G to $\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_r}$ defines a group action of G on $\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_r}$ and hence $(\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_r})/G$ is a manifold.

In both examples, the graph of the equivalence relation induced by σ_G is equal to $\Gamma_\sim$ and hence, $(V_1 \setminus \{0\} \times \cdots \times V_r \setminus \{0\})/_\sim$ and $(\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_r})/_\sim$ are manifolds.

Example 2.2.9 Let $\mathcal{M}_j$ be the manifold of all elements $X_j \in V_j$ of equal norm, i.e.,

$$\mathcal{M}_j = \{X_j \in V_j \mid \|X_j\|_j = c_j,\ c_j > 0\},$$

for $j = 1, \dots, r$.

We will show that $\mathcal{M}^\times/_\sim$ is a manifold, where $\sim$ is the equivalence relation defined by (2.17). On the $\mathbb{K}$-vector space $V_1 \times \cdots \times V_r$ one has the inner product

$$\langle X, Y\rangle = \langle X_1, Y_1\rangle_1 + \cdots + \langle X_r, Y_r\rangle_r,$$

for all $X = (X_1, \dots, X_r),\ Y = (Y_1, \dots, Y_r) \in V_1 \times \cdots \times V_r$. Moreover, for any $X,\ Y \in \mathcal{M}^\times$ one has

$$X \sim Y \iff \exists\, \alpha \in G,\ |\alpha_1| = \cdots = |\alpha_r| = 1 \text{ such that } Y = \alpha \cdot X.$$

The set

$$\mathcal{A} = \{\alpha \in \mathbb{K}^r \mid |\alpha_1| = \cdots = |\alpha_r| = 1,\ \alpha_1\alpha_2\cdots\alpha_r = 1\}$$

is a Lie-subgroup of G and

$$\sigma_{\mathcal{A}} : \mathcal{A} \times \mathcal{M}^\times \to \mathcal{M}^\times,\ (\alpha, X) \mapsto \alpha \cdot X$$

is the Lie-group action which induces the equivalence relation $\sim$, hence, $\mathcal{M}^\times/_\sim$ is a manifold.

2.2.3 Tangent space, Riemannian metric and Levi-Civita connection on the orbit space $\mathcal{M}^\times/\mathcal{A}$

Let $(V_j, \langle\cdot,\cdot\rangle_j)$ be $\mathbb{K}$-inner product spaces, and $(\mathcal{M}_j, g_j)$ be Riemannian submanifolds of V_j, i.e.

$$g_j(\xi_j, \eta_j) = \mathrm{Re}\langle \xi_j, \eta_j\rangle_j, \tag{2.33}$$

for all $\xi_j, \eta_j \in \mathrm{T}_{X_j}\mathcal{M}_j,\ X_j \in \mathcal{M}_j,\ j = 1, \dots, r$. We endow the direct product manifold $\mathcal{M}^\times := \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ with the Riemannian metric

$$g_X(\xi, \eta) = \sum_{j=1}^{r} \frac{1}{\|X_j\|_j^2} g_j(\xi_j, \eta_j) = \sum_{j=1}^{r} \frac{\mathrm{Re}\langle \xi_j, \eta_j\rangle_j}{\|X_j\|_j^2} \tag{2.34}$$

for all $X = (X_1, \dots, X_r) \in \mathcal{M}^\times$, $\xi = (\xi_1, \dots, \xi_r)$, $\eta = (\eta_1, \dots, \eta_r) \in \mathrm{T}_X\mathcal{M}^\times$, where $\|X_j\|_j^2 = \langle X_j, X_j\rangle_j$.

In this section, we describe the Riemannian structure of the quotient manifold $\mathcal{M}^\times/\mathcal{A}$, for a Lie-subgroup $\mathcal{A}$ of G defined by (2.24), which acts on $\mathcal{M}^\times$ according to the group action (2.32). The tangent vectors of $\mathcal{M}^\times/\mathcal{A}$ have a unique representation as vectors in the horizontal space. To avoid confusion we will use $[X]$ to refer to elements in the quotient manifold $\mathcal{M}^\times/\mathcal{A}$ and $\mathcal{O}_{\mathcal{A}}(X)$ for subsets in the manifold $\mathcal{M}^\times$. The vertical space at $Y \in \mathcal{O}_{\mathcal{A}}(X)$ is the tangent space of $\mathcal{O}_{\mathcal{A}}(X)$ at Y and is given by

$$\mathrm{T}_Y\mathcal{O}_{\mathcal{A}}(X) = \{\omega \cdot Y \mid \omega \in \mathfrak{a}\} \subset \mathrm{T}_Y\mathcal{M}^\times, \tag{2.35}$$

where $\mathfrak{a}$ is the Lie-algebra of $\mathcal{A}$. The horizontal space at Y is the orthogonal complement of $\mathrm{T}_Y\mathcal{O}_{\mathcal{A}}(X)$ in $\mathrm{T}_Y\mathcal{M}^\times$ with respect to the Riemannian metric (2.34) and is given by

$$\begin{aligned}\mathcal{H}_Y &:= \{\xi \in \mathrm{T}_Y\mathcal{M}^\times \mid g_Y(\xi,\eta) = 0,\ \forall\, \eta \in \mathrm{T}_Y\mathcal{O}_{\mathcal{A}}(X)\} \\ &= \{\xi \in \mathrm{T}_Y\mathcal{M}^\times \mid g_Y(\xi, \omega\cdot Y) = 0,\ \forall\, \omega \in \mathfrak{a}\}.\end{aligned} \tag{2.36}$$

A representation for a tangent vector $\xi \in \mathrm{T}_{[X]}\mathcal{M}^\times/\mathcal{A}$ is given by a vector ξ_Y^h in the horizontal space $\mathcal{H}_Y$ with $Y \in \mathcal{O}_{\mathcal{A}}(X)$ that satisfies

$$\mathrm{T}_Y\pi_{\mathcal{A}}(\xi_Y^h) = \xi.$$

This representation, called horizontal lift, is unique up to a choice of a representative $Y \in \mathcal{O}_{\mathcal{A}}(X)$.

The following relation between the horizontal lifts of the tangent vectors of elements in the same equivalence class will help us to introduce a Riemannian metric on the manifold $\mathcal{M}^\times/\mathcal{A}$.

Lemma 2.2.10 *Let $\mathcal{A}$ be any Lie-subgroup of G that acts on $\mathcal{M}^\times$ according the group action* (2.32)*. Then, for any $X \in \mathcal{M}^\times$ and $\alpha \in \mathcal{A}$,*

$$\xi^h_{\alpha\cdot X} = \alpha\cdot\xi^h_X, \tag{2.37}$$

where $\xi^h_X \in \mathcal{H}_X$ denotes the horizontal component of $\xi_X \in \mathrm{T}_X\mathcal{M}^\times$. Moreover,

$$g_X(\xi^h_X, \eta^h_X) = g_{\alpha\cdot X}(\xi^h_{\alpha\cdot X}, \eta^h_{\alpha\cdot X}), \tag{2.38}$$

for all $\xi^h_X,\ \eta^h_X \in \mathcal{H}_X$ and $\alpha \in \mathcal{A}$.

Proof. Let $X \in \mathcal{M}^\times$ be a representative for $[X] \in \mathcal{M}^\times/\mathcal{A}$, and let $\xi \in \mathrm{T}_{[X]}\mathcal{M}^\times/\mathcal{A}$. Then, the horizontal lift $\xi^h_X \in \mathcal{H}_X$ of ξ satisfies

$$T_X\pi_{\mathcal{A}}(\xi^h_X) = \xi.$$

Fix $\alpha \in \mathcal{A}$ and consider the function $f_\alpha : \mathcal{M}^\times \to \mathcal{M}^\times$ defined as $f_\alpha(X) = \alpha\cdot X$. Then, $\pi_{\mathcal{A}}(f_\alpha(X)) = \pi_{\mathcal{A}}(X)$ and hence

$$\xi \ = T_X\pi_{\mathcal{A}}(\xi^h_X) = T_X(\pi_{\mathcal{A}}\circ f_\alpha)(\xi^h_X) = T_{\alpha\cdot X}\pi_{\mathcal{A}}(T_Xf_\alpha(X)(\xi^h_X)) = T_{\alpha\cdot X}\pi_{\mathcal{A}}(\alpha\cdot\xi^h_X).$$

Moreover,

$$\begin{aligned} g_{\alpha\cdot X}(\alpha\cdot\xi^h_X, \omega\cdot(\alpha\cdot X)) &= \sum_{j=1}^r \frac{1}{|\alpha_j|^2\|X_j\|_j^2}\,\mathrm{Re}\langle\alpha_j\xi^h_j, \omega_j\alpha_jX_j\rangle_j \\ &= \sum_{j=1}^r \frac{|\alpha_j|^2}{|\alpha_j|^2\|X_j\|^2}\,\mathrm{Re}\langle\xi^h_j, \omega_jX_j\rangle_j \\ &= g_X(\xi^h_X, \omega\cdot X) = 0,\end{aligned}$$

for all $\omega \in \mathfrak{a}$. Hence, from the definition of the horizontal space (2.36) it follows that $\alpha \cdot \xi_X^h \in \mathcal{H}_{\alpha \cdot X}$. The conclusion (2.37) follows from the fact that

$$T_{\alpha \cdot X}\pi_{\mathcal{A}}(\xi_{\alpha \cdot X}^h) = \xi = T_{\alpha \cdot X}\pi_{\mathcal{A}}(\alpha \cdot \xi_X^h) \quad \text{and} \quad \alpha \cdot \xi_X^h \in \mathcal{H}_{\alpha \cdot X}.$$

To prove (2.38), let $\xi_{\alpha \cdot X}^h,\ \eta_{\alpha \cdot X}^h \in \mathcal{H}_{\alpha \cdot X}$. Then,

$$\begin{aligned} g_{\alpha \cdot X}(\xi_{\alpha \cdot X}^h, \eta_{\alpha \cdot X}^h) &= g_{\alpha \cdot X}(\alpha \cdot \xi_X^h, \alpha \cdot \eta_X^h) \\ &= \sum_{j=1}^r \frac{1}{|\alpha_j|^2 \|X_j\|_j^2} \operatorname{Re}\langle \alpha_j \xi_j^h, \alpha_j \eta_j^h \rangle_j \\ &= \sum_{j=1}^r \frac{1}{\|X_j\|^2} \operatorname{Re}\langle \xi_j^h, \eta_j^h \rangle_j \\ &= g_X(\xi_X^h, \eta_X^h), \end{aligned}$$

where $\xi_X^h = (\xi_1^h, \ldots, \xi_r^h)$, $\eta_X^h = (\eta_1^h, \ldots, \eta_r^h) \in \mathcal{H}_X$. □

The above lemma says that, for all $[X] \in \mathcal{M}^\times/\mathcal{A}$ and all $\xi,\ \eta \in \mathrm{T}_{[X]}\mathcal{M}^\times/\mathcal{A}$, the expression $g_Y(\xi_Y^h, \eta_Y^h)$ does not depend on the choice of the representative $Y \in \mathcal{O}_{\mathcal{A}}(X)$. Hence,

$$\widetilde{g}_{[X]}(\xi, \eta) := g_Y(\xi_Y^h, \eta_Y^h) \tag{2.39}$$

defines a *Riemannian metric on the orbit space* $\mathcal{M}^\times/\mathcal{A}$. The vectors $\xi_Y^h,\ \eta_Y^h \in \mathcal{H}_Y$ are the unique vectors in the horizontal space $\mathcal{H}_Y$ such that $\mathrm{T}_Y\pi_{\mathcal{A}}(\xi_Y^h) = \xi$ and $\mathrm{T}_Y\pi_{\mathcal{A}}(\eta_Y^h) = \eta$.

According to (2.3), on the Riemannian manifold $(\mathcal{M}^\times, g)$ the Levi-Civita connection is given by

$$\boldsymbol{\nabla}_\xi \mathcal{X} = \operatorname{proj}_X \Big(D\widetilde{\mathcal{X}_1}(X_1)(\xi_1), \ldots, D\widetilde{\mathcal{X}_r}(X_r)(\xi_r) \Big)$$

for all $\xi = (\xi_1, \ldots, \xi_r) \in \mathrm{T}_X\mathcal{M}^\times$ and all vector fields $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_r) : \mathcal{M}^\times \to \mathrm{T}\mathcal{M}^\times$, where $\widetilde{\mathcal{X}} = (\widetilde{\mathcal{X}}_1, \ldots, \widetilde{\mathcal{X}}_r)$ is a smooth extension of $\mathcal{X}$ to the vector space $V_1 \times \cdots \times V_r$ and proj_X is the orthogonal projector onto $\mathrm{T}_X\mathcal{M}^\times$. From (2.4) it follows that the Levi-Civita connection on $(\mathcal{M}^\times/\mathcal{A}, \widetilde{g})$ is defined as

$$\boldsymbol{\nabla}_\xi^h \mathcal{X}_{\mathcal{A}} = \boldsymbol{\nabla}_{\xi_X^h} \mathcal{X}^h = \operatorname{proj}_X^h \Big(D\widetilde{\mathcal{X}_1^h}(X_1)(\xi_1^h), \ldots, D\widetilde{\mathcal{X}_r^h}(X_r)(\xi_r^h) \Big),$$

for all $\xi \in \mathrm{T}_{[X]}\mathcal{M}^\times/\mathcal{A}$ and all smooth vector fields $\mathcal{X}_{\mathcal{A}} : \mathcal{M}^\times/\mathcal{A} \to \mathrm{T}\mathcal{M}^\times/\mathcal{A}$, where ξ_X^h is the horizontal lift of $\xi_{\mathcal{A}}$ and $\mathcal{X}^h$ is the vector field of the horizontal lift of $\mathcal{X}_{\mathcal{A}}$. Moreover, $\widetilde{\mathcal{X}^h}$ is a smooth extension of $\mathcal{X}^h$ to $V_1 \times \cdots \times V_r$ and proj_X^h is the orthogonal projector onto the horizontal space $\mathcal{H}_X$.

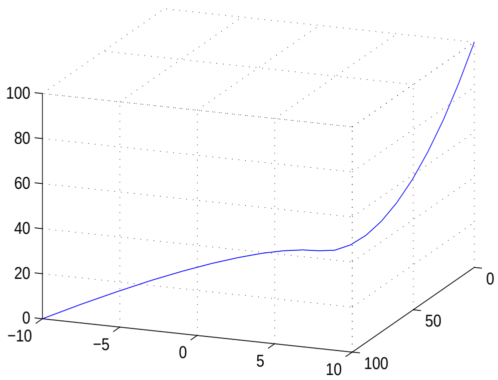


Figure 2.2: The graph of the smooth function f defined in Example 2.2.11.

2.2.4 Submanifold conditions for the r-fold tensor product of manifolds

Recall that $\mathcal{M}_1, \ldots, \mathcal{M}_r$ are Riemannian submanifolds of the $\mathbb{K}$-inner product spaces $V_1, \ldots, V_r$. Let $\mathcal{M}^\times$ denote the direct product manifold $\mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ equipped with the Riemannian metric (2.34), and $\sim$ the equivalence relation on $\mathcal{M}^\times$ given by (2.17). If the quotient space $\mathcal{M}^\times/_\sim$ has a manifold structure, then the one-to-one correspondence with $\mathcal{M}^\otimes$ induces a manifold structure on $\mathcal{M}^{\otimes_\mathbb{K}}$. The interesting question that arises is whether $\mathcal{M}^\otimes_\mathbb{K}$ has the structure of a submanifold of the vector space $V_1 \otimes_\mathbb{K} \cdots \otimes_\mathbb{K} V_r$. This is not always the case, as the next example will prove.

Example 2.2.11 *The graph of the function*

$$f : \mathbb{R} \to \mathbb{R}^2, \quad f(x) = \begin{cases} ((x-1)^2, 0), & x \leq 1 \\ (0, (x-1)^2), & x > 1 \end{cases}$$

is a submanifold of $\mathbb{R}^3$. *Let* $\mathcal{M}_1 = \mathbb{R} \setminus \{0\}$ *and* $\mathcal{M}_2 = \Gamma_f$ *and as before* $\mathcal{M}^\times := \mathcal{M}_1 \times \mathcal{M}_2$. *It is obvious that if* $X_2 \in \mathcal{M}_2$, *then* $\alpha X_2 \in \mathcal{M}_2$ *if and only if* $\alpha = 1$ *and hence,*

$$\mathcal{M}^\times/_\sim = \mathcal{M}^\times/\{1\}.$$

From Theorem 2.2.5 it follows that $\mathcal{M}^\times/_\sim$ *is a manifold diffeomorphic to* $\mathcal{M}^\times$. *Moreover,*

$$\mathcal{M}_1 \otimes \mathcal{M}_2 = \{\alpha(x, f(x)) \mid \alpha \in \mathbb{R} \setminus \{0\},\ x \in \mathbb{R}\} = (xOy \cup xOz) \setminus (K_1 \cup K_2),$$

where K_1 *and* K_2 *are cones in the plane* xOy *and* xOz *respectively. Thus,* $\mathcal{M}_1 \otimes \mathcal{M}_2$ *is not a submanifold of* $\mathbb{R}^3$.

Next, we give sufficient conditions for $\mathcal{M}^{\otimes_\mathbb{K}}$ to be a submanifold of $V_1 \otimes_\mathbb{K} \cdots \otimes_\mathbb{K} V_r$.

Theorem 2.2.12 *Let* $\mathcal{M}_1 \subset V_1 \setminus \{0\}, \dots, \mathcal{M}_r \subset V_r \setminus \{0\}$ *be compact submanifolds of* $V_1, \dots, V_r$, $\sim$ *be the equivalence relation* (2.17) *on* $\mathcal{M}^\times := \mathcal{M}_1 \times \dots \times \mathcal{M}_r$ *and* G *the Lie-group defined in* (2.24) *with Lie algebra* $\mathfrak{g}$*. If the map*

$$h : \mathcal{M}^\times \to V_1 \otimes \dots \otimes V_r, \quad (X_1, \dots, X_r) \mapsto X_1 \otimes \dots \otimes X_r \tag{2.40}$$

is open with respect to the subspace topology and if there exists a Lie-subgroup $\mathcal{A}$ *of* G *such that:*

(1)
$$\mathcal{M}^\times/_\sim = \mathcal{M}^\times/\mathcal{A},$$

(2)
$$\mathrm{T}_X \mathcal{O}_{\mathcal{A}}(X) = \mathrm{T}_X \mathcal{M}^\times \cap \mathrm{T}_X \mathcal{O}_G(X), \quad \textit{for all} \quad X \in \mathcal{M}^\times, \tag{2.41}$$

then $\mathcal{M}^{\otimes_{\mathbb{K}}}$ *is a submanifold of* $V_1 \otimes_{\mathbb{K}} \dots \otimes_{\mathbb{K}} V_r$ *and is diffeomorphic to* $\mathcal{M}^\times/_\sim$*.*

Proof. Let $[X]_{\mathcal{A}} \in \mathcal{M}^\times/\mathcal{A}$ with $X := (X_1, \dots, X_r) \in \pi_{\mathcal{A}}^{-1}([X]_{\mathcal{A}})$ as a representative. We prove that the map

$$f : \mathcal{M}^\times/\mathcal{A} \to V_1 \otimes \dots \otimes V_r, \;\; [X]_{\mathcal{A}} \mapsto X_1 \otimes \dots \otimes X_r$$

is a homeomorphism onto its image $\operatorname{Im} f = \mathcal{M}^{\otimes_{\mathbb{K}}}$.

The following diagram

$$\begin{tikzcd} \mathcal{M}^\times/\mathcal{A} \arrow[r, "f"] & V_1 \otimes \dots \otimes V_r \\ \mathcal{M}^\times \arrow[u, "\pi_{\mathcal{A}}"] \arrow[ur, dashed, "h"'] & \end{tikzcd} \tag{2.42}$$

commutes. For every open set $W \subset V_1 \otimes \dots \otimes V_r$, the preimage $h^{-1}(W)$ is open. Moreover,

$$h^{-1}(W) = (f \circ \pi_{\mathcal{A}})^{-1}(W) = (\pi_{\mathcal{A}}^{-1}(f^{-1}(W))) \tag{2.43}$$

it follows that $f^{-1}(W)$ is open and hence, f is continuous. Since h is open, it follows that f is open as well, and hence a homeomorphism onto its image.

Further, we will prove that if condition (2.41) is satisfied for all $X \in \mathcal{M}^\times$, then f is an embedding. We have shown that f is a homeomorphism onto its image, thus it is left to prove that f is an immersion, i.e. $T_{[X]_{\mathcal{A}}} f : \mathrm{T}_{[X]_{\mathcal{A}}} \mathcal{M}^\times/\mathcal{A} \to V_1 \otimes \dots \otimes V_r$ has full rank equal to the dimension of $\mathcal{M}^\times/\mathcal{A}$. Let $X = (X_1, \dots, X_r) \in \mathcal{M}^\times$ and $\xi_X^h = (\xi_1^h, \dots, \xi_r^h) \in \mathcal{H}_X$ be a representation for $\xi_{\mathcal{A}} \in \mathrm{T}_{[X]_{\mathcal{A}}} \mathcal{M}^\times/\mathcal{A}$. Then

$$T_{[X]_{\mathcal{A}}} f(\xi_{\mathcal{A}}) = \sum_{j=1}^{r} X_1 \otimes \cdots \xi_j^h \otimes \dots \otimes X_r.$$

From Lemma 2.2.1 it follows that

$$\begin{aligned}\ker T_{[X]_{\mathcal{A}}} f &= \{\xi^h := (\xi_1^h, \ldots, \xi_r^h) \in \mathcal{H}_X \mid \sum_{j=1}^{r} X_1 \otimes \cdots \xi_j^h \otimes \cdots \otimes X_r = 0\} \\ &= \{\xi^h := (\xi_1^h, \ldots, \xi_r^h) \in \mathcal{H}_X \mid \xi^h = \omega \cdot X,\ \omega \in \mathfrak{g}\},\end{aligned}$$

where $\mathfrak{g}$ is the Lie-algebra of G defined in (2.24). Further, we have

$$\ker T_{[X]_{\mathcal{A}}} f = \mathcal{H}_X \cap \{\omega \cdot X \mid \omega \in \mathfrak{a}^{\perp}\},$$

where $\mathfrak{a}^{\perp}$ is the orthogonal complement of $\mathfrak{a}$ in $\mathfrak{g}$, i.e. $\mathfrak{g} = \mathfrak{a} \oplus \mathfrak{a}^{\perp}$. From (2.41) it follows that $\omega \cdot X \notin \mathrm{T}_X\mathcal{M}^{\times}$ for all $0 \neq \omega \in \mathfrak{a}^{\perp}$ and hence, $\ker T_{[X]_{\mathcal{A}}} f = \{0\}$. Thus, f is an immersion and therefore an embedding. From Theorem 2.1.2 it follows that $\operatorname{Im} f = \mathcal{M}^{\otimes}$ is a submanifold of $V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ diffeomorphic to $\mathcal{M}^{\times}/_{\sim}$. □

For compact submanifolds, the map

$$f : \mathcal{M}^{\times}/\mathcal{A} \to V_1 \otimes \cdots \otimes V_r,\ \ [X]_{\mathcal{A}} \mapsto X_1 \otimes \cdots \otimes X_r$$

describes a homeomorphism onto its image. The next result is a consequence of Theorem 2.2.12 and an important tool in the next chapters, to prove that the r-fold tensor products of Grassmann manifolds and of Lagrange Grassmann manifolds have a compact submanifold structure.

Corollary 2.2.13 *Let $\mathcal{M}_1 \subset V_1 \setminus \{0\}, \ldots, \mathcal{M}_r \subset V_r \setminus \{0\}$ be compact submanifolds of $V_1, \ldots, V_r$ and $\sim$ the equivalence relation (2.17) on $\mathcal{M}^{\times} := \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ and G the Lie-group defined in (2.24) with Lie algebra $\mathfrak{g}$. If there exists a Lie-subgroup $\mathcal{A}$ of G such that $\mathcal{M}^{\times}/_{\sim} = \mathcal{M}^{\times}/\mathcal{A}$ and the condition (2.41) holds, then $M^{\otimes_{\mathbb{K}}}$ is a compact submanifold of $V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$.*
Furthermore, if $\mathcal{A} = \{(1, \ldots, 1)\}$ is the trivial subgroup of G and if

$$\omega \cdot X \notin \mathrm{T}_X\mathcal{M}^{\times} \tag{2.44}$$

for all $X \in \mathcal{M}^{\times}$ and $0 \neq \omega \in \mathfrak{g}$, then $\mathcal{M}^{\otimes_{\mathbb{K}}}$ is a compact submanifold of $V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$ diffeomorphic to $\mathcal{M}^{\times}$.

Proof. The continuous map

$$h : \mathcal{M}^{\times} \to V_1 \otimes \cdots \otimes V_r,\ \ (X_1, \ldots, X_r) \mapsto X_1 \otimes \cdots \otimes X_r$$

is closed and proper since $\mathcal{M}^{\times}$ is compact. From the commutative diagram (2.42) we obtain that the continuous map

$$f : \mathcal{M}^{\times}/\mathcal{A} \to V_1 \otimes \cdots \otimes V_r,\ \ [X]_{\mathcal{A}} \mapsto X_1 \otimes \cdots \otimes X_r$$

is closed and proper. Furthermore, f is injective, and hence, f is a homeomorphism onto its image. From (2.41) and the proof of Theorem 2.2.12, we conclude that f is

an immersion and hence, an embedding. Then, $\operatorname{Im} f = \operatorname{Im} h = \mathcal{M}^{\otimes_{\mathbb{K}}}$ is a compact submanifold of $V_1 \otimes_{\mathbb{K}} \cdots \otimes_{\mathbb{K}} V_r$.
The second part of the proof follows immediately from the fact that h is injective and the condition (2.41) becomes (2.44). $\square$

Remark 2.2.14 *If the map h (2.40) is injective, i.e. $\mathcal{A} = \{(1, \ldots, 1)\}$, but the manifolds $\mathcal{M}_j$ are not compact, then $\mathcal{M}^{\otimes_{\mathbb{K}}}$ does not necessarily have a submanifold structure, see Example 2.2.11.*

Chapter 3

Tensor products of Grassmannians and Lagrange-Grassmannians

Important tasks in numerical linear algebra such as invariant subspace computation or subspace tracking, can be formulated as optimization tasks on the set of all m-dimensional subspaces of $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$) called the Grassmann manifold, see e.g. [3, 16, 34, 77]. In this chapter, we recall the fundamental properties of the Grassmann manifold and in particular we discuss its representation as the set of rank-m orthogonal or self-adjoint projectors of $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively, called Grassmannian $\mathrm{Gr}(m, n)$. Furthermore, we generalize the Grassmannian to a tensor product of Grassmannians denoted by $\mathrm{Gr}^{\otimes}(\mathbf{m}, \mathbf{n})$. With tools from the previous chapter, we prove that $\mathrm{Gr}^{\otimes}(\mathbf{m}, \mathbf{n})$ can be equipped with a Riemannian submanifold structure and that it is isometric to the direct product of Grassmannians.

In this chapter, we also present the manifold of all Lagrangian subspaces of $\mathbb{R}^{2n}$ and $\mathbb{C}^{2n}$, called the Lagrangian Grassmannian [4]. The task of determining Lagrangian subspaces of Hamiltonian matrices is important in linear optimal control, in Kalman filtering, etc., being closely connected to solutions of an algebraic Riccati equation. For details in this direction we refer to the literature [34, 63]. Similar to the case of the Grassmann manifold, we give a representation of the Lagrangian Grassmannian with orthogonal, respective self-adjoint projectors of rank n that satisfy some properties. We will call this set Lagrange-Grassmannian $\mathrm{LG}(n)$. The r-fold tensor product of Lagrange-Grassmannians $\mathrm{LG}^{\otimes}(\mathbf{n})$ is a submanifold of the $\mathrm{Gr}^{\otimes}(\mathbf{2n})$. Moreover, it is natural to think about what happens when the symplectic form on $\mathbb{C}^{2n}$ is given by a sesquilinear map and not by a bilinear one. In [5], it was proven that in this case the set of all complex Lagrangian subspaces has the structure of a manifold diffeomorphic to the unitary group $\mathrm{U}(n)$, called the complex Lagrangian Grassmannian. Similar to the classical cases, we characterize the complex Lagrangian Grassmannian by a subset of the set of rank-n self–adjoint projectors of $\mathbb{C}^{2n}$ and show that this set is a manifold diffeomorphic to the direct product of complex Lagrange-Grassmannians.

3.1 Preliminaries on the Grassmannian

Let $\mathbb{K}$ denote the field of real or complex numbers, i.e. $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Recall that the $\mathbb{K}$- *Grassmann manifold* is the set of all $\mathbb{K}$-subspaces of dimension m from $\mathbb{K}^n$ and is denoted with $\mathrm{Grass}_{\mathbb{K}}(m,n)$. Its geometric properties are well understood and we refer the interested reader to [8, 34, 55] for specific details. There are several equivalent identifications for $\mathrm{Grass}_{\mathbb{K}}(m,n)$, as a quotient of $\mathbb{K}^{n\times m}$ [3], or as an homogeneous space [34], etc. In our inquiry, we concentrate on the identification of the Grassmann manifold with the set of rank-n self–adjoint projectors, see [32]. For this purpose we recapitulate some important objects that will be used throughout the thesis.

The set of all $n \times n$ symmetric matrices $\mathfrak{sym}_n$ is the subspace of matrices $A \in \mathbb{R}^{n\times n}$ with the property that $A^\top = A$. Similar, the set of $n \times n$ Hermitian matrices $\mathfrak{her}_n$ is the real vector subspace of matrices $A \in \mathbb{C}^{n\times n}$ satisfying $A^\dagger = A$. With $A^\top$ we refer to the *transpose* of A and with $A^\dagger$ to the *transpose conjugate* of A. Fundamental for our work with Grassmann manifolds are the Lie groups of special orthogonal matrices $\mathrm{SO}(n)$ and special unitary matrices $\mathrm{SU}(n)$ and the Lie algebras of $\mathfrak{so}_n$ and $\mathfrak{su}_n$, respectively, i.e.

$$\mathrm{SO}(n) = \{\Theta \in \mathbb{R}^{n\times n} \mid \Theta^\top\Theta = I_n,\ \det\Theta = 1\}, \quad \mathfrak{so}_n = \{\Omega \in \mathbb{R}^{n\times n} \mid \Omega^\top = -\Omega\}.$$

and

$$\mathrm{SU}(n) = \{\Theta \in \mathbb{C}^{n\times n} \mid \Theta^\dagger\Theta = I_n,\ \det\Theta = 1\}, \quad \mathfrak{su}_n = \{\Omega \in \mathbb{C}^{n\times n} \mid \Omega^\dagger = -\Omega,\ \mathrm{tr}(\Omega) = 0\}.$$

The *Grassmannian*,

$$\mathrm{Gr}_{\mathbb{K}}(m,n) := \{P \in \mathbb{K}^{n\times n} \mid P = P^\dagger = P^2,\ \mathrm{tr}(P) = m\}, \tag{3.1}$$

is the set of all rank m self-adjoint projection operators of $\mathbb{K}^n$. In [32], the authors have shown that $\mathrm{Gr}_{\mathbb{R}}(m,n)$ and $\mathrm{Gr}_{\mathbb{C}}(m,n)$ are smooth and compact real submanifolds of $\mathfrak{sym}_n$ and $\mathfrak{her}_n$ of real dimension $m(n-m)$ and $2m(n-m)$, respectively. Here we discuss only the complex Grassmannian $\mathrm{Gr}_{\mathbb{C}}(m,n)$, since its geometric structure can be transfered to the real Grassmannian $\mathrm{Gr}_{\mathbb{R}}(m,n)$ just by replacing $\mathrm{SU}(n)$ and $\mathfrak{su}_n$ with $\mathrm{SO}(n)$ and $\mathfrak{so}_n$, respectively, and transpose with transpose conjugate. From here on, the complex Grassmannian is denoted by $\mathrm{Gr}(m,n)$. For further use, we denote with

$$\Pi_{m,n} := \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.2}$$

the *standard projector* of rank m acting on $\mathbb{C}^n$. Whenever the values of m and n are clear from the context, we will simply write Π.

The Grassmannian $\mathrm{Gr}(m,n)$ is diffeomorphic to the Grassmann manifold, see [34]. It is also diffeomorphic to the homogeneous space $\mathrm{SU}(n)/\mathrm{Stab}_{\mathrm{SU}_n}(\Pi_{m,n})$ according to the map

$$\mathrm{SU}(n) \times \mathrm{Gr}(m,n) \to \mathrm{Gr}(m,n), \quad (\Theta, P) \mapsto \Theta^\dagger P \Theta, \tag{3.3}$$

where

$$\mathrm{Stab}_{\mathrm{SU}(n)}(\Pi_{m,n}) = \{\Theta \in \mathrm{SU}(n) \mid \Theta^\dagger \Pi_{m,n} \Theta = \Pi_{m,n}\}.$$

From the diffeomorphism (3.3) we derive the following representation of the elements P in the Grassmannian and of the vectors in the tangent space $\mathrm{T}_P\mathrm{Gr}(m,n)$. The tangent space at P to $\mathrm{Gr}(m,n)$ is given by

$$\mathrm{T}_P\mathrm{Gr}(m,n) = \{[P,\Omega] \mid \Omega \in \mathfrak{su}_n\}, \tag{3.4}$$

where the *matrix commutator* is defined by

$$[P,\Omega] = P\Omega - \Omega P.$$

Every element $P \in \mathrm{Gr}(m,n)$ and every tangent vector $\xi \in \mathrm{T}_P\mathrm{Gr}(m,n)$ can be written as

$$P = \Theta^\dagger \begin{bmatrix} I_m & 0 \\ & \\ 0 & 0 \end{bmatrix} \Theta \text{ and } \xi = \Theta^\dagger \begin{bmatrix} 0 & Z \\ Z^\dagger & 0 \end{bmatrix} \Theta, \tag{3.5}$$

with $\Theta \in \mathrm{SU}(n)$ and $Z \in \mathbb{C}^{m\times(n-m)}$.

With respect to the Riemannian metric induced by the Hilbert-Schmidt inner product of $\mathfrak{her}_n$

$$\langle X, Y\rangle := \mathrm{tr}(XY), \tag{3.6}$$

the Grassmannian $\mathrm{Gr}(m,n)$ is a Riemannian submanifold of $\mathfrak{her}_n$ and the unique orthogonal projector onto $\mathrm{T}_P\mathrm{Gr}(m,n)$ is given by

$$\mathrm{ad}_P^2 X = [P,[P,X]], \quad X \in \mathfrak{her}_n, \tag{3.7}$$

see [32]. From (2.3) it follows that the Levi-Civita connection $\boldsymbol{\nabla}$ on $\mathrm{Gr}(m,n)$ is given by

$$\boldsymbol{\nabla}_\xi \mathcal{X}(P) = \mathrm{ad}_P^2\Big(D\widetilde{\mathcal{X}}(P)(\xi)\Big), \tag{3.8}$$

for any $\xi \in \mathrm{T}_P\mathrm{Gr}(m,n)$, $P \in \mathrm{Gr}(m,n)$, where $\widetilde{\mathcal{X}}$ is a smooth extension of the vector field $\mathcal{X}$ on $\mathrm{Gr}(m,n)$ to a vector field on $\mathfrak{her}_n$. It is known in the literature (see [32]) that, the curve

$$t \mapsto \gamma(t) = e^{t[\xi,P]} P e^{-t[\xi,P]} \tag{3.9}$$

describes the unique geodesic on $\mathrm{Gr}(m,n)$ with initial conditions $\gamma(0) = P \in \mathrm{Gr}(m,n)$ and $\frac{d}{dt}\gamma(0) = \xi \in \mathrm{T}_P\mathrm{Gr}(m,n)$. One can check that $\gamma(t) = e^{t[\xi,P]}Pe^{-t[\xi,P]}$ satisfies the equation (2.6). Similarly, it can be verified that the parallel transport of $\xi \in \mathrm{T}_P\mathrm{Gr}_{m,n}$ to $\mathrm{T}_{\gamma(t)}\mathrm{Gr}_{m,n}$ along the geodesic γ is given by

$$\xi \mapsto e^{t[\xi,P]}\xi e^{-t[\xi,P]}. \tag{3.10}$$

In general, parametrizations around a point P from a n-dimensional smooth manifold, are defined as inverses of the charts, i.e. smooth maps μ_P from $\mathbb{R}^n$ to the manifold with

$\mu_P(0) = P$. Here, we say that a *local parametrization* around a point $P \in \mathrm{Gr}(m,n)$ is a smooth map

$$\mu_P : \mathrm{T}_P\mathrm{Gr}(m,n) \to \mathrm{Gr}(m,n),$$

which satisfies

$$\mu_P(0) = P, \quad D\mu_P(0) = \mathrm{id}_{\mathrm{T}_P\mathrm{Gr}(m,n)}. \tag{3.11}$$

For the Grassmannian, the authors of [32] have introduced three types of local parametrizations, one given by the exponential map and the other two given by approximations of the exponential map:

(a) *Riemannian normal coordinates*

$$\mu_P^{\exp} : \mathrm{T}_P\mathrm{Gr}(m,n) \to \mathrm{Gr}(m,n), \quad \xi \mapsto e^{[\xi,P]}Pe^{-[\xi,P]}. \tag{3.12}$$

(b) *QR-type coordinates*

$$\mu_P^{\mathrm{QR}} : \mathrm{T}_P\mathrm{Gr}(m,n) \to \mathrm{Gr}(m,n), \quad \xi \mapsto (I_n + [\xi,P])_Q \; P \; (I_n + [\xi,P])_Q^{\dagger}. \tag{3.13}$$

Here $I_n + [\xi, P]$ is the first order approximation of the matrix exponential $\mathrm{e}^{[\xi,P]}$ and $(I_n + [\xi, P])_Q$ denotes the $Q-$factor from the QR decomposition of $I_n + [\xi, P]$. Since, $I_n + [\xi, P]$ is similar to a matrix of the form

$$\begin{bmatrix} I_m & -Z \\ Z^{\dagger} & I_{n-m} \end{bmatrix},$$

with $Z \in \mathbb{C}^{m\times(n-m)}$, it follows that $\det(I_n+[\xi,P]) = \det(I_m+ZZ^{\dagger}) > 0$, i.e. $I_n+[\xi,P]$ is invertible. Thus, the QR decomposition of $I_n + [\xi, P]$ is unique.

(c) *Cayley coordinates*

$$\mu_P^{\mathrm{Cay}} : \mathrm{T}_P\mathrm{Gr}(m,n) \to \mathrm{Gr}(m,n), \quad \xi \mapsto c([\xi,P])Pc(-[\xi,P]), \tag{3.14}$$

where c denotes the *Cayley transform* of a skew-Hermitian matrix Ω, i.e.

$$c : \mathfrak{su}_n \to \mathrm{SU}(n), \quad \Omega \mapsto (2I_n + \Omega)(2I_n - \Omega)^{-1}.$$

In [32] it is shown that the maps $\mu_P^{\exp}$, μ_P^{QR} and μ_P^{Cay} are local parametrizations according to (3.11) and moreover, that they satisfy

$$\frac{d^2}{dt^2}\mu_p^{\exp}(t\xi)|_{t=0} = \frac{d^2}{dt^2}\mu_p^{\mathrm{QR}}(t\xi)|_{t=0} = \frac{d^2}{dt^2}\mu_p^{\mathrm{Cay}}(t\xi)|_{t=0} = \Theta^{\dagger}\begin{bmatrix} -2ZZ^{\dagger} & 0 \\ 0 & 2Z^{\dagger}Z \end{bmatrix}\Theta,$$

for all $P \in \mathrm{Gr}(m,n)$ and $\xi \in \mathrm{T}_P\mathrm{Gr}(m,n)$, where ξ and Z are related by (3.5).

3.2 Riemannian structure of the tensor product of Grassmannians

In this section we introduce an object which arises naturally in many applications as we will see later on, i.e. the tensor product of Grassmannians. We describe its Riemannian structure and show that it is isometric to the direct product of Grassmannians.

Let $\mathbb{K}=\mathbb{R}$ or $\mathbb{K}=\mathbb{C}$. The *$r-$fold tensor product of Grassmannians* is the subset

$$\mathrm{Gr}^{\otimes}_{\mathbb{K}}(\mathbf{m},\mathbf{n}) := \{P_1\otimes\cdots\otimes P_r \mid P_j\in \mathrm{Gr}_{\mathbb{K}}(m_j,n_j),\ j=1,\ldots,r\}\subset \mathrm{Gr}_{\mathbb{K}}(M,N) \quad (3.15)$$

of all rank-M self-adjoint projectors $\mathbf{P}:\mathbb{K}^N\to\mathbb{K}^N$ which have the form of a Kronecker product $\mathbf{P}:=P_1\otimes\cdots\otimes P_r$, where $M:=m_1m_2\cdots m_r$ and $N:=n_1n_2\cdots n_r$. Here, $(\mathbf{m},\mathbf{n})$ denotes the multi index

$$(\mathbf{m},\mathbf{n}) := \Big((m_1,n_1),(m_2,n_2),\ldots,(m_r,n_r)\Big). \quad (3.16)$$

As in the previous section, we present only the case of complex Grassmannians, the real case being easily obtained from the complex one, if not otherwise specified. To avoid complicated notation, we will use $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ for the r-fold tensor product of complex Grassmannians.

The $r-$fold tensor product of Grassmannians can be equipped with a Riemannian submanifold structure of $\mathfrak{her}_N\cong\mathfrak{her}_{n_1}\otimes\cdots\otimes\mathfrak{her}_{n_r}$, as we will prove next. The Riemannian metric of $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ is induced by the Hilbert-Schmidt inner product (3.6) on $\mathfrak{her}_N$.

Remark 3.2.1 *For the real case, we do not have an isomorphism between $\mathfrak{sym}_N$ and the tensor product vector space $\mathfrak{sym}^{\otimes}(\mathbf{n}):=\mathfrak{sym}_{n_1}\otimes\cdots\otimes\mathfrak{sym}_{n_r}$. On the $\mathfrak{sym}^{\otimes}(\mathbf{n})$ we define the inner product*

$$\langle X_1\otimes\cdots\otimes X_r, Y_1\otimes\cdots\otimes Y_r\rangle := \mathrm{tr}(X_1Y_1)\cdots\mathrm{tr}(X_rY_r), \quad (3.17)$$

for all $X_j,Y_j\in\mathfrak{sym}_{n_j}$, for $j=1,\ldots,r$.

Furthermore, the special features of the Grassmannian lead to a diffeomorphism between $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ and the *direct $r-$fold product of Grassmannians*

$$\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n}) := \{(P_1,\ldots,P_r)\mid P_j\in\mathrm{Gr}(m_j,n_j),\ j=1,\ldots,r\}. \quad (3.18)$$

Nothe that, $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ is a smooth and compact Riemannian submanifold of the product space $\mathfrak{her}_{n_1}\times\cdots\times\mathfrak{her}_{n_r}$ with Riemannian metric induced by the inner product

$$\Big\langle (X_1,\ldots,X_r),(Y_1,\ldots,Y_r)\Big\rangle := \mathrm{tr}(X_1Y_1)+\cdots+\mathrm{tr}(X_rY_r), \quad (3.19)$$

for all $X_j,\ Y_j\in\mathfrak{her}_{n_j}$ and $j=1,\ldots,r$. The fundamental geometric objects and concepts of $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ are trivial extensions of the corresponding ones on the Grassmannian. We give the following property of elements in $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ that we will use in the proof of Proposition 3.2.3 and further on in Chapter 4.

Lemma 3.2.2 *Let* $P = (P_1, \ldots, P_r) \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ *and* $(\xi_1, \ldots, \xi_r) \in \mathrm{T}_P\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$. *Then,*

$$\langle P_1 \otimes \cdots \xi_j \otimes \cdots \otimes P_r, P_1 \otimes \cdots \xi_k \otimes \cdots \otimes P_r\rangle = 0, \tag{3.20}$$

for all $j \neq k$, $j,k = 1,\ldots,r$. *The inner product considered is given by* (3.6).

Proof. From (3.6) and the properties of the trace function, it follows that

$$\langle X_1 \otimes \cdots \otimes X_r, Y_1 \otimes \cdots \otimes Y_r\rangle \quad = \mathrm{tr}(X_1Y_1)\cdots\mathrm{tr}(X_rY_r),$$

for all $X_j,\ Y_j \in \mathfrak{her}_{n_j}$ and $j = 1,\ldots,r$. Without loss of generality, in (3.20) we assume that $j = 1$ and $k = 2$. Since $\mathrm{tr}(\xi_jP_j) = 0$ for all $\xi_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$, then

$$\langle \xi_1 \otimes \cdots \otimes P_r, P_1 \otimes \xi_2 \otimes \cdots \otimes P_r\rangle \quad = \mathrm{tr}(\xi_1P_1)\,\mathrm{tr}(P_2\xi_2)\cdots\mathrm{tr}(P_r) = 0.$$

□

Proposition 3.2.3 *The* $r-$*fold tensor product of Grassmannians* $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ *is a smooth and compact submanifold of* $\mathfrak{her}_N$, *diffeomorphic to* $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$, *i.e. the map*

$$\varphi : \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n}) \to \mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})\ , \quad (P_1,\ldots,P_r) \mapsto P_1 \otimes \cdots \otimes P_r \tag{3.21}$$

defines a diffeomorphism. Furthermore, φ *is a global Riemannian isometry if the inner product (3.19) on* $\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$ *is scaled as follows*

$$\langle (X_1,\ldots,X_r),(Y_1,\ldots,Y_r)\rangle = M_1\,\mathrm{tr}(X_1Y_1) + \cdots + M_r\,\mathrm{tr}(X_rY_r), \tag{3.22}$$

with $M_j := \displaystyle\prod_{k=1,\ k\neq j}^{r} m_k$, *for* $X_j,\ Y_j \in \mathfrak{her}_{n_j}$, $j = 1,\ldots,r$.

Proof. First we prove that the map

$$\varphi : \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n}) \to \mathfrak{her}_N\ , \quad (P_1,\ldots,P_r) \mapsto P_1 \otimes \cdots \otimes P_r$$

is injective. Let $(P_1,..,P_r)$, $(Q_1,\ldots,Q_r) \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ such that

$$P_1 \otimes \cdots \otimes P_r = Q_1 \otimes \cdots \otimes Q_r.$$

From Lemma 2.2.1 it follows that there exist $\alpha_1,\ldots,\alpha_r \in \mathbb{C}$ with $\alpha_1\alpha_2\cdots\alpha_r = 1$ such that

$$P_1 = \alpha_1Q_1, \quad ,\ldots, \quad P_r = \alpha_rQ_r.$$

Since P_j and Q_j have only 0 and 1 as eigenvalues it follows that $\alpha_j = 1$ for all $j = 1,\ldots,r$. Thus φ is injective. Let G denote the Lie group $G := \{(\alpha_1,\ldots,\alpha_r) \in \mathbb{C}^r \mid \alpha_1\alpha_2\cdots\alpha_r = 1\}$ and $\mathfrak{g} := \{(\omega_1,\ldots,\omega_r) \in \mathbb{C}^r \mid \omega_1 + \cdots \omega_r = 0\}$ its Lie algebra. For any $P \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ and any $\omega \in \mathfrak{g}$ we have that $\omega \cdot P = (\omega_1P_1,\ldots,\omega_rP_r) \notin$

$\mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$. Since $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ is compact, from Corollary 2.2.13 we conclude that $\mathrm{Gr}^\otimes(\mathbf{m},\mathbf{n})$ is a compact submanifold of $\mathfrak{her}_N$ diffeomorphic to $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$.

For the second part of the proposition, let (3.22) be the inner product on $\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$. The tangent tangent map of φ at P is given by

$$T_P\varphi : \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n}) \to \mathfrak{her}_N$$

$$(\xi_1,\dots,\xi_r) \mapsto \sum_{j=1}^{r} P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r.$$

Then, from the properties of the trace function and the fact that $\mathrm{tr}(P\xi) = 0$ for all $\xi \in \mathrm{T}_P\mathrm{Gr}(m,n)$, we obtain

$$\begin{aligned}\langle T_P\varphi(\xi), T_P\varphi(\eta)\rangle &= \sum_{j=1}^{r}\sum_{k=1}^{r}\langle P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r, P_1 \otimes \cdots \otimes \eta_k \otimes \cdots \otimes P_r\rangle \\ &= \sum_{j=1}^{r}\sum_{k=1}^{r}\mathrm{tr}\Big((P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r)(P_1 \otimes \cdots \otimes \eta_k \otimes \cdots \otimes P_r)\Big) \\ &= \sum_{j=1}^{r} M_j\,\mathrm{tr}(\xi_j\eta_j) = \langle \xi, \eta\rangle,\end{aligned}$$

where $M_j := \prod_{k=1,\, k\neq j}^{r} m_k$, for all $\xi,\ \eta \in \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$. Hence, φ is a global Riemannian isometry. □

The injectivity of the map φ defined by (3.21) is very special and does not hold in general, as can be noticed from the subsequent example

$$\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r} \to \mathfrak{her}_N, \quad (X_1,\dots,X_r) \mapsto X_1 \otimes \cdots \otimes X_r. \tag{3.23}$$

Remark 3.2.4 *It is a well-known fact that the Grassmannian* $\mathrm{Gr}(m,n)$ *is diffeomorphic to the Grassmann manifold* $\mathrm{Grass}(m,n)$*, cf.[34]. Therefore,* $\mathrm{Gr}(m_1,n_1) \otimes \mathrm{Gr}(m_2,n_2)$ *is diffeomorphic to*

$$\{V_1 \otimes V_2 \mid V_1 \in \mathrm{Grass}(m_1,n_1),\ V_2 \in \mathrm{Grass}(m_2,n_2)\} \subset \mathrm{Grass}(M,N),$$

where $M := m_1m_2$ *and* $N := n_1n_2$*. This allows us to use the term* r*-fold tensor product of Grassmann manifolds when talking about the manifold* $\mathrm{Gr}^\otimes(\mathbf{m},\mathbf{n})$*.*

3.3 Symplectic vector spaces and Lagrange-Grassmannians

In this section we review the notions of Lagrangian subspaces of $\mathbb{R}^{2n}$ and $\mathbb{C}^{2n}$ and give a representation of these subspaces by certain orthogonal and self-adjoint projectors, respectively. In the literature [4], the set of all Lagrangian subspaces of $\mathbb{R}^{2n}$ and $\mathbb{C}^{2n}$

is a manifold called the Lagrangian Grassmannian. We show that the set of all projectors representing Lagrangian subspaces is a submanifold of the Grassmann manifold. Naturally, one could be interested in what happens to the complex version of the real Lagrangian Grassmannian, i.e. the set of all complex Lagrangian subspaces of $\mathbb{C}^{2n}$ with respect to a sesquilinear form. In [5], it was proved that the complex Lagrangian Grassmannian is a manifold, which is diffeomorphic to the group of unitary matrices. Similar to the classical cases, we characterize the complex Lagrangian Grassmannian by a subset of rank-n self-adjoint projectors of $\mathbb{C}^{2n}$ show that this subset is a manifold. We start with basic notions from symplectic geometry.

Definition 3.3.1 *A* symplectic vector space *is a pair* (V, ω)*, where* V *is a finite dimensional vector space over* $\mathbb{K}$ $(\mathbb{K} = \mathbb{C}$ *or* $\mathbb{K} = \mathbb{R})$ *and* $\omega : V \times V \to \mathbb{K}$ *is a bilinear form, which satisfies the following:*

(i) nondegeneracy
$$\omega(u, v) = 0 \ \textit{for all}\, v \in V \Rightarrow \ u = 0,$$

(ii) skew-symmetry
$$\omega(u, v) = -\omega(v, u),$$
for all $u,\ v \in V$*.*

A complex symplectic vector space *is a complex vector space* V *endowed with a sesquilinear form* $\omega : V \times V \to \mathbb{C}$*, that is nondegenerate and*

(ii') skew-Hermitian
$$\omega(u, v) = -\overline{\omega(v, u)},$$
for all $u,\ v \in V$*.*

If (V, ω) is a symplectic vector space (or complex symplectic vector space) and $W \subset V$ a subspace (resp. complex subspace), then the *symplectic complement* (resp. *complex symplectic complement*) of W is

$$W^{\omega} = \{v \in V \mid \omega(v, w) = 0, \text{ for all } \ w \in W\}.$$

Moreover, from the nondegeneracy of the bilinear form (resp. sesquilinear form) ω, it follows that
$$\dim W + \dim W^{\omega} = \dim V \quad \text{and} \quad (W^{\omega})^{\omega} = W. \tag{3.24}$$

In contrast to the case of the orthogonal complement of a linear subspace, in general $W \cap W^{\omega} \neq \{0\}$. The following instances may occur:

(a) $W \cap W^{\omega} = \{0\}$ and say that W is a *symplectic subspace* (resp. *complex symplectic subspace*);

(b) $W \subset W^{\omega}$ and say that W is an *isotropic subspace* (resp. *complex isotropic subspace*);

(c) $W^\omega \subset W$ and say that W is a *coisotropic subspace* (resp. *complex coisotropic subspace*);

(d) $W = W^\omega$ and we say that W is a *Lagrangian subspace* (resp. *complex Lagrangian subspace*).

As a consequence of (3.24) it follows that a Lagrangian subspace (resp. complex Lagrangian subspace) $W \subset V$ is an isotropic subspace (resp. complex isotropic subspace) of dimension $(\dim V)/2$.

Classical examples of symplectic spaces are given by the spaces $\mathbb{K}^{2n}$, equipped with **standard symplectic form**

$$\omega(u, v) = u^\top J v, \quad u,\ v \in \mathbb{K}^{2n}, \tag{3.25}$$

where J is the skew-symmetric, nonsingular matrix

$$J := \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}, \tag{3.26}$$

and I_n is the $n \times n$ identity matrix.
The space $\mathbb{C}^{2n}$ can be endowed also with a sesquilinear form that we call the **standard complex symplectic form**

$$\omega(u, v) = u^\dagger J v, \quad u,\ v \in \mathbb{C}^{2n}. \tag{3.27}$$

In what follows, we recall standard results for the classical cases of Lagrangian subspaces of the standard symplectic spaces $\mathbb{R}^{2n}$ and $\mathbb{C}^{2n}$ as well as for the case of complex Lagrangian subspaces of $\mathbb{C}^{2n}$. To this extent, we present some basics on the structure of symplectic groups, compact symplectic groups and their Lie-algebras. For a detailed inquiry we refer to the literature [31, 45].

The *symplectic group* $\mathrm{Sp}(2n, \mathbb{K})$ is defined as

$$\mathrm{Sp}(2n, \mathbb{K}) = \{\Theta \in \mathrm{GL}_{2n}(\mathbb{K}) \mid \Theta J \Theta^\top = J\},$$

and has as Lie-algebra the space of *Hamiltonian matrices*,

$$\mathfrak{sp}(2n, \mathbb{K}) = \{X \in \mathfrak{gl}_{2n}(\mathbb{K}) \mid JX = -X^\top J\}.$$

An easy computation will reveal the following block structure for matrices Θ and X in the symplectic group and in its Lie-algebra, respectively:

$$\Theta = \begin{bmatrix} Q & R \\ S & T \end{bmatrix}, \quad X = \begin{bmatrix} A & B \\ C & -A^\top \end{bmatrix}, \tag{3.28}$$

where $Q, R, S, T, A, B, C \in \mathbb{K}^{n\times n}$, with $Q^\top T - S^\top R = I_n$, $Q^\top S = S^\top Q$, $T^\top R = R^\top T$, $B^\top = B$ and $C^\top = C$. We give also the compact version of the symplectic group, i.e. the *orthogonal symplectic group* for $\mathbb{K} = \mathbb{R}$

$$\mathrm{OSp}(n) = \mathrm{SO}_{2n} \cap \mathrm{Sp}(2n, \mathbb{R}) = \{\Theta \in \mathrm{SO}_{2n} \mid \Theta J \Theta^\top = J\}$$

and the *unitary symplectic group* for $\mathbb{K} = \mathbb{C}$

$$\mathrm{Sp}(n) = \mathrm{SU}_{2n} \cap \mathrm{Sp}(2n, \mathbb{C}) = \{\Theta \in \mathrm{SU}_{2n} \mid \Theta J \Theta^\top = J\}.$$

The Lie algebras of $\mathrm{OSp}(n)$ and $\mathrm{Sp}(n)$ are

$$\mathfrak{osp}(n) = \{X \in \mathfrak{so}_{2n} \mid JX = XJ\}$$

and

$$\mathfrak{sp}(n) = \{X \in \mathfrak{su}_{2n} \mid JX = \bar{X}J\},$$

respectively. Similar to the case of symplectic groups and their Lie-algebras, one obtains the following block-structure for their compact version. Thus, $\Theta \in \mathrm{OSp}(n)$ and $X \in \mathfrak{osp}(n)$ are given by

$$\Theta = \begin{bmatrix} Q & R \\ -R & Q \end{bmatrix}, \quad X = \begin{bmatrix} A & B \\ -B & A \end{bmatrix}, \tag{3.29}$$

with $R, Q, A, B \in \mathbb{R}^{n \times n}$ and $A = -A^\top,\ B = B^\top$. For $\Theta \in \mathrm{Sp}(n)$ and $X \in \mathfrak{sp}(n)$ we have

$$\Theta = \begin{bmatrix} Q & R \\ -\bar{R} & \bar{Q} \end{bmatrix}, \quad X = \begin{bmatrix} A & B \\ -\bar{B} & \bar{A} \end{bmatrix}, \tag{3.30}$$

where $R, Q, A, B \in \mathbb{C}^{n \times n}$ and $A^\top = -\bar{A}, B^\top = B$.

We know from the literature [4], that the set of all Lagrangian subspaces of $\mathbb{R}^{2n}$ and $\mathbb{C}^{2n}$ is a manifold diffeomorphic to $\mathrm{U}(n)/\mathrm{O}(n)$ and $\mathrm{Sp}(n)/\mathrm{U}(n)$, respectively, called the *Lagrangian Grassmannian.* The notation $\mathrm{U}(n)$ stands for the unitary group and $\mathrm{O}(n)$ for the orthogonal group. In what comes, we identify the Lagrangian subspaces with orthogonal and self-adjoint projectors of $\mathbb{R}^{2n}$ and $\mathbb{C}^{2n}$, respectively.

Proposition 3.3.2 *([32]) There is a one-to-one correspondence between the set of Lagrangian subspaces $W \subset \mathbb{K}^{2n}$ and the set of all self-adjoint projectors P of $\mathbb{K}^{2n}$ of rank-n which satisfy $P^\top J P = 0$.*

Proof. We give the proof in the case $\mathbb{K} = \mathbb{C}$ and mention that for $\mathbb{K} = \mathbb{R}$ it is similar. Let $\mathcal{S}$ denote the set $\mathcal{S} := \{P \in \mathbb{C}^{2n \times 2n} \mid P^\dagger = P,\ P^2 = P,\ P^\top J P = 0\}$ and let $P \in \mathcal{S}$. Then $P^\top J P = 0$ and hence, $\operatorname{Im} P$ is a Lagrangian subspace of $\mathbb{C}^{2n}$. Now, let $W \subset \mathbb{C}^{2n}$ be a Lagrangian subspace and $B \in \mathbb{C}^{2n \times n}$ with $B^\dagger B = I_n$ a basis for W, i.e. $B^\top J B = 0$. Since any $\Theta \in \mathrm{Sp}(n)$ is of the form

$$\Theta = \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ -\overline{\Theta}_{12} & \overline{\Theta}_{11} \end{bmatrix},$$

with $\Theta_{11}, \Theta_{12} \in \mathbb{C}^{n \times n}$ such that

$$\Theta_{11}\Theta_{12}^\top - \Theta_{12}\Theta_{11}^\top = [\Theta_{11} \quad \Theta_{12}]\, J \begin{bmatrix} \Theta_{11}^\top \\ \Theta_{12}^\top \end{bmatrix} = 0.$$

Hence, it is always possible to extend B to a basis $\Theta = [B \quad B^{\perp}] \in \mathrm{Sp}(n)$ of the symplectic vector space $\mathbb{C}^{2n}$ and thus, there exist $P = BB^{\dagger}$ such that $W = \operatorname{Im} R$.

□

The *Lagrange–Grassmannian* is the set of all self-adjoint projectors of rank n with the property that $P^{\top}JP = 0$, denoted by $\mathrm{LG}_{\mathbb{K}}(n)$, i.e.

$$\mathrm{LG}_{\mathbb{K}}(n) = \{P \in \mathbb{K}^{2n\times 2n} \mid P = P^{\dagger},\ P = P^2, \operatorname{tr} P = n, P^{\top}JP = 0\}.$$

It is clear that, the Lagrange–Grassmannians is diffeomorph to the Lagrangian Grassmannian. The Lagrange-Grassmannians $\mathrm{LG}_{\mathbb{R}}(n)$ and $\mathrm{LG}_{\mathbb{C}}(n)$ are smooth, compact and connected submanifolds of the Grassmann manifolds $\mathrm{Gr}_{\mathbb{R}}(n, 2n)$ and $\mathrm{Gr}_{\mathbb{C}}(n, 2n)$, with dimension $n(n+1)/2$ and $n(n+1)$, respectively. They are the orbit $\sigma(\Pi)$ of the group actions

$$\sigma : \mathrm{OSp}(n) \times \mathrm{Gr}_{\mathbb{R}}(n, 2n) \to \mathrm{Gr}_{\mathbb{R}}(n, 2n), \quad (\Theta, \Pi) \mapsto \Theta^{\top}\Pi\Theta$$

and

$$\sigma : \mathrm{Sp}(n) \times \mathrm{Gr}_{\mathbb{C}}(n, 2n) \to \mathrm{Gr}_{\mathbb{C}}(n, 2n), \quad (\Theta, \Pi) \mapsto \Theta^{\top}\Pi\overline{\Theta},$$

for $\mathrm{LG}_{\mathbb{R}}(n)$ and $\mathrm{LG}_{\mathbb{C}}(n)$, respectively. We remind that Π is the standard projector of $\mathbb{K}^{2n}$ (3.2). The tangent spaces of $\mathrm{LG}_{\mathbb{R}}(n)$ and $\mathrm{LG}_{\mathbb{C}}(n)$ at P are

$$\mathrm{T}_P\mathrm{LG}_{\mathbb{R}}(n) = \{[P,\ \Omega] \mid \Omega \in \mathfrak{osp}(n)\} \quad \text{and} \quad \mathrm{T}_P\mathrm{LG}_{\mathbb{C}}(n) = \{[P,\ \Omega] \mid \Omega \in \mathfrak{sp}(n)\},$$

respectively. Moreover, every element $P \in \mathrm{LG}_{\mathbb{K}}(n)$ and every tangent vector $\xi \in \mathrm{T}_P\mathrm{LG}_{\mathbb{K}}(n)$ has the structure

$$P = \Theta^{\top}\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}\overline{\Theta}, \quad \xi = \Theta^{\top}\begin{bmatrix} 0 & Z \\ Z & 0 \end{bmatrix}\overline{\Theta}, \tag{3.31}$$

where $\Theta \in \mathrm{OSp}(n),\ Z \in \mathfrak{sym}_n$ and $\Theta \in \mathrm{Sp}(n), Z \in \mathfrak{her}_n$, for $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$, respectively. The projection operator onto $\mathrm{T}_P\mathrm{LG}_{\mathbb{K}}(n)$, $P \in \mathrm{LG}_{\mathbb{K}}(n)$ is

$$\pi_P : \mathrm{T}_P\mathrm{Gr}_{\mathbb{K}}(n, 2n) \to \mathrm{T}_P\mathrm{Gr}_{\mathbb{K}}(n, 2n), \quad \xi \mapsto \frac{\xi + J\xi^{\top}J}{2}. \tag{3.32}$$

Since, $\mathrm{OSp}(n)$ and $\mathrm{Sp}(n)$ are subgroups of $\mathrm{SO}(n)$ and $\mathrm{SU}(n)$, respectively, the geodesics, the exponential map and the parallel transport on the Lagrange-Grassmannians $\mathrm{LG}_{\mathbb{K}}(n)$ are restrictions of the same object on the Grassmann manifold $\mathrm{Gr}_{\mathbb{K}}(n, 2n)$. Furthermore, in [32] it was shown that this is still true for the proposed approximations of the exponential map, i.e. the QR-coordinates and the Cayley-coordinates.

We know that the set of complex Lagrangian subspaces of the complex symplectic space $\mathbb{C}^{2n}$ has a manifold structure diffeomorphic to $\mathrm{U}(n)$. In the sequel, we use the same techniques as in the case of classical Lagrangian subspaces of $\mathbb{K}^{2n}$ to define a geometric structure for the set of complex Lagrangian subspaces by means of certain self-adjoint projectors of $\mathbb{C}^{2n}$. First, we introduce the Lie-group

$$\widehat{\mathrm{Sp}}(2n, \mathbb{C}) = \{\Theta \in \mathrm{GL}_{2n}(\mathbb{C}) \mid \Theta J\Theta^{\dagger} = J\},$$

and call it the *complex symplectic group*. Its Lie-algebra is

$$\widehat{\mathfrak{sp}}(2n, \mathbb{C}) = \{X \in \mathfrak{gl}_{2n}(\mathbb{C}) \mid JX = -X^\dagger J\}.$$

and we call it the space of *complex Hamiltonian matrix*. The block-structure of $\Theta \in \widehat{\mathrm{Sp}}(n)$ and $X \in \widehat{\mathfrak{sp}}(2n, \mathbb{C})$ is

$$\Theta = \begin{bmatrix} Q & R \\ -R & Q \end{bmatrix}, \quad X = \begin{bmatrix} A & B \\ C & -A^\dagger \end{bmatrix}, \tag{3.33}$$

where $Q, R, A, B, C \in \mathbb{C}^{n\times n}$, with $B^\dagger = B$ and $C^\dagger = C$. Thus, the space of complex Hamiltonian matrices is a real vector space of dimension $4n^2$. The *compact complex symplectic group* is the Lie-group

$$\widehat{\mathrm{Sp}}(n) \quad = \mathrm{SU}_{2n} \cap \widehat{\mathrm{Sp}}(2n, \mathbb{C}) = \{\Theta \in \mathrm{SU}_{2n} \mid \Theta J \Theta^\dagger = J\}$$

with Lie-algebra

$$\widehat{\mathfrak{sp}}(n) = \{X \in \mathfrak{su}_{2n} \mid XJ = JX\}.$$

The matrices $\Theta \in \widehat{\mathrm{Sp}}(n)$ and $X \in \widehat{\mathfrak{sp}}(n)$ have the following block structure

$$\Theta = \begin{bmatrix} Q & R \\ -R & Q \end{bmatrix}, \quad X = \begin{bmatrix} A & B \\ -B & A \end{bmatrix}, \tag{3.34}$$

where $Q, R, A, B, C \in \mathbb{C}^{n\times n}$, with $A^\dagger = -A$ and $B^\dagger = B$.

Next, we prove that complex Lagrangian subspaces can be represented by self-adjoint projectors of $\mathbb{C}^{2n}$ with some properties.

Proposition 3.3.3 *There is a one-to-one correspondence between the set of complex Lagrangian subspaces $W \subset \mathbb{C}^{2n}$ and the set of all self-adjoint projectors P of $\mathbb{C}^{2n}$ of rank-n which satisfy $PJP = 0$.*

Proof. First we show that, for every self-adjoint projector $P \in \mathbb{C}^{2n\times 2n}$ of rank n with $PJP = 0$, the subspace $\operatorname{Im} P$ is a Lagrangian subspace. Since $\dim_{\mathbb{C}} \operatorname{Im} P = \operatorname{rank} P = n$, we have to prove only that $\operatorname{Im} P$ is an isotropic subspace, i.e.

$$u^\dagger J v = 0, \text{ for all } \ u, \ v \in \operatorname{Im} P.$$

From $PJP = 0$, for $u = Px, \ v = Py$, it follows that $u^\dagger J v = x^\dagger PJPy = 0$, for all $x, \ y \in \mathbb{C}^{2n}$. Second, we show that for every complex Lagrangian subspace $W \subset \mathbb{C}^{2n}$ there exists a self-adjoint projector $P \in \mathbb{C}^{2n\times sn}$ with $PJP = 0$ such that $W = \operatorname{Im} P$. There exists a rank n self-adjoint projector P of $\mathbb{C}^{2n}$ such that $W = \operatorname{Im} P$. Hence, it remains to show that $PJP = 0$. Since $W = W^\omega$ it follows that

$$u^\dagger J v = 0, \text{ for all } \ u, v \in \operatorname{Im} P$$

and thus, $x^\dagger PJPy = 0$, for all $x, y \in \mathbb{C}^{2n}$. Hence, $PJP = 0$. □

We denote with $\widehat{\mathrm{LG}}(n)$ the set of all complex Lagrangian subspaces of $\mathbb{C}^{2n}$ and call it the *complex Lagrange-Grassmannian*, i.e.

$$\widehat{\mathrm{LG}}(n) := \{P \in \mathbb{C}^{2n\times 2n} \mid P^\dagger = P,\ P = P^2, \operatorname{tr} P = n, PJP = 0\}.$$

In the subsequent statement we show that $\widehat{\mathrm{LG}}(n)$ is indeed a manifold.

Theorem 3.3.4 *The complex Lagrange-Grassmannian* $\widehat{\mathrm{LG}}(n)$ *is a smooth, compact and connected real submanifold of the Grassmann manifold* $\mathrm{Gr}_{\mathbb{C}}(n, 2n)$ *and has dimension* n^2*.*

Proof. The manifold $\widehat{\mathrm{LG}}(n)$ is given as the orbit of the standard projector Π of $\mathbb{C}^{2n}$ with respect to the group action

$$\sigma : \widehat{\mathrm{Sp}}(n) \times \mathrm{Gr}_{\mathbb{C}}(n, 2n) \to \mathrm{Gr}_{\mathbb{C}}(n, 2n), \quad (\Theta, \Pi) \mapsto \Theta^\dagger \Pi \Theta.$$

□

The tangent space of $\widehat{\mathrm{LG}}(n)$ at P is given by

$$\mathrm{T}_P\widehat{\mathrm{LG}}(n) = \{[P, \Omega] \mid \Omega \in \widehat{\mathfrak{sp}}(n)\}$$

and each $P \in \widehat{\mathrm{LG}}(n)$ and $\xi \in \mathrm{T}_P\widehat{\mathrm{LG}}(n)$ can be represented as

$$P = \Theta^\dagger \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \Theta, \quad \xi = \Theta^\dagger \begin{bmatrix} 0 & Z \\ Z & 0 \end{bmatrix} \Theta, \tag{3.35}$$

with $\Theta \in \widehat{\mathrm{Sp}}(n)$ and $Z \in \mathfrak{her}_n$. The tangent vectors of $\widehat{\mathrm{LG}}(n)$ at P can be obtained from the tangent vectors in $\mathrm{T}_P\mathrm{Gr}_{\mathbb{C}}(n, 2n)$ with the following projection

$$\widehat{\pi}_P : \mathrm{T}_P\mathrm{Gr}_{\mathbb{C}}(n, 2n) \to \mathrm{T}_P\mathrm{Gr}_{\mathbb{C}}(n, 2n), \quad \xi \mapsto \frac{\xi + J\xi J}{2}. \tag{3.36}$$

Since $\widehat{\mathrm{LG}}(n)$ is generated by the action of a subgroup of $\mathrm{SU}(n)$ on the complex Grassmannian, it follows that the geodesics, the exponential map and the parallel transport are the restrictions of the same objects from the Grassmannian $\mathrm{Gr}_{\mathbb{C}}(n, 2n)$. In the same way as for the classical real Lagrange-Grassmannians it can be shown that also the QR-coordinates and the Cayley-coordinates on $\widehat{\mathrm{LG}}(n)$ are restrictions of the QR-coordinates and Cayley-coordinates, respectively, from the Grassmann manifold $\mathrm{Gr}_{\mathbb{C}}(n, 2n)$.

3.4 Riemannian structure of the tensor product of Lagrange-Grassmannians

In this section, we extend the classical Lagrange-Grassmannians $\mathrm{LG}_{\mathbb{K}}(n)$ and the complex Lagrange-Grassmannian $\widehat{\mathrm{LG}}(n)$ to a tensor product of Lagrange-Grassmannians and complex Lagrange-Grassmannians, respectively. We will prove that these tensor products are smooth and compact manifolds. Furthermore, we specify an isometry with the direct product of Lagrange-Grassmannians.

Let $J_1, \ldots, J_r$ denote the standard symplectic forms (3.26) on $\mathbb{K}^{2n_1}, \ldots, \mathbb{K}^{2n_r}$ respectively. For simplicity we use the following notations:

$$N := 2n_1 \cdots 2n_r, \quad \mathbf{n} := (n_1, \ldots, n_r), \quad (\mathbf{n}, \mathbf{2n}) := ((n_1, 2n_1), \ldots, (n_r, 2n_r)). \tag{3.37}$$

If r is odd then, the skew-symmetric nondegenerate matrix $J_1 \otimes \cdots \otimes J_r$ defines a symplectic bilinear form on $\mathbb{K}^N \cong \mathbb{K}^{2n_1} \otimes \cdots \otimes \mathbb{K}^{2n_r}$ as

$$\omega : \mathbb{K}^N \times \mathbb{K}^N \to \mathbb{R}, \quad \omega(x, y) = x^\top (J_1 \otimes \cdots \otimes J_r) y. \tag{3.38}$$

If J denotes the standard symplectic form on $\mathbb{K}^N$, then there exists a permutation matrix $\Gamma \in \mathbb{K}^{N \times N}$ such that

$$J = \Gamma (J_1 \otimes \cdots \otimes J_r) \Gamma^\top. \tag{3.39}$$

If r is even, then $J_1 \otimes \cdots \otimes J_r$ is symmetric and hence it does no longer define a symplectic bilinear form on $\mathbb{K}^N$. Thus, from now on, when discussing about symplectic spaces, we will consider r odd.

If $P_1 \in \mathrm{LG}_{\mathbb{K}}(n_1), \ldots, P_r \in \mathrm{LG}_{\mathbb{K}}(n_r)$ are orthogonal projectors corresponding to Lagrangian subspaces $W_1, \ldots, W_r$ in $\mathbb{K}^{2n_1}, \ldots, \mathbb{K}^{2n_r}$, then, the Kronecker product $P_1 \otimes \cdots \otimes P_r$ is an orthogonal projector that corresponds to an isotropic subspace W of $\mathbb{K}^N$. From the rank of $P_1 \otimes \cdots \otimes P_r$ it is evident that W is not a Lagrangian subspace of $\mathbb{K}^N$ nor with respect to $J_1 \otimes \cdots \otimes J_r$, nor with J. We introduce the following notion.

Definition 3.4.1 *A subspace $W \subset \mathbb{K}^N$ is* decomposable Lagrangian subspace *of $\mathbb{K}^N$ if it is the image of a projector of the form $P_1 \otimes \cdots \otimes P_r$ with $P_j \in \mathrm{LG}_{\mathbb{K}}(n_j)$, for $j = 1, \ldots, r$.*

We show that the set of all decomposable Lagrangian subspaces of $\mathbb{K}^N$ that we call *$r-$fold tensor product of Lagrange-Grassmannians*

$$\mathrm{LG}_{\mathbb{K}}^{\otimes}(\mathbf{n}) := \{P_1 \otimes \cdots \otimes P_r \mid P_j \in \mathrm{LG}_{\mathbb{K}}(n_j),\ j = 1, \ldots, r\} \tag{3.40}$$

has a manifold structure. In fact $\mathrm{LG}_{\mathbb{K}}^{\otimes}(\mathbf{n})$ is a Riemannian submanifold of $\mathrm{Gr}_{\mathbb{K}}^{\otimes}(\mathbf{n}, \mathbf{2n})$. The Riemannian metric on $\mathrm{LG}_{\mathbb{R}}^{\otimes}(\mathbf{n})$ and $\mathrm{LG}_{\mathbb{C}}^{\otimes}(\mathbf{n})$ is induced by the inner product (3.17) and (3.6) on $\mathfrak{sym}_{2n_1} \otimes \cdots \otimes \mathfrak{sym}_{2n_r}$ and $\mathfrak{her}_N$, respectively.

Theorem 3.4.2 *The $r-$fold tensor product of Lagrange-Grassmannians $\mathrm{LG}_{\mathbb{K}}^{\otimes}(\mathbf{n})$ is a smooth and compact submanifold of $\mathrm{Gr}_{\mathbb{K}}^{\otimes}(\mathbf{n}, \mathbf{2n})$ with*

$$\dim\, \mathrm{LG}_{\mathbb{R}}^{\otimes}(\mathbf{n}) = \sum_{i=1}^{r} \frac{n_i(n_i+1)}{2} \quad \textit{and} \quad \dim_{\mathbb{R}}\, \mathrm{LG}_{\mathbb{C}}^{\otimes}(\mathbf{n}) = \sum_{i=1}^{r} n_i(n_i+1). \tag{3.41}$$

Moreover, the map

$$\varphi : \mathrm{LG}_{\mathbb{K}}^{\times}(\mathbf{n}) \to \mathrm{LG}_{\mathbb{K}}^{\otimes}(\mathbf{n}), \quad (X_1, \ldots, X_r) \mapsto X_1 \otimes \cdots \otimes X_r,$$

defines a global Riemannian isometry.

Proof. The map

$$\widetilde{\varphi} : \mathrm{Gr}_{\mathbb{K}}^{\times}(\mathbf{n}, \mathbf{2n}) \to \mathrm{Gr}_{\mathbb{K}}^{\otimes}(\mathbf{n}, \mathbf{2n}), \quad (P_1, \dots, P_r) \mapsto P_1 \otimes \cdots \otimes P_r$$

is injective, it follows that also its restriction to the set $\mathrm{LG}_{\mathbb{K}}^{\times}(\mathbf{n})$ remains injective. Moreover, the Lagrange-Grassmannians $\mathrm{LG}_{\mathbb{K}}(n_j)$ are compact manifolds and also $\alpha_j P_j \notin \mathrm{T}_{P_j}\mathrm{LG}_{\mathbb{K}}(n_j)$, for all $1 \neq \alpha \in \mathbb{K}$ and all $j = 1, \dots, r$. From Corollary 2.2.13 it follows that $\mathrm{LG}_{\mathbb{K}}^{\otimes}(\mathbf{n})$ is a submanifold of $\mathrm{Gr}_{\mathbb{K}}^{\otimes}(\mathbf{n}, \mathbf{2n})$. Since $\widetilde{\varphi}$ is a Riemannian global isometry, then φ as well is a global isometry. The dimension of $\mathrm{LG}_{\mathbb{K}}^{\otimes}(\mathbf{n})$ is equal to the dimension of $\mathrm{LG}_{\mathbb{K}}^{\times}(\mathbf{n})$. $\square$

For r odd, we define the symplectic sesquilinear form

$$\widehat{\omega} : \mathbb{C}^N \times \mathbb{C}^{2n} \to \mathbb{C}, \quad \widehat{\omega}(x, y) = x^{\dagger}(J_1 \otimes \cdots \otimes J_r)y.$$

A subspace $W \subset \mathbb{C}^N$ is *decomposable complex Lagrangian subspace* if it is the image of of a projector of the form $P_1 \otimes \cdots \otimes P_r$ with $P_j \in \widehat{LG}(n_j)$, for $j = 1, \dots, r$.

Analogous to the r-fold tensor product of classical Lagrange-Grassmannians, we show that the set of all decomposable complex Lagrangian subspaces of $\mathbb{C}^{2n}$, called the *r-fold tensor product of complex Lagrange-Grassmannians*

$$\widehat{\mathrm{LG}}^{\otimes}(\mathbf{n}) := \{P_1 \otimes \cdots \otimes P_r \mid P_j \in \widehat{\mathrm{LG}}_{n_j},\ j = 1, \dots, r\}$$

is a submanifold of $\mathrm{Gr}_{\mathbb{C}}^{\otimes}(\mathbf{n}, \mathbf{2n})$. Moreover, $\widehat{\mathrm{LG}}^{\otimes}(\mathbf{n})$ is isometric to the *r-fold direct product of complex Lagrange-Grassmannians*

$$\widehat{\mathrm{LG}}^{\times}(\mathbf{n}) := \{(P_1, \dots, P_r) \mid P_j \in \widehat{\mathrm{LG}}_{n_j},\ j = 1, \dots, r\}.$$

Theorem 3.4.3 *The set* $\widehat{\mathrm{LG}}^{\otimes}(\mathbf{n})$ *is a smooth and compact real submanifold of* $\mathrm{Gr}_{\mathbb{C}}^{\otimes}(\mathbf{n}, \mathbf{2n})$ *of dimension*

$$\dim \widehat{\mathrm{LG}}^{\otimes}(\mathbf{n}) = \sum_{i=1}^{r} n_i^2.$$

Moreover, the map

$$\varphi : \widehat{\mathrm{LG}}^{\times}(\mathbf{n}) \to \widehat{\mathrm{LG}}^{\otimes}(\mathbf{n})\ , \qquad (P_1, \dots, P_r) \mapsto P_1 \otimes \cdots \otimes P_r$$

describes a Riemannian isometry between $\widehat{\mathrm{LG}}^{\otimes}(\mathbf{n})$ *and* $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$.

The proof is similar to the proof of Theorem 3.4.2.

Chapter 4

Generalized Rayleigh-quotient on Grassmannians

Several applications in signal processing, data compression, quantum computing, image processing, etc., have a natural description as optimization tasks on the tensor product of Grassmannians. In particular, we show that a certain generalization of the classical Rayleigh-quotient map relates to important optimization problems from numerical linear algebra, which will be detailed in section 4.2. The isometry between the r-fold tensor product of Grassmannians and the direct product of Grassmannians enables to formulate any optimization task on the tensor product equivalently on the direct product of Grassmannians.

The structure of this chapter is as follows: Section 4.1 is dedicated to the problem of optimizing the generalized Rayleigh-quotient ρ_A, including a detailed discussion on the computation of its Riemannian gradient and its Hessian, as well as necessary conditions for the nondegeneracy of its critical points. Moreover, we make an analogy to the classcial Rayleigh-quotient and show by some examples, that unlike the classical case, the generalized Rayleigh-quotient has also local optima.

In section 4.2, we discuss in detail several applications for the optimization task of the generalized Rayleigh-quotient: (i) the best approximation of a tensor with a tensor of lower rank from signal processing, statistics, and pattern recognition [50, 66]; (ii) the Euclidean entanglement measure from quantum computation [15, 56]; (iii) a "chicken-and-egg" problem in computer vision, namely the problem of determining subspaces from noisy data [73]; (iv) a combinatorial problem, which is a generalization of the well-known Brockett matching problems [10].

Using techniques from the transversality theory, we prove in section 4.3 that, the critical points of the generalized Rayleigh-quotient ρ_A are nondegenerate when the parameter A is taken generically in the space of Hermitian matrices. Furthermore, we extend the result to handle also situations when one would like to exclude some of the critical points from the genericity quest. Such a problem is encountered in applications from computer vision, signal processing, image processing, quantum information, which can be expressed as optimizations of ρ_A, where A is only from a thin subset of the space of Hermitian matrices.

The optimization of the generalized Rayleigh-quotient of semi-positive matrices is the topic of section 4.4. An important result in this section underlies a particular property of the Hessian of ρ_A for A semi-positive definite, and is given in Theorem 4.4.2. An immediate corollary of Theorem 4.4.2 states necessary conditions for the nondegeneracy of the critical points of ρ_A. Moreover, we compute a lower bound for the rank of the parameters, such that the critical points of the generalized Rayleigh-quotient are generically nondegenerate. Another key point of section 4.4 concerns the critical points of ρ_A when A is semi-positive definite and of rank-1, i.e. problems (i) and (ii). The result stated in Theorem 4.4.10 says that the critical points of the generalized Rayleigh-quotient satisfying a certain property are generically nondegenerate, and in particular, the global maximizers are generically nondegenerate. As a consequence of Theorem 4.4.10, for problem (ii) we have a complete characterization, i.e. the critical points except the global minimizers are generically nondegenerate and the global minimizers are always degenerate.

4.1 The generalized Rayleigh-quotient

In this section, we introduce the central task of our work, i.e. the optimization of a generalization of the classical Rayleigh-quotient on the r-fold tensor product of Grassmannians. We motivate our optimization task by the large area of applications and stress its difficulty by pointing out crucial differences to the optimization of the classical Rayleigh-quotient.

Let $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ be the $r-$fold tensor product of Grassmannians with $(\mathbf{m},\mathbf{n})$ as in (3.16) and let $A \in \mathfrak{her}_N$, $N = n_1 n_2 \cdots n_r, M = m_1 m_2 \cdots m_r$. In the following, we analyze the constrained optimization problem

$$\max_{\mathbf{P}\in \mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})} \mathrm{tr}(A\mathbf{P}). \tag{4.1}$$

For this purpose, we define the **generalized Rayleigh-quotient** of a matrix A as

$$\rho_A : \mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n}) \to \mathbb{R}, \quad \mathbf{P} \mapsto \mathrm{tr}(A\mathbf{P}). \tag{4.2}$$

We justify the term "generalized Rayleigh-quotient" for the map (4.2) by pointing out that for only one Grassmannian in the tensor product we obtain the classical Rayleigh-quotient $\rho_A(P) = \mathrm{tr}(AP),\ P \in \mathrm{Gr}(m,n)$. The generalized Rayleigh-quotient can also be regarded as the restriction of the classical Rayleigh-quotient on $\mathrm{Gr}(M,N)$ to the subset $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$. In the sequel, we discuss about the similarities and differences between the generalized and the classical Rayleigh-quotient.

Since $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ is a compact manifold, it follows that the optimization problem (4.1) is well-defined and

$$\max_{\mathbf{P}\in \mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})} \mathrm{tr}(A\mathbf{P}) \leq \max_{\mathbf{P}\in \mathrm{Gr}(M,N)} \mathrm{tr}(A\mathbf{P}). \tag{4.3}$$

The next example will show that the inequality (4.3) can be strict.

Example 4.1.1 *We take a matrix* $A \in \mathfrak{her}_8$ *of the form* $A = \mathrm{diag}(1\ 2\ 3\ 7\ 8\ 4\ 5\ 6)$. *We want to compare the maximal value of* $\mathrm{tr}(A\mathbf{P})$ *on* $\mathrm{Gr}(2,8)$ *with the maximal value of* $\mathrm{tr}(A\mathbf{P})$ *on* $\mathrm{Gr}(2,4) \otimes \mathrm{Gr}(1,2)$. *It is know that*

$$\max_{\mathbf{P} \in \mathrm{Gr}(2,8)} \mathrm{tr}(A\mathbf{P}) = 7 + 8.$$

Hence, there exists $\mathbf{P} = \mathrm{diag}(0\ 0\ 0\ 1\ 1\ 0\ 0\ 0) \in \mathrm{Gr}(2,8)$ *such that* $\rho_A(\mathbf{P}) = 15$. *However, there exist no matrices of the form* $P_1 = \mathrm{diag}(1\ 1\ 0\ 0)$ *and* $P_2 = \mathrm{diag}(1\ 0)$ *and any possible permutation of the* 0*s and* 1*s in* P_1 *and* P_2*, such that* $\mathbf{P} = P_1 \otimes P_2$. *Thus, the maximal value of the generalized Rayleigh-quotient is strictly smaller than* 15.

It is well known that under the assumption that there is a spectral gap between the eigenvalues of $A \in \mathfrak{her}_N$, there is a unique maximizer and a unique minimizer of the classical Rayleigh-quotient of A. Unfortunately, this is no longer the case for the generalized Rayleigh-quotient ρ_A. Global maximizers and global minimizers exist since the generalized Rayleigh-quotient is defined on a compact manifold, but unlike the classical case, it admits also local extrema as we will show in the next example. Because we do not want to enter into too many details in order to be able to give the example, we have to clarify first some relations between the objects we work with.

Problem (4.1) comprises problems from different areas, such as *multilinear low-rank approximations of a tensor*, *geometric measures of entanglement*, *subspace clustering* and *combinatorial optimization*. These applications are naturally stated on a tensor product space. However, for the special case of the Grassmannian they can be reformulated on a direct product space. By abuse of notation we will define the map

$$\rho_A : \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n}) \to \mathbb{R}, \quad \rho_A(P_1,\dots,P_r) := \mathrm{tr}\left(A(P_1 \otimes \cdots \otimes P_r)\right) \tag{4.4}$$

and call it as well *the generalized Rayleigh-quotient* of A. From now on, whenever we discuss about the generalized Rayleigh-quotient, we have in mind the map (4.4). Based on the isometry (3.21) between $\mathrm{Gr}^{\otimes}(\mathbf{m},\mathbf{n})$ and $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$, the optimization problem (4.1) has the equivalent statement

$$\max_{(P_1,\dots,P_r) \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})} \rho_A(P_1,\dots,P_r). \tag{4.5}$$

Now, we can give our example to prove that the generalized Rayleigh-quotient has also local extrema. For the case when A is of rank-1 we refer to Example 3 in [51].

Example 4.1.2 *Let* $A = \mathrm{diag}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) \in \mathfrak{her}_4$ *be a diagonal matrix with* $\lambda_2 > \lambda_3 > \lambda_4 > \lambda_1$ *and* $P_1^*,\ P_2^* \in \mathrm{Gr}_{1,2}$ *of the form*

$$P_1^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad P_2^* = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

The maximum of ρ_A *is obvious less or equal to* λ_2. *Since* $\rho_A(P_1^*,P_2^*) = \lambda_2$*, we have* (P_1^*,P_2^*) *as the global maximizer of* ρ_A. *From* (4.13) *it follows that all* $(P_1,P_2) \in$

$\mathrm{Gr}(1,2) \times \mathrm{Gr}(1,2)$ *with* P_1 *and* P_2 *diagonal, are critical points of* ρ_A*. In particular* (P_2^*, P_1^*) *is a critical point of* ρ_A *with* $\rho_A(P_2^*, P_1^*) = \lambda_3 < \lambda_2$*. Moreover, one can check by computing the Hessian of* ρ_A *at* (P_2^*, P_1^*) *, see* (4.19)*, that* (P_2^*, P_1^*) *is actually a local maximizer of* ρ_A*.*
This implies that the generalized Rayleigh-quotient has local extrema, unlike the classical Rayleigh-quotient. This is a result of the fact that not all 4×4 *permutation matrices are of the form* $\Theta_1 \otimes \Theta_2$*, with* $\Theta_1,\ \Theta_2 \in \mathrm{SU}(2)$*.*

While for the classical Rayleigh-quotient one knows that the maximizer and minimizer are orthogonal projectors onto the space spanned by the eigenvectors corresponding to the largest and smallest eigenvalues of A, respectively, it is difficult to provide an analog characterization for the global extrema of the generalized Rayleigh-quotient for an arbitrary matrix A. Hence, in what follows, we give a detailed description of the fundamental geometric objects necessary to develop Riemannian algorithms to tackle the optimization problem (4.5).

4.1.1 Riemannian optimization of the generalized Rayleigh-quotient

In this section we derive clear expressions for the gradient and the Hessian of the generalized Rayleigh-quotient of a matrix $A \in \mathfrak{her}_N$ on $\mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$. Thus, in the following lemma we establish multilinear maps $\Psi_{A,j}$, which will help us to achieve this purpose. As before, $(\mathbf{m}, \mathbf{n})$ stands for the multi-index $((m_1, n_1), \dots, (m_r, n_r))$ and $N = n_1 n_2 \cdots n_r$.

Lemma 4.1.3 *Let* $A \in \mathbb{C}^{N \times N}$ *and* $(X_1, \dots, X_r) \in \mathbb{C}^{n_1 \times n_1} \times \cdots \times \mathbb{C}^{n_r \times n_r}$*. Then, for all* $j = 1, \dots, r$ *there exists a unique map* $\Psi_{A,j} : \mathbb{C}^{n_1 \times n_1} \times \cdots \times \mathbb{C}^{n_r \times n_r} \to \mathbb{C}^{n_j \times n_j}$ *such that*

$$\mathrm{tr}\Big(A^\dagger (X_1 \otimes \cdots \otimes X_j Z^\dagger \otimes \cdots \otimes X_r)\Big) = \mathrm{tr}\Big(\Psi_{A,j}(X_1, \dots, X_r)^\dagger Z\Big) \tag{4.6}$$

holds for all $Z \in \mathbb{C}^{n_j \times n_j}$*. In particular, one has*

$$\begin{aligned} \mathrm{tr}\Big(A^\dagger (X_1 \otimes \cdots \otimes X_r)\Big) &= \mathrm{tr}\Big(\Psi_{A,1}(I_{n_1}, X_2, \dots, X_r)^\dagger X_1\Big) \\ &= \cdots = \mathrm{tr}\Big(\Psi_{A,r}(X_1, \dots, X_{r-1}, I_{n_r})^\dagger X_r\Big). \end{aligned} \tag{4.7}$$

Moreover, for $A := A_1 \otimes \cdots \otimes A_r$ *the maps* $\Psi_{A,j}$ *exhibit the explicit form*

$$\Psi_{A,j}(X_1, \dots, X_r) = \Bigg(\prod_{k=1,\, k \neq j}^{r} \mathrm{tr}(X_k^\dagger A_k)\Bigg) A_j. \tag{4.8}$$

Proof. Fix j and consider the linear functional

$$Z \mapsto \lambda_A(Z) := \mathrm{tr}\Big(A^\dagger (X_1 \otimes \cdots \otimes X_j Z \otimes \cdots \otimes X_r)\Big).$$

By the Riesz representation theorem, there exists a unique $B_j \in \mathbb{C}^{n_j \times n_j}$ such that $\lambda_A(Z) = \operatorname{tr}\left(B_j^\dagger Z\right)$ for all $Z \in \mathbb{C}^{n_j \times n_j}$. Therefore, the map $\Psi_{A,j}$ is given by $(X_1, \ldots, X_r) \mapsto \Psi_{A,j}(X_1, \ldots, X_r) := B_j$. It is straightforward to show that $\Psi_{A,j}$ is multilinear in $X_1, \ldots, X_r$. Now, choosing $Z := X_j$ and $X_j := I_{n_j}$ in (4.6) immediately yields (4.7). Moreover, (4.8) follows from the trace equality

$$\operatorname{tr}\left(A_1^\dagger X_1 \otimes \cdots \otimes A_j^\dagger X_j Z \otimes \cdots \otimes A_r^\dagger X_r\right) = \Bigg(\prod_{k=1,\, k\neq j}^{r} \operatorname{tr}(A_k^\dagger X_k)\Bigg) \operatorname{tr}(A_j^\dagger X_j Z).$$

Thus the proof of Lemma 4.1.3 is complete. $\square$

Remark 4.1.4 *The linear maps $\Psi_{A,j}$ constructed in the above proof are almost identical to the so-called partial trace operators — a well-known concept from multilinear algebra and quantum mechanics (see [7]).*

Next, we show how to compute $\Psi_{A,j}(X_1, \ldots, X_r)$ for given $(X_1, \ldots, X_r) \in \mathbb{C}^{n_1 \times n_1} \times \cdots \times \mathbb{C}^{n_r \times n_r}$ if A is not a pure tensor product $A_1 \otimes \cdots \otimes A_r$.

Lemma 4.1.5 *Let $A \in \mathbb{C}^{N \times N}$ and $(X_1, \ldots, X_r) \in \mathbb{C}^{n_1 \times n_1} \times \cdots \times \mathbb{C}^{n_r \times n_r}$. Then, the (s,t)-entry of $\Psi_{A,j}(X_1, \ldots, X_r) \in \mathbb{C}^{n_j \times n_j}$ is given by*

$$\sum_{\substack{i_q=1,\, q\neq j \\ q=1,\ldots,r}}^{n_l} (\mathrm{e}_{i_1}^\top \otimes \cdots \otimes \mathrm{e}_s^\top \otimes \cdots \otimes \mathrm{e}_{i_r}^\top) A^\dagger (X_1 \otimes \cdots \otimes X_r)(\mathrm{e}_{i_1} \otimes \cdots \otimes \mathrm{e}_t \otimes \cdots \otimes \mathrm{e}_{i_r}), \quad (4.9)$$

where $\{\mathrm{e}_t\}_{t=1}^{n_l}$ denotes the standard basis of $\mathbb{C}^{n_l}$.

Proof. Let $1 \leq s,\ t \leq n_j$. Then, the element in the (s,t) position of the transpose conjugate of the matrix $\Psi_{A,j}(X_1, \ldots, X_r)$ is given by

$$\begin{aligned}
\mathrm{e}_t^\top (\Psi_{A,j}(X_1, \ldots, X_r))^\dagger \mathrm{e}_s &= \operatorname{tr}\Big(\Psi_{A,j}(X_1, \ldots, X_r)^\dagger \mathrm{e}_s \mathrm{e}_t^\top\Big) \\
&= \operatorname{tr}\Big(A^\dagger (X_1 \otimes \cdots \otimes X_j \mathrm{e}_s \mathrm{e}_t^\top \otimes \cdots \otimes X_r)\Big) \\
&= \operatorname{tr}\Big(A^\dagger (X_1 \otimes \cdots \otimes X_r)(I_{k_1} \otimes \cdots \otimes \mathrm{e}_s \mathrm{e}_t^\top \otimes \cdots \otimes I_{k_r})\Big).
\end{aligned}$$

Hence, (4.9) follows from the identity $I_{n_q} = \sum_{i_q=1}^{n_q} \mathrm{e}_{i_q} \mathrm{e}_{i_q}^\top$. $\square$

A useful property of the multilinear map $\Psi_{A,j}$ is given in what follows.

Lemma 4.1.6 *Let $A \in \mathfrak{her}_N$ and $(X_1, \ldots, X_r) \in \mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$. Then,*

$$\Psi_{A,j}(X_1, \ldots, I_{n_j}, \ldots, X_r)$$

is a Hermitian matrix for all $j = 1, \ldots, r$.

Proof. The result is a straightforward consequence of the identity

$$\operatorname{tr}\Big(A(X_1 \otimes \cdots \otimes Z \otimes \cdots \otimes X_r)\Big)^\dagger = \operatorname{tr}\Big(A(X_1 \otimes \cdots \otimes Z^\dagger \otimes \cdots \otimes X_r)\Big), \tag{4.10}$$

for all $Z \in \mathbb{C}^{n_j \times n_j}$. □

Now, we can give an explicit formula for the Riemannian gradient of ρ_A and derive necessary and sufficient critical point conditions for it. For simplicity of writing, whenever $(P_1, \ldots, P_r) \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ is understood from the context, we use the following shortcut

$$\widehat{A}_j := \Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r). \tag{4.11}$$

Theorem 4.1.7 *Let $A \in \mathfrak{her}_N$, $P := (P_1, \ldots, P_r) \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ and let ρ_A be the generalized Rayleigh-quotient on $\mathrm{Gr}^{\times r}(\mathbf{m}, \mathbf{n})$. Then, one has the following:*

(i) The gradient of ρ_A at P with respect to the Riemannian metric (3.19) is

$$\operatorname{grad}\rho_A(P) = \left(\operatorname{ad}^2_{P_1}\widehat{A}_1, \ldots, \operatorname{ad}^2_{P_r}\widehat{A}_r\right). \tag{4.12}$$

(ii) The critical points of ρ_A on $\mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ are characterized by

$$[P_j, \widehat{A}_j] = 0, \tag{4.13}$$

for $j = 1, \ldots, r$, i.e. P_j is the orthogonal projector onto an m_j-dimensional invariant subspace of $\widehat{A}_j$.

Proof. (i) Fix $P := (P_1, \ldots, P_r) \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ and let $\widetilde{\rho}_A$ denote the canonical smooth extension of ρ_A to $\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$. Then,

$$D\widetilde{\rho}_A(P)(X) = \sum_{j=1}^{r} \operatorname{tr}\Big(A(P_1 \otimes \cdots \otimes X_j \otimes \cdots \otimes P_r)\Big) = \sum_{j=1}^{r} \operatorname{tr}(\widehat{A}_j X_j),$$

for all $X := (X_1, \ldots, X_r) \in \mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$. From (3.19), we obtain that the gradient of $\widetilde{\rho}_A$ at P is given by

$$\nabla\widetilde{\rho}_A(P) = (\widehat{A}_1, \ldots, \widehat{A}_r).$$

Thus, according to (3.7) and (2.9),

$$\operatorname{grad}\,\rho_A(P) = \left(\operatorname{ad}^2_{P_1}\widehat{A}_1, \ldots, \operatorname{ad}^2_{P_r}\widehat{A}_r\right).$$

(ii) $P := (P_1, \ldots, P_r) \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ is a critical point of ρ_A if and only if

$$\operatorname{grad}\,\rho_A(P) = 0.$$

This is equivalent to

$$P_j[P_j, \widehat{A}_j] = [P_j, \widehat{A}_j]P_j, \tag{4.14}$$

for all $j = 1, \ldots, r$. By multiplying (4.14) once from the left with P_j and once from the right with P_j, we obtain that $P_j\widehat{A}_j = P_j\widehat{A}_jP_j$ and $\widehat{A}_jP_j = P_j\widehat{A}_jP_j$. Hence, the conclusion $[P_j, \widehat{A}_j] = 0$ holds for all $j = 1, \ldots, r$. $\square$

As a consequence of Theorem 4.1.7, we obtain the following necessary and sufficient critical point condition.

Corollary 4.1.8 *Let $A \in \mathfrak{her}_N$, $P := (P_1, \ldots, P_r) \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ and let $\Theta_j \in \mathrm{SU}(n_j)$ be such that $\Theta_j P_j \Theta_j^{\dagger} = \Pi_j$, where Π_j is the standard projector in Gr_{m_j,n_j}. We write*

$$\Theta_j \widehat{A}_j \Theta_j^{\dagger} = \begin{bmatrix} \Psi_j' & \Psi_j''' \\ \Psi_j'''^{\dagger} & \Psi_j'' \end{bmatrix}, \tag{4.15}$$

with $\Psi_j' \in \mathfrak{her}_{m_j}$, $\Psi_j'' \in \mathfrak{her}_{n_j - m_j}$, and $\Psi_j''' \in \mathbb{C}^{m_j \times (n_j - m_j)}$. Then, P is a critical point of ρ_A if and only if

$$\Psi_j''' = 0, \tag{4.16}$$

for all $j = 1, \ldots, r$. Moreover, for any $P \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ the following holds

$$\mathrm{tr}(\Psi_1') = \cdots = \mathrm{tr}(\Psi_r') = \rho_A(P). \tag{4.17}$$

In the case when $A \in \mathfrak{her}_N$ can be diagonalized by elements in

$$\mathrm{SU}(\mathbf{n}) = \{\Theta_1 \otimes \cdots \otimes \Theta_r \mid \Theta_j \in \mathrm{SU}(n_j),\ j = 1, \ldots, r\}, \tag{4.18}$$

it is possible to give an explicit characterization of the critical points of ρ_A. We can assume without loss of generality that A is diagonal and distinguish two possibilities:

(1) A can be written as $\Lambda_1 \otimes \cdots \otimes \Lambda_r$, with Λ_j diagonal;

(2) A cannot be written as a Kronecker product of diagonal matrices

The first case arises when $A = A_1 \otimes \cdots \otimes A_r$, $A_j \in \mathfrak{her}_{n_j}$, and in this instance, the generalized Rayleigh-quotient becomes a product of r decoupled classical Rayleigh-quotients with one maximizer and one minimizer. However, there is a dramatic change if A cannot be written as a Kronecker product of diagonal matrices. This situation is encountered in Example 4.1.2 and the conclusion is that the generalized Rayleigh–qutient has also local extrema. Next, we give a sufficient critical point condition for the case when the matrix A is diagonal. Before that, we underlie a property of the Kronecker product that will be used to prove the mentioned critical point condition.

Lemma 4.1.9 *Let $X_1 \in \mathbb{C}^{n_1 \times n_1} \setminus \{0\}, \ldots, X_r \in \mathbb{C}^{n_r \times n_r} \setminus \{0\}$. Then, $X_1 \otimes \cdots \otimes X_r$ is diagonal if and only if $X_1, \ldots, X_r$ are diagonal.*

Proof. The proof follows by induction over r. For $r = 1$ the conclusion is obvious. Assume that the conclusion of the lemma holds for r and we show it for $r+1$. Then, from $(X_1 \otimes \cdots \otimes X_r) \otimes X_{r+1}$ diagonal with $X_{r+1} \neq 0$, it follows that $X_1 \otimes \cdots \otimes X_r$ and X_{r+1} are diagonal. From the previous induction step we know that $X_1, \dots, X_r$ are also diagonal and hence the proof. $\square$

Corollary 4.1.10 *Let $A \in \mathfrak{her}_N$ be diagonal. Then, $(P_1, \dots, P_r) \in \mathrm{Gr}^{\times r}(\mathbf{m}, \mathbf{n})$ with P_j any permutation of the standard projector Π_j, for $j = 1, \dots, r$, is a critical point of ρ_A.*

Proof. If $D \in \mathfrak{her}_N$ is diagonal, then

$$\mathrm{tr}(D(X_1 \otimes \cdots \otimes X_r)) = \mathrm{tr}(D\ \mathrm{diag}(X_1 \otimes \cdots \otimes X_r)).$$

Moreover, from Lemma 4.1.9 it follows that $X_1 \otimes \cdots \otimes X_r \neq 0$ is diagonal if and only if $X_1 \in \mathfrak{her}_{n_1} \setminus \{0\}, \dots, X_r \in \mathfrak{her}_{n_r} \setminus \{0\}$ are diagonal. Let $P_1, \dots, P_r$ be permutations of the standard projectors, i.e. diagonal. Since A is diagonal, it follows that $\widehat{A}_1, \dots, \widehat{A}_r$ are diagoanl and hence,

$$[P_j, \widehat{A}_j] = 0,$$

for $j = 1, \dots, r$. The conclusion follows from Theorem 4.1.7. $\square$

For the rest of this section we are concerned with the computation of the Riemannian Hessian of ρ_A and give also necessary conditions for its nondegeneracy at critical points.

Theorem 4.1.11 *Let $A \in \mathfrak{her}_N$ and $P := (P_1, \dots, P_r) \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$. Then, the Riemannian Hessian of ρ_A at P is the unique self-adjoint operator*

$$\begin{aligned} &\mathbf{H}_{\rho_A}(P) : \mathrm{T}_P\mathrm{Gr}^{\times r}(\mathbf{m}, \mathbf{n}) \to \mathrm{T}_P\mathrm{Gr}^{\times r}(\mathbf{m}, \mathbf{n}), \\ &\xi := (\xi_1, \dots, \xi_r) \mapsto \mathbf{H}_{\rho_A}(P)(\xi) := \Big(\mathbf{H}_1(\xi), \dots, \mathbf{H}_r(\xi)\Big), \end{aligned} \tag{4.19}$$

defined by

$$\mathbf{H}_j(\xi) := -\mathrm{ad}_{P_j}\mathrm{ad}_{\widehat{A}_j}\xi_j + \sum_{k=1, k\neq j}^{r} \mathrm{ad}^2_{P_j}\Psi_{A,j}(P_1, \dots, I_{n_j}, \dots, \xi_k, \dots, P_r), \tag{4.20}$$

where $\widehat{A}_j := \Psi_{A,j}(P_1, \dots, I_{n_j}, \dots, P_r)$.

Proof. Let $(\widetilde{\mathfrak{X}}_1, \dots, \widetilde{\mathfrak{X}}_r)$ denote a smooth extension of $\mathrm{grad}\rho_A$ to $\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$. According to (4.12), we can choose

$$P \mapsto \widetilde{\mathfrak{X}}_j(P) = \mathrm{ad}^2_{P_j}\widehat{A}_j.$$

Then,

$$\begin{aligned} D\widetilde{\mathfrak{X}}_j(P)(X) &= \mathrm{ad}_{X_j}\mathrm{ad}_{P_j}\widehat{A}_j + \mathrm{ad}_{P_j}\mathrm{ad}_{X_j}\widehat{A}_j \\ &+ \sum_{k=1,k\neq j}^{r} \mathrm{ad}^2_{P_j}\Psi_{A,j}(P_1,\dots,I_{n_j},\dots,X_k,\dots,P_r), \end{aligned}$$

for all $P := (P_1,\dots,P_r)$ and $X := (X_1,\dots,X_r)$ in $\mathfrak{her}_{n_1}\times\cdots\times\mathfrak{her}_{n_r}$. Notice that the derivative of the linear map $P_k \mapsto \Psi_{A,j}(P_1,\cdots,I_{n_j},\dots,P_k,\dots,P_r)$ in direction $X_k \in \mathfrak{her}_{n_k}$ $(k\neq j)$ is $\Psi_{A,j}(P_1,\dots,I_{n_j},\dots,X_k,\dots,P_r)$. Applying (3.7) and (2.9), the Riemannian Hessian of ρ_A at $P\in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ is given by

$$\mathbf{H}_j(\xi) = -\mathrm{ad}_{P_j}\mathrm{ad}_{\widehat{A}_j}\xi_j + \sum_{k=1,k\neq j}^{r} \mathrm{ad}^2_{P_j}\Psi_{A,j}(P_1,\dots,I_{n_j},\dots,\xi_k,\dots,P_r),$$

for all $\xi := (\xi_1,\dots,\xi_r)\in \mathrm{T}_P\mathrm{Gr}^{\times r}(\mathbf{m},\mathbf{n})$. Here, we have used the following two facts:

(i) Clearly, $\mathrm{ad}_{\widehat{A}_j}\xi_j$ is skew-hermitian and hence

$$-\mathrm{ad}_{P_j}\mathrm{ad}_{\widehat{A}_j}\xi_j = \mathrm{ad}_{P_j}\mathrm{ad}_{\xi_j}\widehat{A}_j$$

is in the tangent space $\mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$ for all $\xi_j\in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$.

(ii) A straightforward computation shows that $\mathrm{ad}_{\xi_j}\mathrm{ad}_{P_j}\widehat{A}_j$ is in the orthogonal complement of $\mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$ and hence

$$\mathrm{ad}^2_{P_j}\mathrm{ad}_{\xi_j}\mathrm{ad}_{P_j}\widehat{A}_j = 0$$

for all $\xi_j\in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$. □

Recall from [32] that for the classical Rayleigh-quotient, the Hessian of ρ_A at $P\in\mathrm{Gr}_{m,n}$ is nondegenerate if and only if the Sylvester equation

$$\Psi' Z - Z\Psi'' = 0$$

has only the trivial solution $0 = Z\in\mathbb{C}^{m\times(n-m)}$. The matrices Ψ' and Ψ'' are defined in (4.15) by taking into account that $r=1$. Exploiting this fact, in the case of the generalized Rayleigh-quotient we obtain only a necessary condition for the nondegeneracy of the Hessian in a local maximizer or a local minimizer, as the next theorem will prove.

Theorem 4.1.12 *Let $A\in\mathfrak{her}_N$, and $P\in\mathrm{Gr}^{\times r}(\mathbf{m},\mathbf{n})$ be a local maximizer (local minimizer) of ρ_A. If $\mathbf{H}_{\rho_A}(P)$ is nondegenerate, then for all $j=1,\dots,r$ the equality*

$$\sigma(\Psi_j')\cap\sigma(\Psi_j'') = \emptyset, \tag{4.21}$$

holds with Ψ_j' and Ψ_j'' as in (4.15). Here, $\sigma(X)$ denotes the spectrum of X.

Proof. Let $P := (P_1, \ldots, P_r) \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ be a local maximizer of ρ_A. Since $\mathbf{H}_{\rho_A}(P)$ is nondegenerate it follows that is negative definite and hence, $\mathbf{H}_j$ restricted to $\{(0, \ldots, \xi_j, \ldots, 0) \mid \xi_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}\}$ is negative definite. Hence

$$\mathbf{H}_j(0, \ldots, \xi_j, \ldots, 0) \neq 0,$$

for all $0 \neq \xi_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$ and $j = 1, \ldots, r$. From (4.20) it follows that

$$\mathrm{ad}_{P_j}\mathrm{ad}_{\widehat{A}_j}\xi_j \neq 0, \tag{4.22}$$

for all $0 \neq \xi_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$. Now, according to Corollary 4.1.8, there exists $\Theta_j \in \mathrm{SU}(n_j)$ for all $j = 1, \ldots, r$ such that $P_j = \Theta_j^{\dagger}\Pi_j\Theta_j$ and

$$\Theta_j\widehat{A}_j\Theta_j^{\dagger} = \begin{bmatrix} \Psi_j' & 0 \\ 0 & \Psi_j'' \end{bmatrix}, \ \Psi_j' \in \mathfrak{her}_{m_j}, \ \Psi_j'' \in \mathfrak{her}_{n_j-m_j}.$$

Hence, from (3.31) it follows that (4.22) is equivalent to the fact that the Sylvester equation

$$\Psi_j'Z_j - Z_j\Psi_j'' = 0 \tag{4.23}$$

has only the trivial solution $0 = Z_j \in \mathbb{C}^{m_j\times(n_j-m_j)}$, for all $j = 1, \ldots, r$. This, in turn, is well-known to hold if and only if the intersection of the spectra of Ψ_j' and Ψ_j'' is empty. □

Remark 4.1.13 *If $A \in \mathfrak{her}_N$ can be diagonalized by elements in $\mathrm{SU}(\mathbf{n})$, then condition (4.21) is also sufficient for the nondegeneracy of the Hessian of ρ_A at local extrema. In this situation the Hessian of ρ_A at critical points is block-diagonal.*

4.2 Applications of the generalized Rayleigh-quotient

There is a wide range of applications for problem (4.5) in areas such as signal processing, computer vision and quantum information. We illustrate the broad potential of (4.5) by four examples.

4.2.1 Best multilinear rank-$(m_1, \ldots, m_r)$ tensor approximation

The problem of best approximation of a tensor by a tensor of lower rank is important in areas such as statistics, signal processing and pattern recognition. Unlike in the matrix case, there are several rank concepts for a higher order tensor, [47, 51, 66]. For the scope of this paper, we focus on the *multilinear rank* case.

A finite dimensional complex tensor $\mathcal{A}$ of order r is an element of a tensor product $V_1 \otimes \cdots \otimes V_r$, where $V_1, \ldots, V_r$ are complex vector spaces with dim $V_j = n_j$. Such an element can have various representations, a common one is the description as an $r-$way array, i.e. after a choice of bases for $V_1, \ldots, V_r$, the tensor $\mathcal{A}$ is identified with

$[a_{i_1\dots i_r}]_{i_1=1,\dots,i_r=1}^{n_1,\dots,n_r} \in \mathbb{C}^{n_1\times n_2\times\cdots\times n_r}$, see e.g. [66]. The $j-$th way of the array is referred to as the $j-$th *mode* of $\mathcal{A}$. A matrix $X \in \mathbb{C}^{q_j\times n_j}$ acts on a tensor $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times\cdots\times n_r}$ via *mode$-j$ multiplication* $\times_j$ and the result is a tensor in $\mathbb{C}^{n_1\times\cdots\times q_j\times\cdots\times n_r}$, i.e.

$$(\mathcal{A}\times_j X)_{i_1\dots i_{j-1}k_1 i_{j+1}\dots i_r} = \sum_{k_2=1}^{n_j} a_{i_1\dots i_{j-1}k_2 i_{j+1}\dots i_r} x_{k_1 k_2}, \tag{4.24}$$

cf. [50, 66]. Moreover, given $\mathcal{A} \in \mathbb{C}^{n_1\times n_2\times\cdots\times n_r}$, $X \in \mathbb{C}^{q_j\times n_j}$ and $Y \in \mathbb{C}^{q_k\times n_k}$, from (4.24) one has

$$\mathcal{A}\times_j X\times_k Y = (\mathcal{A}\times_j X)\times_k Y = (\mathcal{A}\times_k Y)\times_j X. \tag{4.25}$$

It is always possible to rearrange the elements of $\mathcal{A}$ along one or, more general, several modes such that they form a matrix. Let $l_1,\dots,l_q$ and $c_1,\dots,c_p$ be ordered subsets of $1,\dots,r$ such that $\{l_1,\dots,l_q\}\cup\{c_1,\dots,c_p\} = \{1,\dots,r\}$. Moreover, consider the products $N_k := n_{l_{k+1}}\cdots n_{l_q}$, $N'_k := n_{c_{k+1}}\cdots n_{c_p}$, for $k = 0,\dots,q-1$ and $k = 0,\dots,p-1$, respectively. Then, the *matrix unfolding* of $\mathcal{A}$ along $(l_1,\dots,l_q)$ is a matrix $A_{(l_1,\dots,l_q)}$ of size $N_0\times N'_0$ such that the element in position $(i_1,\dots,i_r)$ of $\mathcal{A}$ moves to position (s,t) in $A_{(l_1,\dots,l_q)}$, where

$$s := i_{l_q} + \sum_{k=1}^{q-1}(i_{l_k}-1)N_k \qquad \text{and} \qquad t := i_{c_p} + \sum_{k=1}^{p-1}(i_{c_k}-1)N'_k. \tag{4.26}$$

As an example, for a third order tensor $\mathcal{A} \in \mathbb{C}^{2\times2\times2}$ we obtain the following matrix unfoldings as in [50]

$$A_{(1)} = \begin{bmatrix} a_{111} & a_{112} & a_{121} & a_{122} \\ a_{211} & a_{212} & a_{221} & a_{222} \end{bmatrix},\quad A_{(2)} = \begin{bmatrix} a_{111} & a_{112} & a_{211} & a_{212} \\ a_{121} & a_{122} & a_{221} & a_{222} \end{bmatrix},$$

$$A_{(3)} = \begin{bmatrix} a_{111} & a_{121} & a_{211} & a_{221} \\ a_{112} & a_{122} & a_{212} & a_{222} \end{bmatrix}.$$

The *multilinear rank* of $\mathcal{A} \in \mathbb{C}^{n_1\times\cdots\times n_r}$ is the $r-$tuple $(m_1,\dots,m_r)$ such that

$$m_1 = \text{rank } A_{(1)}\ ,\ \dots\ ,\ m_r = \text{rank } A_{(r)}. \tag{4.27}$$

To refer to the multilinear rank of $\mathcal{A}$ we will use the notation rank-$(m_1,\dots,m_r)$ or rank $\mathcal{A} = (m_1,\dots,m_r)$. Given a tensor $\mathcal{A} \in \mathbb{C}^{n_1\times\cdots\times n_r}$, we are interested in finding the best rank-$(m_1,\dots,m_r)$ approximation of $\mathcal{A}$ (assuming that $\mathcal{A}$ has rank greater or equal to $(m_1,\dots,m_r)$), i.e.

$$\min_{\text{rank}(\mathcal{B})\le(m_1,\dots,m_r)} \|\mathcal{A}-\mathcal{B}\|. \tag{4.28}$$

Here, $\|\mathcal{A}\|$ is the Frobenius norm of a tensor, i.e. $\|\mathcal{A}\|^2 = \langle\mathcal{A},\mathcal{A}\rangle$ with

$$\langle\mathcal{A},\mathcal{B}\rangle = \text{vec}(\mathcal{A})^\dagger\text{vec}(\mathcal{B}) = \sum_{i_1,\dots,i_r=1}^{n_1,\dots,n_r} \bar{a}_{i_1\dots i_r} b_{i_1\dots i_r}. \tag{4.29}$$

Here, $\mathrm{vec}(\mathcal{A})$ refers to the matrix unfolding $A_{(1,\ldots,r)} \in \mathbb{C}^{N\times 1}$.
In the matrix case, the solution of the optimization problem (4.28) is given by a truncated SVD, cf. Eckart-Young theorem [18]. What would be a good generalization of the singular value decomposition such that one could formulate a similar Eckart-Young result for higher order tensors? There are several types of decompositions for tensors available in the literature, see [11, 50, 71], however, there is no equivalent of the Eckart-Young theorem for the higher-order case.
According to the Tucker decomposition [71] or its generalization, the higher order singular value decomposition (HOSVD) [50], any rank-$(m_1,\ldots,m_r)$ tensor can be written as a product of a *core* tensor $\mathcal{S}$ and r Stiefel matrices $X_1 \in \mathbb{C}^{n_1\times m_1},\ldots,\ X_r \in \mathbb{C}^{n_r\times m_r}$, i.e.

$$\mathcal{B} = \mathcal{S} \times_1 X_1 \times_2 \cdots \times_r X_r, \quad X_j^\dagger X_j = I_{m_j},\ j = 1,\ldots,r.$$

Using vec−operation and Kronecker product language, one has

$$\mathrm{vec}(\mathcal{S} \times_1 X_1 \times_2 \cdots \times_r X_r) = (X_1 \otimes \cdots \otimes X_r)\mathrm{vec}(\mathcal{S}). \tag{4.30}$$

Thus, from (4.29) it follows that

$$\begin{aligned} \|\mathcal{A} - \mathcal{B}\|^2 &= \mathrm{vec}(\mathcal{A})^\dagger \mathrm{vec}(\mathcal{A}) - 2\,\mathrm{Re}\Big(\mathrm{vec}(\mathcal{A})^\dagger \mathrm{vec}(\mathcal{B})\Big) + \mathrm{vec}(\mathcal{B})^\dagger \mathrm{vec}(\mathcal{B}) \\ &= \mathrm{vec}(\mathcal{A})^\dagger \mathrm{vec}(\mathcal{A}) - 2\,\mathrm{Re}\Big(\mathrm{vec}(\mathcal{A})^\dagger (X_1 \otimes \cdots \otimes X_r)\mathrm{vec}(\mathcal{S})\Big) + \mathrm{vec}(\mathcal{S})^\dagger \mathrm{vec}(\mathcal{S}). \end{aligned}$$

Let $a := \mathrm{vec}(\mathcal{A}) \in \mathbb{C}^{n_1 n_2 \cdots n_r}$ and $s := \mathrm{vec}(\mathcal{S}) \in \mathbb{C}^{m_1 m_2 \cdots m_r}$ and consider for fixed $(X_1,\ldots,X_r) \in \mathbb{C}^{n_1\times m_1} \times \cdots \times \mathbb{C}^{n_r\times m_r}$ the function

$$f : \mathbb{C}^{m_1 m_2 \cdots m_r} \to \mathbb{R},\ \ f(s) = \|a\|^2 - 2\,\mathrm{Re}\Big(a^\dagger (X_1 \otimes \cdots \otimes X_r) s\Big) + \|s\|^2.$$

The minimal value of f on $\mathbb{C}^{m_1 m_2 \cdots m_r}$ is

$$\|a\|^2 - \|a^\dagger (X_1 \otimes \cdots \otimes X_r)\|^2$$

and hence, solving (4.28) is equivalent to solving the maximization problem problem

$$\max_{X_1,\ldots,X_r} \|\mathrm{vec}(\mathcal{A})^\dagger (X_1 \otimes \cdots \otimes X_r)\|^2,$$

with $X_j^\dagger X_j = I_{m_j},\ j = 1,\ldots,r$. From the properties of the *trace* function, the best multilinear rank-$(m_1,\ldots,m_r)$ approximation problem becomes

$$\max_{(P_1,\ldots,P_r)\in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})} \mathrm{tr}\,\Big(A(P_1 \otimes \cdots \otimes P_r)\Big), \tag{4.31}$$

with $A = \mathrm{vec}(\mathcal{A})\mathrm{vec}(\mathcal{A})^\dagger$ and $P_j = X_j X_j^\dagger,\ j = 1,\ldots,r$.

4.2.2 A geometric measure of entanglement

The task of characterizing and quantifying entanglement is a central theme in quantum information theory. There exist various ways to measure the difference between entangled and product states and we refer to [39, 59] for detailed reviews on quantifying entanglement. Here, we discuss a geometric measure of entanglement, which is given by the Euclidean distance of $z \in \mathbb{C}^N$ with $\|z\| = 1$ to the set of all product states $\mathcal{P} = \{x_1 \otimes \cdots \otimes x_r \mid x_j \in \mathbb{C}^{n_j},\ \|x_j\| = 1\}$, i.e.

$$\delta_{\mathrm{E}}(z) := \min_{x \in \mathcal{P}} \|z - x\|^2. \tag{4.32}$$

Since

$$\|z - x\|^2 = \|z\|^2 - z^\dagger x - x^\dagger z + \|x\|^2 = \|z\|^2 - 2\,\mathrm{Re}(z^\dagger x) + 1,$$

it follows that any minimizer of δ_E is also a maximizer of

$$\max_{x_j \in \mathbb{C}^{n_j},\ \|x_j\|=1} \mathrm{Re}\Big(z^\dagger (x_1 \otimes \cdots \otimes x_r)\Big), \tag{4.33}$$

and vice versa. Moreover, since the maximal value is ≥ 0 and in the critical points we have $z^\dagger (x_1 \otimes \cdots \otimes x_r) \in \mathbb{R}$ it follows that (4.33) is equivalent to

$$\max_{x_j \in \mathbb{C}^{n_j},\ \|x_j\|=1} |z^\dagger (x_1 \otimes \cdots \otimes x_r)|. \tag{4.34}$$

Hence, computing the entanglement measure (4.32) is equivalent to solving

$$\max_{(P_1,\ldots,P_r) \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})} \mathrm{tr}\,\Big(A(P_1 \otimes \cdots \otimes P_r)\Big), \tag{4.35}$$

with $A = zz^\dagger$ and $P_1 = x_1 x_1^\dagger, \ldots, P_r = x_r x_r^\dagger$. Note that (4.35) actually constitutes a best rank$-(1,\ldots,1)$ tensor approximation problem [15].

4.2.3 Subspace clustering

Subspace segmentation is a fundamental problem in many applications in computer vision (e.g. image segmentation) and image processing (e.g. image representation and compression), see [73, 75]. The problem of clustering data lying on multiple subspaces of different dimensions can be stated as follows:

Given a set of data points $X = \{x_l \in \mathbb{R}^n\}_{j=1}^L$ which lie approximately in $r \geq 1$ distinct subspaces S_k of dimension d_k, $1 \leq d_k < n$, identify the subspaces S_k without knowing in advance which points belong to which subspace.

Every d_k dimensional subspace $S_k \subset \mathbb{R}^n$ can be defined as the kernel of a rank $m_k = n - d_k$ orthogonal projector P_k of $\mathbb{R}^{n_k}$, with $n_k = n$ as

$$S_k = \{x \in \mathbb{R}^n \mid P_k x = 0\}.$$

Therefore, any point $x \in \bigcup_{k=1}^{r} S_k$ satisfies

$$\|P_1 x\| \cdot \|P_2 x\| \cdots \|P_r x\| = 0,$$

which is equivalent to

$$\operatorname{tr}(xx^\top P_1)\operatorname{tr}(xx^\top P_2)\cdots\operatorname{tr}(xx^\top P_r) = \operatorname{tr}\Big((xx^\top \otimes \cdots \otimes xx^\top)(P_1 \otimes \cdots \otimes P_r)\Big) = 0.$$

Thus, the problem of recovering the subspaces S_k from the data points X can be treated as the following optimization task:

$$\min_{P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})} \sum_{l=1}^{L} \prod_{k=1}^{r} \|P_k x_l\|^2 \quad = \min_{P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})} \operatorname{tr}\Big(A(P_1 \otimes \cdots \otimes P_r)\Big), \tag{4.36}$$

with $P := (P_1, ..., P_r)$ and

$$A := \sum_{l=1}^{L} \underbrace{x_l x_l^\top \otimes \cdots \otimes x_l x_l^\top}_{r \text{ times}}. \tag{4.37}$$

We mention that here we have used the same notation $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ to refer to the direct $r-$fold product of *real* Grassmannians.

4.2.4 A combinatorial problem

Let $\Lambda = (\lambda_{jk})_{j=1,k=1}^{n_2,n_1}$ be a given array of positive real numbers and let $m_1 \leq n_1,\ m_2 \leq n_2$ be fixed. Find m_1 columns and m_2 rows such that the sum of the corresponding entries λ_{jk} is maximal, i.e. solve the combinatorial maximization problem

$$\max_{\substack{J \subset \{1,\dots,n_2\} \\ |J|=m_2}} \max_{\substack{K \subset \{1,\dots,n_1\} \\ |K|=m_1}} \sum_{j \in J,\, k \in K} \lambda_{jk}. \tag{4.38}$$

We can permute m_1 columns and m_2 rows of Λ by right and left multiplication with permutations of the standard projectors Π_1 and Π_2, respectively. Hence, problem (4.38) is solved by finding permutation matrices σ_1 and σ_2 which maximize:

$$\sum_{i,j} (\Pi_{\sigma_2} \Lambda \Pi_{\sigma_1})_{ij}, \tag{4.39}$$

where $\sum_{i,j}$ is the sum over all entries and $\Pi_{\sigma_1} := \sigma_1^\top \Pi_1 \sigma_1,\ \Pi_{\sigma_2} := \sigma_2^\top \Pi_2 \sigma_2$. The sum in (4.39) can be written as

$$\sum_{i,j} (\Pi_{\sigma_2} \Lambda \Pi_{\sigma_1})_{ij} = \sum_{i,j} \Big((\Pi_{\sigma_1} \otimes \Pi_{\sigma_2})\operatorname{vec}(\Lambda)\Big)_{ij} = \operatorname{tr}\Big(A(\Pi_{\sigma_1} \otimes \Pi_{\sigma_2})\Big), \tag{4.40}$$

where $A := \operatorname{diag}(\operatorname{vec}(\Lambda))$. The last equality in (4.40) holds since $\Pi_{\sigma_1} \otimes \Pi_{\sigma_2}$ is diagonal, too. According to Corollary 4.1.10, we have the following equivalence

$$\max_{\sigma_1,\,\sigma_2} \operatorname{tr}\Big(A(\Pi_{\sigma_1} \otimes \Pi_{\sigma_2})\Big) \equiv \max_{(P_1,P_2)\in \operatorname{Gr}^{\times 2}(\mathbf{m},\mathbf{n})} \operatorname{tr}\Big(A(P_1 \otimes P_2)\Big). \tag{4.41}$$

Hence, we can embed the combinatorial maximization problem (4.38) into our continuous optimization task (4.5). The generalization of (4.38) to Λ being an arbitrary multi-array is straight-forward.

Problems of this type arise in multi-decision processes such as the following. Assume that a company has n_1 branches and each branch produces n_2 goods. If λ_{jk} denotes the gain of the $j-$th branch with the $k-$th good, then one could be interested to reduce the number of producers and goods to m_1 and m_2, respectively, which give maximum benefit.

4.3 Generic nondegeneracy of the critical points

One could be interested in knowing with which "certainty" the critical points of a real-valued function are nondegenerate. This type of question is common in differential topology and the basic tool is the Morse-Sard theorem, which is used to prove various transversality theorems, see e.g. [36]. In this subsection we derive a genericity statement concerning the critical points of the generalized Rayleigh-quotient as a consequence of the parametric transversality theorem [36]. We say that a property holds *generically* if it holds on a *residual set*, i.e. on a subset of a topological space that contains the intersection of a countable family of dense and open sets.

Let V, $\mathcal{M}$, $\mathcal{N}$ be finite dimensional smooth manifolds and $F : \mathrm{V} \times \mathcal{M} \to \mathcal{N}$ a smooth map. Moreover, let $\mathrm{T}_{(A,P)}F : V \times \mathrm{T}_P\mathcal{M} \to \mathrm{T}_{F(A,P)}\mathcal{N}$ denote the tangent map of F at $(A,P) \in V \times \mathcal{M}$. We say that F is *transversal* to a submanifold $S \subset \mathcal{N}$ and write $F \pitchfork S$ if

$$\operatorname{Im} \mathrm{T}_{(A,P)}F + \mathrm{T}_{F(A,P)}S = \mathrm{T}_{F(A,P)}\mathcal{N}, \tag{4.42}$$

for all $(A,P) \in F^{-1}(S)$. Then, the parametric transversality theorem states the following.

Theorem 4.3.1 *([36]) Let V, $\mathcal{M}$, $\mathcal{N}$ be smooth manifolds and S a submanifold of $\mathcal{N}$. Let $F : V \times \mathcal{M} \to \mathcal{N}$ be a smooth map, let $A \in V$ and define $F_A : \mathcal{M} \to \mathcal{N}$, $F_A(P) := F(A,P)$. If $F \pitchfork S$, then the set*

$$\{A \in V \mid F_A \pitchfork S\} \tag{4.43}$$

is residual. If moreover S is closed, then the set (4.43) *is open and dense.*

Now, let $f_A : \mathcal{M} \to \mathbb{R}$ be a smooth function depending on a parameter $A \in V$ and consider the map

$$F : \mathrm{V} \times \mathcal{M} \to \mathbb{R} \times \mathrm{T}^*\mathcal{M}, \quad F(A,P) = (f_A(P), \mathrm{d}f_A(P)), \tag{4.44}$$

where $\mathrm{T}^*\mathcal{M}$ is the cotangent bundle of $\mathcal{M}$ and $\mathrm{d}f_A(P)$ denotes the differential of f_A at $P \in \mathcal{M}$.
The relation between the nondegenerate critical points of f_A and the transversality of F is given in the next result.

Theorem 4.3.2 *Let M, V and F be as above and let $S := I \times \mathcal{M}_0$, where I is an open subset of $\mathbb{R}$ and $\mathcal{M}_0$ is the image of the zero section in $\mathrm{T}^*\mathcal{M}$. If $F \pitchfork S$, then the critical points $P \in \mathcal{M}$ of the smooth function $f_A : \mathcal{M} \to \mathbb{R}$ for which $f_A(P) \in I$, are generically nondegenerate.*

Proof. Fix $A \in V$ and define

$$F_A : \mathcal{M} \to \mathrm{T}^*\mathcal{M}, \quad F_A(P) := F(A, P). \tag{4.45}$$

From the Transversality Theorem 4.3.1 it follows that the set

$$R := \{A \in V \mid F_A \pitchfork S\}$$

is residual in V if $F \pitchfork S$. In the following, we will prove that $F_A \pitchfork S$ is equivalent to the fact that the Hessian of f_A is nondegenerate in the critical points $P \in \mathcal{M}$ for which $f_A(P) \in I$. This will prove the theorem.

First, notice that $P_c \in F_A^{-1}(S)$ if and only if $P_c \in \mathcal{M}$ is a critical point of f_A and $f_A(P_c) \in I$. Therefore, the transversality condition for F_A is

$$\operatorname{Im} \mathrm{T}_{P_c} F_A + \mathrm{T}_{F_A(P_c)} S = \mathrm{T}_{F_A(P_c)}(\mathbb{R} \times \mathrm{T}^*\mathcal{M}). \tag{4.46}$$

To rewrite (4.46) in local coordinates, we choose a coordinate chart φ on an open subset $U \subset \mathcal{M}$ around P_c with values in $\mathrm{T}_{P_c}\mathcal{M}$, such that $\varphi^{-1}(0) = P_c$ and $D\varphi^{-1}(0) = \mathrm{id}$. Then define

$$\widetilde{f}_A := f_A \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}.$$

Moreover, φ induces a chart $\psi : \mathbb{R} \times \pi^{-1}(U) \to \mathbb{R} \times \varphi(U) \times \mathrm{T}^*_{P_c}\mathcal{M} \subset \mathbb{R} \times \mathrm{T}_{P_c}\mathcal{M} \times \mathrm{T}^*_{P_c}\mathcal{M}$ around $F_A(P_c)$ via

$$\psi(\alpha, \gamma) = (\alpha, x, (D\varphi^{-1}(x))^*(\gamma)), \quad x := \varphi \circ \pi(\gamma),$$

Here, $\pi : \mathrm{T}^*\mathcal{M} \to \mathcal{M}$ refers to the natural projection and $(D\varphi^{-1}(x))^*(\gamma) := \gamma \circ D\varphi^{-1}(x)$. Thus, for

$$\widetilde{F}_A := \psi \circ F_A \circ \varphi^{-1} : \varphi(U) \to \mathbb{R} \times \varphi(U) \times \mathrm{T}^*_{P_c}\mathcal{M}$$

one has $\widetilde{F}_A(x) = (\widetilde{f}_A(x), x, d\widetilde{f}_A(x))$. Since transversality of F_A to S is preserved in local coordinates, (4.46) is equivalent to

$$\operatorname{Im} D\widetilde{F}_A(0) + \mathbb{R} \times \mathrm{T}_{P_c}\mathcal{M} \times \{0\} = \mathbb{R} \times \mathrm{T}_{P_c}\mathcal{M} \times \mathrm{T}^*_{P_c}\mathcal{M}. \tag{4.47}$$

Since $D\widetilde{F}_A(0) = (0, \mathrm{id}, d^2\widetilde{f}_A(0))$ yields that (4.47) is fulfilled if and only if $d^2\widetilde{f}_A(0)$ is nonsingular. Finally, the conclusion follows from the identity $\mathrm{Hess}_{f_A}(P_c) = d^2\widetilde{f}_A(0)$

which is satisfied due to the fact that P_c is a critical point and $D\varphi^{-1}(0) = \mathrm{id}$. Here, $\mathrm{Hess}_{f_A}(P_c)$ denotes the Hessian form corresponding to the Hessian operator via $\mathrm{Hess}_{f_A}(P_c)(x,y) = \langle \mathbf{H}_{f_A}(P_c)x, y\rangle$ for all $x, y \in \mathrm{T}_{P_c}\mathcal{M}$. □

An immediate consequence of the above theorem we obtain that if $S = \mathbb{R} \times \mathcal{M}_0$ and $F \pitchfork S$, then the set

$$\{A \in V \mid \text{all critical points of } f_A \text{ are nondegenerate}\}$$

is open and dense, which implies that the critical points of f_A are nondegenerate for a generic $A \in V$. In this situation, the validity of Theorem 4.3.2 does not change if instead of the map F defined in (4.44) we take the following

$$F : V \times \mathcal{M} \to \mathrm{T}^*\mathcal{M}, \quad F(A,P) := \mathrm{d}f_A(P).$$

However, if one wants to exclude certain critical points of f_A, as is e.g. the case in the problem of best low-rank tensor approximation, it is useful to have F defined by (4.44).

Now, the important conclusion about the critical points of the generalized Rayleigh-quotient is given.

Theorem 4.3.3 *The critical points of the generalized Rayleigh-quotient are generically nondegenerate, i.e. the set*

$$\{A \in \mathfrak{her}_N \mid \textit{all critical points of } \rho_A \textit{ are nondegenerate}\}$$

is open and dense.

Proof. Set $\mathcal{M} := \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$, $V := \mathfrak{her}_N$. For simplicity, we identify the cotangent bundle $\mathrm{T}^*\mathcal{M}$ with the tangent bundle $\mathrm{T}\mathcal{M}$ and work with the map

$$F : \mathrm{V} \times \mathcal{M} \to \mathbb{R} \times \mathrm{T}\mathcal{M}, \quad (A,P) \mapsto (\rho_A(P), \mathrm{grad}\rho_A(P)), \tag{4.48}$$

where $\mathrm{grad}\rho_A(P)$ is the Riemannian gradient of ρ_A at P. We show that $F \pitchfork S$, where $S := \mathbb{R} \times \mathcal{M}_0$ and $\mathcal{M}_0$ is now the image of the zero section in $\mathrm{T}\mathcal{M}$, i.e.

$$\mathrm{Im}\,\mathrm{T}_{(A,P)}F + \mathrm{T}_{F(A,P)}S = \mathrm{T}_{F(A,P)}(\mathbb{R} \times \mathrm{T}\mathcal{M}), \tag{4.49}$$

for all $(A,P) \in V \times \mathcal{M}$ with $\mathrm{grad}\rho_A(P) = 0$. As in the proof of Theorem 4.3.2, we rewrite the transversality condition (4.49) in local coordinates, i.e.

$$\mathrm{Im}\, D\widetilde{F}(A,0) + \mathbb{R} \times (\mathrm{T}_P\mathcal{M} \times \{0\}) = \mathbb{R} \times (\mathrm{T}_P\mathcal{M} \times \mathrm{T}_P\mathcal{M}), \tag{4.50}$$

where

$$\widetilde{F} := \psi \circ F \circ (\mathrm{id} \times \varphi^{-1}) : V \times W \to \mathbb{R} \times W \times \mathrm{T}_P\mathcal{M}.$$

Here, $\varphi : U \to W \subset \mathrm{T}_P\mathcal{M}$ is a chart around P with $\varphi^{-1}(0) = P$ and $D\varphi^{-1}(0) = \mathrm{id}$ and $\psi : \mathbb{R} \times \pi^{-1}(U) \to \mathbb{R} \times W \times \mathrm{T}_P\mathcal{M} \subset \mathbb{R} \times \mathrm{T}_P\mathcal{M} \times \mathrm{T}_P\mathcal{M}$ is the corresponding induced chart around $F(A,P)$. With this choice of charts, we obtain

$$\widetilde{F}(A,x) = (\widetilde{\rho}_A(x), x, \nabla\widetilde{\rho}_A(x)),$$

where $\widetilde{\rho}_A := \rho_A \circ \varphi^{-1} : W \to \mathbb{R}$. Since $A \mapsto \widetilde{\rho}_A(0)$ is linear and $\nabla\widetilde{\rho}_A(0) = \mathrm{grad}\rho_A(P) = 0$, one has

$$D\widetilde{F}(A,0)(X,\xi) = \Big(\widetilde{\rho}_X(0), \xi, \nabla\widetilde{\rho}_X(0) + d^2\widetilde{\rho}_A(0)\xi\Big).$$

Thus, condition (4.50) holds if and only if

$$\mathrm{Im}\,\nabla\widetilde{\rho}_{(\cdot)}(0) + \mathrm{Im}\,d^2\widetilde{\rho}_A(0) = \mathrm{T}_P\mathcal{M}. \tag{4.51}$$

Finally, we will show that $\mathrm{Im}\,\nabla\widetilde{\rho}_{(\cdot)}(0) = \mathrm{T}_P\mathcal{M}$ which clearly guarantees (4.51). Let $\xi := (\xi_1, \ldots, \xi_r) \in \big(\mathrm{Im}\,\nabla\widetilde{\rho}_{(\cdot)}(0)\big)^\perp$, then we obtain

$$\begin{aligned} 0 &= \langle \nabla\widetilde{\rho}_X(0), \xi\rangle = d\widetilde{\rho}_X(0)\xi = d\rho_X(P)\xi \\ &= \mathrm{tr}\Big(X(\sum_{j=1}^{r} P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r)\Big), \end{aligned}$$

for all $X \in \mathfrak{her}_N$. Notice, that the equality $d\widetilde{\rho}_X(0)\xi = d\rho_X(P)\xi$ follows from $D\varphi^{-1}(0) = \mathrm{id}$. Therefore,

$$\sum_{j=1}^{r} P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r = 0 \tag{4.52}$$

and this holds if and only if $\xi_1 = 0, \ldots, \xi_r = 0$, since according to Lemma 3.2.2, all summands in (4.52) are orthogonal to each other. We have proven that $\widetilde{F} \pitchfork \mathbb{R} \times \mathrm{T}_P\mathcal{M} \times \{0\}$ and hence $F \pitchfork S$. From the Theorem 4.3.2 it follows immediately that the critical points of the generalized Rayleigh-quotient are generically nondegenerate. □

From the proof of the above corollary, we extract the following equivalent formulation for the transversality condition (4.49):
Let $F : \mathfrak{her}_N \times \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n}) \to \mathrm{TGr}^\times(\mathbf{m}, \mathbf{n})$ be the map defined by $(A, P) \mapsto \mathrm{grad}\ \rho_A(P)$, and for fix $P \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ define

$$F_P : \mathfrak{her}_N \to \mathrm{TGr}^\times(\mathbf{m}, \mathbf{n}), \quad A \mapsto \mathrm{grad}\ \rho_A(P).$$

Then, $F \pitchfork S$ if and only if

$$\mathrm{Im}\,\mathrm{T}_A F_P + \mathrm{Im}\,\mathbf{H}_{\rho_A}(P) = \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m}, \mathbf{n}), \tag{4.53}$$

for all $(A, P) \in \mathfrak{her}_N \times \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ with $\mathrm{grad}\ \rho_A(P) = 0$. As before S is the zero section in the tangent bundle $\mathrm{TGr}^\times(\mathbf{m}, \mathbf{n})$.

4.4 Critical point discussion in the case of the best low-rank tensor approximation problem

In the previous section we have proved that the critical points of the generalized Rayleigh-quotient ρ_A are nondegenerate for A in an open and dense subset of $\mathfrak{her}_N$.

A key point was the fact that the generalized Rayleigh-quotient depends linearly on the parameter $A \in \mathfrak{her}_N$. However, for the problem of best approximation of a tensor with a tensor of lower rank, A is a positive semidefinite matrix of rank-one, i.e. $A = xx^\dagger$, $x \in \mathbb{C}^N$, and hence no longer from a vector space. With x as the new parameter, the generalized Rayleigh-quotient is no longer linear in the parameter, but quadratic. In what follows, we analyze the general situation when the generalized Rayleigh-quotient depends quadratically on the parameter, particularly

$$\rho_{XX^\dagger}(P_1,\dots,P_r) := \operatorname{tr}\Big(XX^\dagger(P_1\otimes\cdots\otimes P_r)\Big), \tag{4.54}$$

where $X \in \mathbb{C}^{N\times K}$, $K \leq N$ and $N := n_1 n_2 \cdots n_r$.

Let $\mathrm{St}_{K,N} := \{X \in \mathbb{C}^{N\times K} \mid \operatorname{rank} X = K, K \leq N\}$ denote the noncompact Stiefel manifold, i.e. the set of all full rank matrices $X \in \mathbb{C}^{N\times K}$, with $K \leq N$, and define the map

$$F : \mathrm{St}_{K,N} \times \mathrm{Gr}^\times(\mathbf{m},\mathbf{n}) \to \mathrm{T}\mathrm{Gr}^\times(\mathbf{m},\mathbf{n}), \quad (X,P) \mapsto \operatorname{grad} \rho_{XX^\dagger}(P). \tag{4.55}$$

By fixing $P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ in F, we obtain the following map

$$F_P : \mathrm{St}_{K,N} \to \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n}),\ X \mapsto \operatorname{grad} \rho_{XX^\dagger}(P), \tag{4.56}$$

with the following tangent map at $X \in \mathrm{St}_{K,N}$

$$\mathrm{T}_X F_P : \mathbb{C}^{K\times N} \to \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n}),\ \mathrm{T}_X F_P(Y) = \operatorname{grad} \rho_{XY^\dagger + YX^\dagger}(P).$$

To avoid confusion, we accentuate that $\operatorname{grad} \rho_{XX^\dagger}(P)$ and $\mathbf{H}_{\rho_{XX^\dagger}}(P)$ denote the Riemannian gradient and the Hessian of $\rho_{XX^\dagger}$ at $P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$.

Recall the equivalent transversality condition (4.53) for the generalized Rayleigh-quotient ρ_A, $A \in \mathfrak{her}_N$. We prove in the sequel that, $F \pitchfork S$ if and only if $\operatorname{Im} \mathrm{T}_X F_P = \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$, for all $(X,P) \in F^{-1}(S)$, with S the zero section in the tangent bundle of $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$. We denote with $(\operatorname{Im} \mathrm{T}_X F_P)^\perp$ the orthogonal complement of $\operatorname{Im} \mathrm{T}_X F_P$ in $\mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$. For simplicity, we give first a characterization of the elements in $(\operatorname{Im} \mathrm{T}_X F_P)^\perp$.

Lemma 4.4.1 *Let $X \in \mathrm{St}_{K,N}$, $P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ and let F_P be the map defined by* (4.56). *Then $0 \neq \xi = (\xi_1,\dots,\xi_r) \in (\operatorname{Im} \mathrm{T}_X F_P)^\perp$ if and only if*

$$(\sum_{j=1}^r P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r)X = 0. \tag{4.57}$$

Proof. On the vector space $\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$ we consider the inner product $\langle \cdot,\cdot \rangle$ defined by (3.19). Let $A \in \mathfrak{her}_N$ and denote with $\widetilde{\rho}_A$ a smooth extension of ρ_A to $\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$. For $P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ the following hold

$$\begin{aligned} D\widetilde{\rho}_A(P)(\eta) &= \operatorname{tr}\Big(A(\sum_{j=1}^r P_1\otimes\cdots\otimes Z_j \otimes\cdots\otimes P_r)\Big) \\ &= \langle \nabla\widetilde{\rho}_A(P), \eta\rangle = \langle \operatorname{grad} \rho_A(P), \eta\rangle, \end{aligned} \tag{4.58}$$

for all $\eta \in \mathrm{T}_P\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$. Here, $\nabla\widetilde{\rho}_A(P)$ is the gradient of $\widetilde{\rho}_A$ at P and grad $\rho_A(P)$ is the Riemannian gradient of ρ_A at P.

Let $X \in \mathrm{St}_{K,N},\ P \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ and $\xi := (\xi_1,\dots,\xi_r) \in \mathrm{T}_P(\mathrm{Im}\,\mathrm{T}_X F_P)^{\perp}$. From the definition of the orthogonal complement in $\mathrm{T}_P\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ we have that $\xi := (\xi_1,\dots,\xi_r) \in \mathrm{T}_P(\mathrm{Im}\,\mathrm{T}_X F_P)^{\perp}$ if and only if

$$\langle \xi, \mathrm{grad}\ \rho_{XY^\dagger+YX^\dagger}(P)\rangle = 0, \tag{4.59}$$

for all $Y \in \mathrm{T}_X\mathrm{St}_{K,N} = \mathbb{C}^{N\times K}$. From (4.58) and (4.59) we have that $\xi \in \mathrm{T}_P(\mathrm{Im}\,\mathrm{T}_X F_P)^{\perp}$ if and only if

$$\mathrm{tr}\Big((XY^\dagger + YX^\dagger)(\sum_{j=1}^{r} P_1\otimes\cdots\otimes\xi_j\otimes\cdots\otimes P_r)\Big) = \langle \xi, \mathrm{grad}\ \rho_{XY^\dagger+YX^\dagger}(P)\rangle = 0,$$

for all $Y \in \mathbb{C}^{N\times K}$, which is equivalent to

$$(\sum_{j=1}^{r} P_1\otimes\cdots\otimes\xi_j\otimes\cdots\otimes P_r)X = 0.$$

□

Theorem 4.4.2 *Let $X \in \mathrm{St}_{K,N}$, $P \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ and let F_P be the map defined by* (4.56). *Then,*

$$(\mathrm{Im}\,\mathrm{T}_X F_P)^{\perp} \subset \ker\ \mathbf{H}_{\rho_{XX^\dagger}}(P).$$

Proof. Let $X \in \mathrm{St}_{K,N},\ P \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ and $\xi := (\xi_1,\dots,\xi_r) \in \mathrm{T}_P(\mathrm{Im}\,\mathrm{T}_X F_P)^{\perp}$. To prove that $\xi \in \ker\ \mathbf{H}_{\rho_{XX^\dagger}}(P)$, it is necessary and sufficient to show that

$$\langle \mathbf{H}_{\rho_{XX^\dagger}}(P)\xi, \eta\rangle = 0,$$

for all $\eta \in \mathrm{T}_P\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$. Explicitly, it is left to show that

$$\begin{aligned}&\sum_{j=1}^{r}\mathrm{tr}\left(XX^\dagger(P_1\otimes\cdots\otimes([\xi_j,P_j]\eta_j + [\eta_j,P_j]\xi_j)\otimes\cdots\otimes P_r\right)\\ &+\sum_{j=1}^{r}\sum_{k=1,\,k\neq j}^{r}\mathrm{tr}\left(XX^\dagger(P_1\otimes\cdots\otimes\eta_j\otimes\cdots\otimes\xi_k\otimes\cdots\otimes P_r)\right) \quad = 0,\end{aligned} \tag{4.60}$$

for all $\eta \in \mathrm{T}_P\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$. With the following notations

$$\begin{aligned}A_j &:= \mathrm{tr}\left(XX^\dagger(P_1\otimes\cdots\otimes[\xi_j,P_j]\eta_j\otimes\cdots\otimes P_r)\right),\\ B_j &:= \mathrm{tr}\left(XX^\dagger(P_1\otimes\cdots\otimes[\eta_j,P_j]\xi_j\otimes\cdots\otimes P_r)\right)\end{aligned}$$

equation (4.60) becomes

$$\sum_{j=1}^{r} A_j + B_j = -\sum_{j=1}^{r} \sum_{k=1,k\neq j}^{r} \operatorname{tr}\Big(XX^\dagger(P_1 \otimes \cdots \otimes \eta_j \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r)\Big), \qquad (4.61)$$

for all $\eta \in \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$. By multiplying (4.57) from the left with $(I_{n_1} \otimes \cdots \otimes P_j \otimes \cdots \otimes I_{n_r})$ and substracting the new equation from (4.57), we obtain that

$$\Big(P_1 \otimes \cdots \otimes (\xi_j - P_j\xi_j) \otimes \cdots \otimes P_r\Big)X = 0, \qquad (4.62)$$

for all $j = 1,\ldots,r$. Moreover, since $\xi_j = P_j\xi_j + \xi_j P_j$, from equality (4.62) we obtain

$$\Big(P_1 \otimes \cdots \otimes \xi_j P_j \otimes \cdots \otimes P_r\Big)X = 0, \qquad (4.63)$$

for all $j = 1,\ldots,r$. Furthermore, from (4.57) it follows that

$$XX^\dagger\Big(\sum_{k=1}^{r} P_1 \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r\Big)\Big(I_{n_1} \otimes \cdots \otimes P_j\eta_j \otimes \cdots \otimes I_{n_r}\Big) \quad = 0, \qquad (4.64)$$

for all $j = 1,\ldots,r$. Hence,

$$\begin{aligned} A_j \quad &= \operatorname{tr}\Big(XX^\dagger(P_1 \otimes \cdots \otimes \xi_j P_j \eta_j \otimes \cdots \otimes P_r)\Big) \\ &= \operatorname{tr}\Bigg(XX^\dagger\Big(\sum_{k=1}^{r} P_1 \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r\Big)\Big(I_{n_1} \otimes \cdots \otimes P_j\eta_j \otimes \cdots \otimes I_{n_r}\Big)\Bigg) \\ &- \sum_{k=1,k\neq j}^{r} \operatorname{tr}\Big(XX^\dagger(P_1 \otimes \cdots \otimes P_j\eta_j \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r)\Big) \\ &= -\sum_{k=1,k\neq j}^{r} \operatorname{tr}\Big(XX^\dagger(P_1 \otimes \cdots \otimes P_j\eta_j \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r)\Big), \end{aligned}$$

for all $j = 1,\ldots,r$. Similar to (4.64) we obtain that

$$\Big(I_{n_1} \otimes \cdots \otimes [\eta_j, P_j] \otimes \cdots \otimes I_{n_r}\Big)\Big(\sum_{k=1}^{r} P_1 \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r\Big)XX^\dagger \quad = 0, \qquad (4.65)$$

and hence,

$$B_j \quad = -\sum_{k=1,k\neq j}^{r} \operatorname{tr}\Big(XX^\dagger(P_1 \otimes \cdots \otimes [\eta_j, P_j]P_j \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r)\Big),$$

for all $j = 1,\ldots,r$. Moreover, since $P_j\eta_j P_j = 0$ for all $\eta_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$, it follows that

$$B_j \quad = -\sum_{k=1,k\neq j}^{r} \operatorname{tr}\Big(XX^\dagger(P_1 \otimes \cdots \otimes \eta_j, P_j \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r)\Big),$$

for all $j = 1, \ldots, r$. Recalling that $\eta_j = P_j\eta_j + \eta_j P_j$ for every $\eta_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$, the conclusion follows

$$\sum_{j=1}^{r} A_j + B_j = -\sum_{j=1} r \sum_{k=1,k\neq j}^{r} \mathrm{tr}\left(XX^\dagger(P_1 \otimes \cdots \otimes \eta_j \otimes \cdots \otimes \xi_k \otimes \cdots \otimes P_r)\right).$$

Thus $\langle \mathbf{H}_{\rho_{XX^\dagger}}(P)\xi, \eta\rangle = 0$ for all $\eta \in \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ and hence, $\xi \in \ker\mathbf{H}_{\rho_{XX^\dagger}}(P)$. $\square$

As an immediate consequence of the above theorem we have the following result about the critical points of $\rho_{XX^\dagger}$, $X \in \mathrm{St}_{K,N}$.

Corollary 4.4.3 *For $X \in \mathrm{St}_{K,N}$, all critical points $P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ of $\rho_{XX^\dagger}$ for which $(\mathrm{Im}\,\mathrm{T}_X F_P)^\perp \neq \{0\}$ are degenerate.*

Another consequence of Theorem 4.4.2 is a necessary and sufficient condition for the generic nondegeneracy of the critical points of the generalized Rayleigh-quotient (4.54).

Corollary 4.4.4 *Let F be the map defined by* (4.55) *and S the zero section in* $\mathrm{TGr}^\times(\mathbf{m},\mathbf{n})$. *Then, $F \pitchfork S$ if and only if*

$$\mathrm{Im}\,\mathrm{T}_X F_P = \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n}), \tag{4.66}$$

for all $(X,P) \in F^{-1}(S)$, with F_P defined by (4.56)*. Moreover, from Theorem 4.3.1 it follows that if $F \pitchfork S$, then the set*

$$\{X \in \mathrm{St}_{K,N} \mid \textit{all critical points of } \rho_{XX^\dagger} \textit{ are nondegenerate}\}$$

is open and dense.

Depending on the rank of the parameter X of $\rho_{XX^\dagger}$, we prove next that the critical points of the generalized Rayleigh-quotient are generically nondegenerate.

Theorem 4.4.5 *Let F be the map defined by* (4.55) *and S be the zero section in the tangent bundle of* $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$*. If*

$$K > N - 2\min\{\prod_{k=1,\, k\neq j}^{r} m_k \mid j = 1, \ldots, r\}, \tag{4.67}$$

then F is transversal to S. Moreover, the set

$$\{X \in \mathrm{St}_{K,N} \mid \textit{all critical points } P \in \mathrm{Gr}^\times(\mathbf{m},\mathbf{n}) \textit{ of } \rho_{XX^\dagger} \textit{ are nondegenerate}\}$$

is open and dense.

Proof. We will prove that when condition (4.67) holds, then $(\operatorname{Im} \mathrm{T}_X F_P)^\perp = \{0\}$ holds for all $(P, X) \in F^{-1}(S)$. Hence, the conclusion follows from Corollary 4.4.4 and Theorem 4.3.1.

Without loss of generality we can assume that $P = (\Pi_1, \ldots, \Pi_r)$, where Π_j are the standard projectors of $\mathbb{C}^{n_j}$, $j = 1, \ldots, r$. Let $X \in \mathrm{St}_{K,N}$ such that $(\Pi, X) \in F^{-1}(S)$. Assume that there exist $0 \neq \xi \in (\operatorname{Im} \mathrm{T}_X F_P)^\perp$. From Lemma 4.4.1 we have

$$(\sum_{j=1}^{r} \Pi_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes \Pi_r)X = 0. \tag{4.68}$$

Next, we show that

$$\dim \ \ker \left(\sum_{j=1}^{r} \Pi_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes \Pi_r \right) \leq N - 2 \min\{ \prod_{k=1,\ k \neq j}^{r} m_k \mid j = 1, \ldots, r\}. \tag{4.69}$$

Let $0 \neq \eta \in \mathrm{T}_P \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ and denote with $W := \sum_{j=1}^{r} \Pi_1 \otimes \cdots \otimes \eta_j \otimes \cdots \otimes \Pi_r$, then

$$\begin{aligned} W &= \Pi_1 \otimes \left(\sum_{k=2}^{r} \Pi_2 \otimes \cdots \otimes \eta_k \otimes \cdots \otimes \Pi_r \right) + \eta_1 \otimes \Pi_2 \otimes \cdots \otimes \Pi_r \\ &= \begin{bmatrix} I_{m_1} \otimes \left(\sum_{k=2}^{r} \Pi_2 \otimes \cdots \otimes \eta_k \otimes \cdots \otimes \Pi_r \right) & Z_1 \otimes \Pi_2 \otimes \cdots \otimes \Pi_r \\ Z_1^\dagger \otimes \Pi_2 \otimes \cdots \otimes \Pi_r & 0 \end{bmatrix}, \end{aligned}$$

where

$$\eta_1 = \begin{bmatrix} 0 & Z_1 \\ Z_1^\dagger & 0 \end{bmatrix} \in \mathrm{T}_{\Pi_1} \mathrm{Gr}_{m_1, n_1},$$

and $Z_1 \in \mathbb{C}^{m_1 \times (n_1 - m_1)}$. Thus,

$$\operatorname{rank} W \geq \operatorname{rank} \eta_1 \otimes \Pi_2 \otimes \cdots \otimes \Pi_r.$$

Similar, we obtain that

$$\operatorname{rank} W \geq \operatorname{rank} \Pi_1 \otimes \cdots \otimes \eta_j \otimes \cdots \otimes \Pi_r, \tag{4.70}$$

for all $j = 1, \ldots, r$.

By the Rank-Nullity Theorem we have

$$\dim \ \ker W = N - \operatorname{rank} W$$

and by (4.70) it follows that

$$N - \operatorname{rank} W \leq N - \max\{\operatorname{rank} \Pi_1 \otimes \cdots \otimes \eta_j \otimes \cdots \otimes \Pi_r \mid j = 1, \ldots, r\}.$$

Moreover, since $\eta \neq 0$ there exists j between 1 and r such that rank $\eta_j \geq 2$. Thus, we obtain that

$$N - \text{rank } W \leq N - 2\min\{\prod_{k=1,\, k\neq j}^{r} m_k \mid j = 1, \ldots, r\}, \tag{4.71}$$

which proves (4.69), which implies that $(\operatorname{Im} \mathrm{T}_X F_P)^{\perp} = \{0\}$. □

In view of applications, one is in particular interested in what can be said about the nondegeneracy of the critical points of the generalized Rayleigh-quotient in the case when $K = 1$, i.e. $\rho_{xx^\dagger}$ with $x \in \mathbb{C}^N$. Recall from Section 4.2.1 that when $K = 1$, the maximization of the generalized Rayleigh-quotient $\rho_{xx^\dagger}$ is equivalent to finding the best approximation of a tensor $\mathcal{T} \in \mathbb{C}^{n_1 \times \cdots \times n_r}$ with a tensor of lower rank (4.28). The tensor $\mathcal{T}$ and the vector x are related by the fact that x is a lexicographical representation of $\mathcal{T}$. At this point we recall some of the important notions when dealing with tensors, and refer to Section 4.2.1 and the literature therein for details. If $\mathcal{T}$ is a tensor in $\mathbb{C}^{n_1 \times \cdots \times n_r}$, then $T_{(j)} \in \mathbb{C}^{n_j \times N_j}$ with $N_j = N/n_j$ denotes its matrix unfolding along the direction j, for $j = 1, \ldots, r$. The rank of $\mathcal{T}$ is rank $\mathcal{T} = (\text{rank } T_{(1)}, \ldots, \text{rank } T_{(r)})$. Moreover, we say that a tensor $\mathcal{T}$ has *full rank* if the unfoldings $T_{(1)}, \ldots, T_{(r)}$ have full rank.

It is known that for a tensor of order 2 (matrix) the column rank and the row rank are equal, hence a best approximation with a tensor of rank $\leq (m_1, m_2)$ is equivalent to a best approximation with a tensor of rank (m, m), where $m = \min\{m_1, m_2\}$. For higher-order tensors there is no similar statement known. In the following lemma, we give a relation between the values $m_1, \ldots, m_r$ from the best $(m_1, \ldots, m_r)$-rank approximation of a tensor, which is in a certain sense a generalization of the statement about the best rank-(m_1, m_2) approximation of a matrix.

Lemma 4.4.6 *Let $m_1 \leq n_1/2, \ldots, m_r \leq n_r/2$ be given natural numbers and let $N := n_1 n_2 \cdots n_r$. If there exists a permutation $\beta \in \mathrm{S}_r$ such that*

$$m_{\beta(1)} > m_{\beta(2)} m_{\beta(3)} \cdots m_{\beta(r)}, \tag{4.72}$$

then, for any $x \in \mathbb{C}^N$, the local extrema of the generalized Rayleigh-quotient $\rho_{xx^\dagger}$ are degenerate.

Proof. Without loss of generality we assume that $\beta = \mathrm{id}$, thus

$$m_1 > m_2 m_3 \cdots m_r.$$

The proof is given for $r = 3$, as the generalization is straight-forward. Let $x \in \mathbb{C}^N$ with $N := n_1 n_2 n_3$ and let $(P_1, P_2, P_3) \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ be a critical point of $\rho_{xx^\dagger}$. Then

$$\operatorname{tr}\Big(xx^\dagger(\xi_1 \otimes P_2 \otimes P_3)\Big) = 0 \tag{4.73}$$

for all $\xi_1 \in \mathrm{T}_{P_1}\mathrm{Gr}_{m_1,n_1}$. Let $X \in \mathbb{C}^{n_1 \times n_2 n_3}$ be a matrix such that $x = \mathrm{vec}(X)$. With the Kronecker-vec formalism we can write

$$(\xi_1 \otimes P_2 \otimes P_3)x = \mathrm{vec}\Big(\xi_1 X(P_2 \otimes P_3)\Big)$$

and hence (4.73) becomes

$$\mathrm{tr}\Big(X(P_2 \otimes P_3)X^\dagger \xi_1\Big) = 0,$$

for all $\xi_1 \in \mathrm{T}_{P_1}\mathrm{Gr}_{m_1,n_1}$. Here we have used the following property of the trace function and vec operator

$$\mathrm{tr}(A^\dagger B) = \mathrm{vec}(A)^\dagger \mathrm{vec}(B),$$

for any $A, B \in \mathbb{C}^{n \times n}$.

Let $\Theta_1 \in \mathrm{SU}(n_1)$ such that $P_1 = \Theta_1^\dagger \Pi_1 \Theta_1$. Then, from Corollary 4.1.8 it follows that

$$\Theta_1\Big(X(P_2 \otimes P_3)X^\dagger\Big)\Theta_1^\dagger = \begin{bmatrix} \Psi_1' & 0 \\ 0 & \Psi_1'' \end{bmatrix}, \tag{4.74}$$

where $\Psi_1' \in \mathfrak{her}_{m_1},\ \Psi_1'' \in \mathfrak{her}_{n_1-m_1}$. Since,

$$\mathrm{rank}\,\Big(X(P_2 \otimes P_3)X^\dagger\Big) \leq m_2 m_3 < m_1$$

it follows that rank $\Psi_1' < m_1$ and rank $\Psi_1'' < m_1$ and hence

$$\{0\} \subset \sigma(\Psi_1') \cap \sigma(\Psi_1'') \neq \emptyset,$$

where $\sigma(\Psi_1'),\ \sigma(\Psi_1'')$ denote the spectrum of Ψ_1' and Ψ_1'' respectively. From Theorem 4.1.12 it follows that the Hessian $\mathbf{H}_{\rho_{xx^\dagger}}(P)$ is degenerate. □

From the above result it can be easily noticed that for the best rank-(m_1, m_2) approximation of a second order tensor, the local extrema of the generalized Rayleigh-quotient are degenerate if $m_1 \neq m_2$, i.e. $m_1 > m_2$ or $m_2 > m_1$. In this sense, the relation (4.72) is a generalization of the fact that it is required the best rank-(m_1, m_2) of a matrix with $m_1 = m_2$.

We enclose the analysis of the problem of best approximating a tensor by a tensor of lower rank, with some genericity results on the critical points. We provide sufficient conditions which guarantee that the critical points of the generalized Rayleigh-quotient $\rho_{xx^\dagger}$ that satisfy a certain condition are nondegenerate for x from a residual set. As before, x is a vector representation of a tensor $\mathcal{T}$. In particular, we obtain that the global maximizers of $\rho_{xx^\dagger}$ are nondegenerate for a generic choice of x.

We know from Corollary 4.4.4 that the map F defined by (4.55) is transversal to the zero section in $\mathrm{TGr}^\times(\mathbf{m}, \mathbf{n})$ if and only if $\mathrm{Im}\,\mathrm{T}_x F_P = \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ or equivalent

$(\operatorname{Im} \mathrm{T}_x F_P)^\perp = \{0\}$, for all $(x, P) \in F^{-1}(S)$ (S is the zero section in $\mathrm{TGr}^\times(\mathbf{m}, \mathbf{n})$). From Lemma 4.4.1 we have that $\xi \in (\operatorname{Im} \mathrm{T}_x F_P)^\perp$ if and only if

$$(\sum_{j=1}^{r} P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r)x = 0. \tag{4.75}$$

has only the trivial solution $\xi = 0$.

In what follows, we want to find an equivalent formulation for (4.75). To simplify the exposure, we will present in detail the linear system of equations (4.75) in the case when $r = 3$, since the general case follows straight-forwardly. Choose unitary matrices $\Theta_j \in \mathrm{SU}(n_j)$ such that P_j and ξ_j have the standard representation

$$P_j = \Theta_j^\dagger \Pi_j \Theta_j \quad \text{and} \quad \xi_j = \Theta_j^\dagger \begin{bmatrix} 0 & Z_j \\ Z_j^\dagger & 0 \end{bmatrix} \Theta_j, \tag{4.76}$$

respectively. Then, rewrite the system of equations (4.75) with $(Z_1, Z_2, Z_3) \in \mathbb{C}^{m_1\times(n_1-m_1)} \times \mathbb{C}^{m_2\times(n_2-m_2)} \times \mathbb{C}^{m_3\times(n_3-m_3)}$ as variables

$$\Big(\zeta_1 \otimes \Pi_2 \otimes \Pi_3 + \Pi_1 \otimes \zeta_2 \otimes \Pi_3 + \Pi_1 \otimes \Pi_2 \otimes \zeta_3\Big)\widetilde{x} = 0. \tag{4.77}$$

Here, $\zeta_j = \begin{bmatrix} 0 & Z_j \\ Z_j^\dagger & 0 \end{bmatrix}$ and $\widetilde{x} = (\Theta_1 \otimes \Theta_2 \otimes \Theta_3)x$. If $\mathcal{T} \in \mathbb{C}^{n_1\times n_2\times n_3}$ is the tensor corresponding to $\widetilde{x}$ ($\widetilde{x} = \operatorname{vec}(\mathcal{T})$), then (4.77) has the equivalent formulation in terms of tensors

$$\mathcal{T} \times_1 \zeta_1 \times_2 \Pi_2 \times_3 \Pi_3 + \mathcal{T} \times_1 \Pi_1 \times_2 \zeta_2 \times_3 \Pi_3 + \mathcal{T} \times_1 \Pi_1 \times_2 \Pi_2 \times_3 \zeta_3 = 0. \tag{4.78}$$

Denote

$$\begin{aligned}
\mathcal{A} &:= \mathcal{T} \times_1 Q_1 \times_2 Q_2 \times_3 Q_3 \in \mathbb{C}^{m_1\times m_2\times m_3}, \\
\mathcal{B} &:= \mathcal{T} \times_1 Q_1^\perp \times_2 Q_2 \times_3 Q_3 \in \mathbb{C}^{(n_1-m_1)\times m_2\times m_3}, \\
\mathcal{C} &:= \mathcal{T} \times_1 Q_1 \times_2 Q_2^\perp \times_3 Q_3 \in \mathbb{C}^{m_1\times(n_2-m_2)\times m_3}, \\
\mathcal{D} &:= \mathcal{T} \times_1 Q_1 \times_2 Q_2 \times_3 Q_3^\perp \in \mathbb{C}^{m_1\times m_2\times(n_3-m_3)},
\end{aligned} \tag{4.79}$$

where

$$Q_j = \begin{bmatrix} I_{m_j} \\ 0 \end{bmatrix} \in \mathbb{C}^{n_j\times m_j} \quad \text{and} \quad Q_j^\perp = \begin{bmatrix} 0 \\ I_{(n_j-m_j)} \end{bmatrix} \in \mathbb{C}^{n_j\times(n_j-m_j)},$$

for $j = 1, 2, 3$. Recalling that $T_{(j)}$ denotes the unfolding of $\mathcal{T}$ along direction j and that

$$(\mathcal{T} \times_j V)_{(j)} = VT_{(j)},$$

for all $V \in \mathbb{C}^{k\times n_j},\ j = 1, 2, 3$, the equation (4.78) can be written in matrix form along direction 1 as follows

$$\zeta_1(\mathcal{T} \times_2 \Pi_2 \times_3 \Pi_3)_{(1)} + \Pi_1(\mathcal{T} \times_2 \zeta_2 \times_3 \Pi_3)_{(1)} + \Pi_1(\mathcal{T} \times_2 \Pi_2 \times_3 \zeta_3)_{(1)} = 0.$$

This is equivalent to

$$[\,0 \quad Z_1\,]\,(\mathcal{T} \times_2 \Pi_2 \times_3 \Pi_3)_{(1)} + Q_1^\dagger(\mathcal{T} \times_2 \zeta_2 \times_3 \Pi_3)_{(1)} + Q_1^\dagger(\mathcal{T} \times_2 \Pi_2 \times_3 \zeta_3)_{(1)} = 0,$$

$$\begin{bmatrix} Z_1^\dagger & 0 \end{bmatrix}(\mathcal{T} \times_2 \Pi_2 \times_3 \Pi_3)_{(1)} = 0$$

and in tensor formalism to

$$\mathcal{T} \times_1 (Q_1^\perp)^\dagger Z_1 \times_2 \Pi_2 \times_3 \Pi_3 + \mathcal{T} \times_1 Q_1 \times_2 \zeta_2 \times_3 \Pi_3 + \mathcal{T} \times_1 Q_1 \times_2 \Pi_2 \times_3 \zeta_3 = 0, \tag{4.80}$$

$$(\mathcal{T} \times_1 Q_1 \times_2 \Pi_2 \times_3 \Pi_3) \times_1 Z_1^\dagger = 0. \tag{4.81}$$

Using the same techinque for direction 2 on (4.80), we obtain

$$\mathcal{T} \times_1 (Q_1^\perp)^\dagger Z_1 \times_2 Q_2 \times_3 \Pi_3 + \mathcal{T} \times_1 Q_1 \times_2 (Q_2^\perp)^\dagger Z_2 \times_3 \Pi_3$$

$$+ \mathcal{T} \times_1 Q_1 \times_2 Q_2 \times_3 \zeta_3 = 0, \tag{4.82}$$

$$(\mathcal{T} \times_1 Q_1 \times_2 Q_2 \times_3 \Pi_3) \times_2 Z_2^\dagger = 0. \tag{4.83}$$

One more time for the direction 3 and (4.82) and it follows

$$\mathcal{T} \times_1 (Q_1^\perp)^\dagger Z_1 \times_2 Q_2 \times_3 Q_3 + \mathcal{T} \times_1 Q_1 \times_2 (Q_2^\perp)^\dagger Z_2 \times_3 Q_3$$

$$+\mathcal{T} \times_1 Q_1 \times_2 Q_2 \times_3 (Q_3^\perp)^\dagger Z_3 = 0,$$

$$(\mathcal{T} \times_1 Q_1 \times_2 Q_2 \times_3 Q_3) \times_3 Z_3^\dagger = 0.$$

With the notation given by (4.79), the solutions (Z_1, Z_2, Z_3) of (4.77) are solutions of the system

$$\mathcal{A} \times_1 Z_1^\dagger = 0 \tag{4.84}$$

$$\mathcal{A} \times_2 Z_2^\dagger = 0 \tag{4.85}$$

$$\mathcal{A} \times_3 Z_3^\dagger = 0 \tag{4.86}$$

$$\mathcal{B} \times_1 Z_1 + \mathcal{C} \times_2 Z_2 + \mathcal{D} \times_3 Z_3 = 0 \tag{4.87}$$

and the other way around. Since the first three equations above have the matrix form

$$Z_1^\dagger A_{(1)} = 0 \tag{4.88}$$

$$Z_2^\dagger A_{(2)} = 0 \tag{4.89}$$

$$Z_3^\dagger A_{(3)} = 0, \tag{4.90}$$

it follows that if $\mathcal{A}$ has full rank, then the system (4.77) has only the trivial solution. Recall the multilinear map $\Psi_{\widetilde{xx^\dagger},j}$ defined by (4.6) and denote

$$\widehat{X}_1 := \Psi_{\widetilde{xx^\dagger},1}(I_{n_1}, \Pi_2, \Pi_3) = T_{(1)}(\Pi_2 \otimes \Pi_3)T_{(1)}^\dagger, \tag{4.91}$$

$$\widehat{X}_2 := \Psi_{\widetilde{xx^\dagger},2}(\Pi_1, I_{n_2}, \Pi_3) = T_{(2)}(\Pi_1 \otimes \Pi_3)T_{(2)}^\dagger, \tag{4.92}$$

$$\widehat{X}_3 := \Psi_{\widetilde{xx^\dagger},3}(\Pi_1, \Pi_2, I_{n_3}) = T_{(3)}(\Pi_1 \otimes \Pi_2)T_{(3)}^\dagger, \tag{4.93}$$

where $T_{(j)}$ is the matrix unfolding of $\mathcal{T}$ along direction j. From the notation (4.79) we obtain that

$$\widehat{X}_1 = \begin{bmatrix} A_{(1)} \\ B_{(1)} \end{bmatrix} \begin{bmatrix} A_{(1)}^\dagger & B_{(1)}^\dagger \end{bmatrix}, \tag{4.94}$$

$$\widehat{X}_2 = \begin{bmatrix} A_{(2)} \\ C_{(2)} \end{bmatrix} \begin{bmatrix} A_{(2)}^\dagger & C_{(2)}^\dagger \end{bmatrix}^\dagger, \tag{4.95}$$

$$\widehat{X}_3 = \begin{bmatrix} A_{(3)} \\ D_{(3)} \end{bmatrix} \begin{bmatrix} A_{(3)}^\dagger & D_{(3)}^\dagger \end{bmatrix}^\dagger, \tag{4.96}$$

where $B_{(1)}$, $C_{(2)}$ and $D_{(3)}$ are the matrix unfoldings of $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ along direction 1, 2 and 3 respectively. With the above specifications, the critical point condition (4.1.8) ca be equivalently given as follows.

Lemma 4.4.7 *Let $x \in \mathbb{C}^N$, $P \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ and $A_{(1)}, A_{(2)}, A_{(3)}, B_{(1)}, C_{(2)}, D_{(3)}$ as defined before. If P is a critical point of $\rho_{xx^\dagger}$, then the critical point condition* (4.1.8) *is equivalent to*

$$B_{(1)}A_{(1)}^\dagger = 0, \quad C_{(2)}A_{(2)}^\dagger = 0, \quad D_{(3)}A_{(3)}^\dagger = 0. \tag{4.97}$$

Moreover,

$$\begin{aligned} \rho_{xx^\dagger}(P) &= \mathrm{tr}(\widehat{X_1}\Pi_1) = \mathrm{tr}(\widehat{X_2}\Pi_2) = \mathrm{tr}(\widehat{X_3}\Pi_3) \\ &= \mathrm{tr}(A_{(1)}A_{(1)}^\dagger) = \mathrm{tr}(A_{(2)}A_{(2)}^\dagger) = \mathrm{tr}(A_{(3)}A_{(3)}^\dagger). \end{aligned} \tag{4.98}$$

In particular, for the best low rank approximation of a matrix $X \in \mathbb{C}^{n_1 \times n_2}$ with a matrix of lower rank, from Theorem 4.4.2 we obtain the following conclusions about the critical points.

Theorem 4.4.8 *Let $X \in \mathbb{C}^{n_1 \times n_2}$ be a full-rank matrix with the property that there exists a spectral gap between its m-th and $m+1$ singular values, and let $x := \mathrm{vec}(X)$. Then, the generalized Rayleigh-quotient $\rho_{xx^\dagger}$ has only one minimizer and only one maximizer. The maximizer $(P_1^*, P_2^*) \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ of ρ_A is given by the orthogonal projectors onto the space spanned by the $m_* := \min\{m_1, m_2\}$ $(m_1 \leq n_1/2,\ m_2 \leq n_2/2)$ left, respective right singular vectors corresponding to the largest m_* singular values, and it is nondegenerate. The minimal value of $\rho_{xx^\dagger}$ is 0 and the minimizer is degenerate.*

Proof. Since $x = \mathrm{vec}(X)$, the generalized Rayleigh-quotient can be written as

$$\rho_{xx^\dagger} = \mathrm{tr}(xx^\dagger(P_1 \otimes P_2)) = \mathrm{tr}(X^\dagger P_1 X P_2).$$

Let $(P_1, P_2) \in \mathrm{Gr}(m_1, n_1) \times \mathrm{Gr}(m_2, n_2)$ be a critical point of $\rho_{xx^\dagger}$ with corresponding unitary matrices Θ_1 and Θ_2 and let

$$\Theta_1 X \Theta_2^\dagger = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix},$$

where $X_{11} \in \mathbb{C}^{m \times m}$, $X_{12} \in \mathbb{C}^{m \times (n_2 - m)}$, $X_{21} \in \mathbb{C}^{(n_1 - m) \times m}$, $X_{22} \in \mathbb{C}^{(n_1 - m) \times (n_2 - m)}$. According to Theorem 4.4.2 it is desired to cancel out all those critical points P of $\rho_{xx^\dagger}$ for which $(\mathrm{Im}\, \mathrm{T}_x F_P)^\perp \neq \{0\}$, i.e. the system (4.84)–(4.87) has only the trivial solution. In the matrix case, this means that we are interested in those P for which the system

$$X_{12} Z_1^\dagger + Z_2 X_{21} = 0 \tag{4.99}$$

$$X_{11} Z_1 = 0 \tag{4.100}$$

$$X_{11}^\dagger Z_2 = 0 \tag{4.101}$$

has only the solution $Z_1 = 0,\ Z_2 = 0$. If X_{11} is invertible, then the system (4.99) – (4.101) has $(0, 0)$ as the unique solution. What happens when X_{11} is not invertible? In this situation, let $U \in \mathbb{C}^{k \times m}$ with orthogonal columns span the kernel of X_{11} and $V \in \mathbb{C}^{m \times k}$ with orthogonal lines span the kernel of $X_{11}^\dagger$. From (4.100) and (4.101) it follows that there exists $Y_1 \in \mathbb{C}^{k \times (n_1 - m)}$, $Y_2 \in \mathbb{C}^{k \times (n_2 - m)}$ such that Z_1 and Z_2 can be given as

$$Z_1 = U^\dagger Y_1, \qquad Z_2 = V Y_2.$$

Moreover, since P is a critical point, from Corollary 4.1.8 it follows that

$$X_{12}^\dagger X_{11} = 0, \quad X_{11} X_{21}^\dagger = 0,$$

which means that there exist matrices $A_1 \in \mathbb{C}^{(n_1 - m) \times k}$ and $A_2 \in \mathbb{C}^{k \times (n_2 - m)}$ such that X_{12} and X_{21} are represented as

$$X_{12} = V A_2, \quad X_{21} = A_1 U.$$

Hence, the system (4.99) – (4.101) becomes

$$VA_2Y_1^\dagger U + VY_2A_1U = 0 \Longleftrightarrow A_2Y_1^\dagger + Y_2A_1 = 0,$$

which is an underdetermined $k^2 \times k(n_1 + n_2 - 2m)-$system and hence there exist nontrivial solution $Y_1,\ Y_2$. Thus, the system (4.99) – (4.101) has a unique solution $(Z_1, Z_2) = 0$ if and only if X_{11} is invertible. In this situation, $X_{12} = 0$ and $X_{21} = 0$. Hence, all critical points P of $\rho_{xx^\dagger}$ which satisfy $(\operatorname{Im} \mathrm{T}_x F_P)^\perp = \{0\}$ bring X in the form

$$\Theta_1 X \Theta_2^\dagger = \begin{bmatrix} X_{11} & 0 \\ 0 & X_{22} \end{bmatrix}, \tag{4.102}$$

with X_{11} invertible. The set of all full rank matrices with this property is open and dense in $\mathbb{C}^{n_1 \times n_2}$. Now, let $P := (P_1, P_2)$ be a local maximizer (or local minimizer) of $\rho_{xx^\dagger}$ with corresponding Θ_1 and Θ_2 such that (4.102) holds. For which P the system

$$\mathbf{H}_{\rho_{xx^\dagger}}(P)\xi = 0 \tag{4.103}$$

has only the trivial solution $\xi = (\xi_1, \xi_2) \in \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$? From the Theorem 4.1.11 and equations (4.102), (3.5) and (3.31), the system (4.103) becomes a system in $(Z_1, Z_2) \in \mathbb{C}^{m\times(n_1-m)} \times \mathbb{C}^{m\times(n_2-m)}$, i.e.

$$\begin{aligned} X_{11}X_{11}^\dagger Z_1 - X_{11}Z_2X_{22}^\dagger &= 0 \\ X_{11}^\dagger Z_1 X_{22} - X_{11}^\dagger X_{11} Z_2 &= 0. \end{aligned} \tag{4.104}$$

Using Kronecker-vec formalism, the matrix of the system (4.104) can be written as

$$H = \begin{bmatrix} I_{(n_1-m)} \otimes X_{11}X_{11}^\dagger & -X_{22} \otimes X_{11} \\ -X_{22}^\dagger \otimes X_{11}^\dagger & I_{(n_2-m)} \otimes X_{11}^\dagger X_{11} \end{bmatrix}.$$

Then,

$$\begin{aligned} \det(H) &= \det\Big(I_{(n_1-m)} \otimes X_{11}X_{11}^\dagger\Big) \\ &\quad \det\Big(I_{(n_2-m)} \otimes X_{11}^\dagger X_{11} - X_{22}^\dagger \otimes X_{11}^\dagger (I_{(n_1-m)} \otimes X_{11}X_{11}^\dagger)^{-1} X_{22} \otimes X_{11}\Big) \\ &= \det\Big(I_{(n_1-m)} \otimes X_{11}X_{11}^\dagger\Big)\det\Big(I_{(n_2-m)} \otimes X_{11}^\dagger X_{11} - X_{22}^\dagger X_{22} \otimes I_{(n_1-m)}\Big). \end{aligned}$$

It follows that H is invertible if and only if $\sigma(X_{11}) \cap \sigma(X_{22}) = \emptyset$, where $\sigma(X_{11})$ and $\sigma(X_{22})$ is the set of singular values of X_{11} and X_{22} respectively. Hence, global maximizers are nondegenerate.

When X_{11} is not invertible, the critical point (P_1, P_2) is degenerate. In particular, the global minimizers are degenerate and correspond to those projectors (P_1, P_2) that bring the matrix X in the following form

$$\Theta_1 X \Theta_2^\dagger = \begin{bmatrix} 0_m & X_{12} \\ X_{21} & X_{22} \end{bmatrix}.$$

Here we have assumed that $m \leq \min\{n_1, n_2\}/2$. Hence, for the matrix case, the minimal value of $\rho_{xx^\dagger}$ is $\operatorname{tr}(X_{11}) = 0$. □

Further, we formulate a result on the nondegeneracy of the critical points of the generalized Rayleigh-quotient in the case of rank-one tensor approximations.

Theorem 4.4.9 *Let $m_1 = \cdots = m_r = 1$ and $N = n_1 n_2 \cdots n_r$. Then, for x from a residual subset of $\mathbb{C}^N$, the critical points $(P_1, \ldots, P_r) \in \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ of $\rho_{xx^\dagger}$ which are not global minimizers are nondegenerate. The global minimizers of $\rho_{xx^\dagger}$ are degenerate for all $x \in \mathbb{C}^N$.*

Proof. Set $\mathcal{M} := \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ and define the map

$$F : \mathbb{C}^N \times \mathcal{M} \to \mathbb{R} \times \mathrm{T}\mathcal{M}, \quad (x, P) \mapsto (\rho_{xx^\dagger}(P), \operatorname{grad} \rho_{xx^\dagger}(P)). \tag{4.105}$$

Moreover, let $S := (\mathbb{R} \setminus \{0\}) \times \mathcal{M}_0$, where $\mathcal{M}_0$ is the zero section in the tangent bundle of $\mathcal{M}$. Since $\rho_{xx^\dagger}(P) = \|(P_1 \otimes \cdots \otimes P_r)x\|^2 \geq 0$ and there exists $P \in \mathcal{M}$ such that $\rho_{xx^\dagger}(P) = 0$, it follows that the global minimum is zero, for any $x \in \mathbb{C}^N$. Theorem 4.3.2 states that if $F \pitchfork S$ then, for a generic $x \in \mathbb{C}^N$, the critical points $(P_1, \ldots, P_r)$ of $\rho_{xx^\dagger}$ such that $\rho_{xx^\dagger}(P_1, \ldots, P_r) \neq 0$ different from the global minimizers are nondegenerate. To show that $F \pitchfork S$, we will proceed as in the proof of Corollary 4.3.3 and write the transversality condition in local coordinates.
Let $(x, P) \in \mathbb{C}^N \times \mathcal{M}$ with $\operatorname{grad} \rho_{xx^\dagger}(P) = 0$ and $\rho_{xx^\dagger}(P) \neq 0$ and consider the coordinates φ and ψ around P and $F(x, P)$ exactly as defined in the proof of Corollary 4.3.3. Then, the transversality condition in local coordinates reads as follows

$$\operatorname{Im} D\widetilde{F}(x, 0) + \mathbb{R} \times (\mathrm{T}_P\mathcal{M} \times \{0\}) = \mathbb{R} \times (\mathrm{T}_P\mathcal{M} \times \mathrm{T}_P\mathcal{M}), \tag{4.106}$$

where

$$\widetilde{F}(x, t) = (\widetilde{\rho}_{xx^\dagger}(t), t, \nabla \widetilde{\rho}_{xx^\dagger}(t))$$

and $\widetilde{\rho}_x = \rho_{xx^\dagger} \circ \varphi^{-1}$. One has

$$D\widetilde{F}(x, 0)(y, \xi) = \left(\widetilde{\rho}_{xy^\dagger + yx^\dagger}(0), \xi, \nabla \widetilde{\rho}_{xy^\dagger + yx^\dagger}(0) + d^2 \widetilde{\rho}_{xx^\dagger}(0)\xi \right)$$

and hence, condition (4.106) holds if and only if

$$\operatorname{Im} \nabla \widetilde{\rho}_{x(\cdot)^\dagger + (\cdot)x^\dagger}(0) + \operatorname{Im} d^2 \widetilde{\rho}_{xx^\dagger}(0) = \mathrm{T}_P\mathcal{M}. \tag{4.107}$$

We will prove that $\operatorname{Im} \nabla \widetilde{\rho}_{x(\cdot)^\dagger + (\cdot)x^\dagger}(0) = \mathrm{T}_P \mathcal{M}$. Let $\xi = (\xi_1, \dots, \xi_r) \in \left(\operatorname{Im} \nabla \widetilde{\rho}_{x(\cdot)^\dagger + (\cdot)x^\dagger}(0)\right)^\perp$, then

$$\begin{aligned} 0 &= \langle \nabla \widetilde{\rho}_{xy^\dagger + yx^\dagger}(0), \xi \rangle = \mathrm{d}\widetilde{\rho}_{xy^\dagger + yx^\dagger}(0)(\xi) = \mathrm{d}\rho_{xy^\dagger + yx^\dagger}(P)(\xi) \\ &= 2x^\dagger (\sum_{j=1}^r P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r) y, \end{aligned}$$

for all $y \in \mathbb{C}^N$. This is equivalent to

$$(\sum_{j=1}^r P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r) x = 0. \tag{4.108}$$

We will give the proof for $r = 3$, since it can be straight-forwardly extended to the general case. Let $\Theta_1, \Theta_2, \Theta_3$ be the unitary matrices related to P_1, P_2 and P_3, respectively according to (4.76) and let $\mathcal{T} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ be the tensor corresponding to $\widetilde{x} := (\Theta_1 \otimes \Theta_2 \otimes \Theta_3) x$. As mentioned before, solving (4.108) is equivalent to solving the system (4.84) – (4.87). Since we are interested in rank-one approximations, it follows that $\mathcal{A}$ in (4.84) – (4.86) is a scalar, i.e. $\mathcal{A} = A_{(1)} = A_{(2)} = A_{(3)} = \alpha \in \mathbb{C}$. If $\alpha = 0$, then from (4.98) it follows that $\rho_{xx^\dagger}(P) = 0$, and this contradicts the hypothesis. Hence, from Theorem 4.3.2 the conclusion follows. □

We generalize the above result for the problem of best approximating a tensor with a tensor of lower rank, not necessarily rank–one. In fact, we prove that critical points of the generalized Rayleigh-quotient that satisfy a certain property (that will be mention next) are generically nondegenerate. For this purpose, we define the following map

$$\begin{aligned} F &: \mathbb{C}^N \times \mathrm{Gr}^\times(\mathbf{m}, \mathbf{n}) \to \mathbb{R}^r \times \mathrm{TGr}^\times(\mathbf{m}, \mathbf{n}), \\ &(x, P) \mapsto (\det(L_1), \dots, \det(L_r), \operatorname{grad} \rho_{xx^\dagger}(P)). \end{aligned} \tag{4.109}$$

The matrices L_j are defined next. Let $\Psi_{xx^\dagger, j}$ be the multilinear map defined by (4.6), then

$$\widehat{\Psi}_1 := \Psi_{xx^\dagger,1}(I_{n_1}, P_2, P_3),\ \widehat{\Psi}_2 := \Psi_{xx^\dagger,2}(P_1, I_{n_2}, P_3),\ \widehat{\Psi}_3 := \Psi_{xx^\dagger,1}(P_1, P_2, I_{n_3}).$$

Now, the matrices L_j are given as

$$L_j := \begin{bmatrix} I_{m_j} & 0 \end{bmatrix} P_j \widehat{\Psi}_j P_j \begin{bmatrix} I_{m_j} \\ 0 \end{bmatrix} \in \mathbb{C}^{m_j \times m_j}, \tag{4.110}$$

for $j = 1, \dots, r$.

Theorem 4.4.10 *Let $L_1, \dots, L_r$ be the matrices given by* (4.110)*. If F is the defined by* (4.109) *and $S := (\mathbb{R} \setminus \{0\})^r \times \mathcal{M}_0$, where $\mathcal{M}_0$ is the image of the zero section in the tangent bundle of* $\mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$*, then $F \pitchfork S$. In particular, the set of all $x \in \mathbb{C}^N$ for which the critical points of $\rho_{xx^\dagger}$ with* $\det(L_1) \neq 0, \dots, \det(L_r) \neq 0$*, are nondegenerate, is residual.*

Proof. Let $\mathcal{M} := \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$. Similar to the proof of Theorem 4.4.9, in order to prove that F is transversal to S, we show that the equation

$$(\sum_{j=1}^{r} P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r)x = 0. \tag{4.111}$$

has only the trivial solution $\xi := (\xi_1, \ldots, \xi_r) \in \mathrm{T}_P\mathcal{M}$, for all $(x, P) \in F^{-1}(S)$. As before, we give the proof for the case $r = 3$, the general case follows immediately form this. Let $(x, P) \in F^{-1}(S)$, i.e., grad $\rho_{xx^\dagger}(P) = 0$ and det $(L_1) \neq 0, \ldots,$ det $(L_r) \neq 0$, with $L_1, \ldots, L_r$ given by (4.110). With the notations and technical details already given ((3.5), (4.79)), solving (4.111) is equivalent to solving the system

$$\mathcal{A} \times_1 Z_1^\dagger = 0 \tag{4.112}$$

$$\mathcal{A} \times_2 Z_2^\dagger = 0 \tag{4.113}$$

$$\mathcal{A} \times_3 Z_3^\dagger = 0 \tag{4.114}$$

$$\mathcal{B} \times_1 Z_1 + \mathcal{C} \times_2 Z_2 + \mathcal{D} \times_3 Z_3 = 0, \tag{4.115}$$

for $Z_j \in \mathbb{C}^{m_j \times (n_j - m_j)}$, $j = 1, 2, 3$. We are interested only in the first three equations of the system and we write them in matrix form

$$Z_1^\dagger A_{(1)} = 0 \tag{4.116}$$

$$Z_2^\dagger A_{(2)} = 0 \tag{4.117}$$

$$Z_3^\dagger A_{(3)} = 0, \tag{4.118}$$

where $A_{(j)}$ is the matrix unfolding of the tensor $\mathcal{A}$ along direction j. From

$$\det(L_j) = \det(A_{(j)} A_{(j)}^\dagger) \neq 0$$

it follows that $A_{(j)}$ has full rank m_j, for $j = 1, 2, 3$ and hence, the system (4.112)–(4.115) has only the trivial solution $(Z_1, Z_2, Z_3) = (0, 0, 0)$. Generalizing the Theorem 4.3.2 we prove that the critical points P of the generalized Rayleigh-quotient which satisfy det $(L_1) \neq 0, \ldots,$ det $(L_r) \neq 0$ are generically nondegenerate. $\square$

It follows that the global maximizers of the generalized Rayleigh-quotient are generically nondegenerate.

Corollary 4.4.11 *Let $m_1 \leq n_1/2, \ldots, m_r \leq n_r/2$ be given natural numbers such that* (4.72) *holds for any permutation $\beta \in \mathrm{S}_r$. For generic $x \in \mathbb{C}^N$, the global maximizers of $\rho_{xx^\dagger}$ are nondegenerate.*

Proof. Let $\mathcal{W} \in \mathbb{C}^{n_1 \times \cdots \times n_r}$ denote the tensor corresponding to $x \in \mathbb{C}^N$ according to the lexicographical order. If

$$M_j = \prod_{k=1,\, k \neq j}^{r} m_k, \tag{4.119}$$

and $c_j := \min\{n_j, M_j\}$ for $j = 1, \dots, r$, then the set

$$\mathrm{T}(\mathbf{m},\mathbf{n}) = \bigcap_{j=1}^{r} \left\{ \mathcal{T} \in \mathbb{C}^{n_1\times\cdots\times n_r} \mid \text{any } c_j \text{ columns of } T_{(j)} \text{ are linearly independent } \right\} \tag{4.120}$$

is open and dense in $\mathbb{C}^{n_1\times\cdots\times n_r}$. Let $x \in \mathbb{C}^N$ be such that $\mathcal{W} \in \mathrm{T}(\mathbf{m},\mathbf{n})$ and let P be a global maximizer for $\rho_{xx^\dagger}$.
Let Θ_j be the unitary matrices corresponding to P_j, denote $\widetilde{x} := (\Theta_1 \otimes \Theta_2 \otimes \Theta_3)x$, with $\mathcal{T} = \mathrm{vec}(x)$. Recall matrices $\widehat{X}_1,\ \widehat{X}_2,\ \widehat{X}_3$ given by (4.91), (4.92) and (4.93) respectively. Since $\mathcal{T} \in \mathrm{T}(\mathbf{m},\mathbf{n})$, it follows that

$$\text{rank } \widehat{X}_1 \geq m_1, \quad \text{rank } \widehat{X}_2 \geq m_2, \quad \text{rank } \widehat{X}_3 \geq m_3.$$

From the critical point condition (4.97) we have

$$\widehat{X}_1 = \begin{bmatrix} A_{(1)}A_{(1)}^\dagger & 0 \\ 0 & B_{(1)}B_{(1)}^\dagger \end{bmatrix}, \tag{4.121}$$

$$\widehat{X}_2 = \begin{bmatrix} A_{(2)}A_{(2)}^\dagger & 0 \\ 0 & C_{(2)}C_{(2)}^\dagger \end{bmatrix}, \tag{4.122}$$

$$\widehat{X}_3 = \begin{bmatrix} A_{(3)}A_{(3)}^\dagger & 0 \\ 0 & D_{(3)}D_{(3)}^\dagger \end{bmatrix}, \tag{4.123}$$

where $B_{(1)},\ C_{(2)}$ and $D_{(3)}$ are the matrix unfoldings of $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ along direction 1, 2 and 3 respectively. Moreover, recall that $\rho_{xx^\dagger}(P) = \mathrm{tr}(A_{(1)}A_{(1)}^\dagger) = \mathrm{tr}(A_{(2)}A_{(2)}^\dagger) = \mathrm{tr}(A_{(3)}A_{(3)}^\dagger)$. Since rank $X_j \geq m_j$ and P is a global maximizer for $\rho_{xx^\dagger}$ it follows that $A_{(j)}$ has full rank, for $j = 1, 2, 3$. Thus, $\det\ L_1 \neq 0, \det\ L_2 \neq 0, \det\ L_3 \neq 0$ and according to Theorem 4.4.10, the global maximizers of the generalized Rayleigh-quotient are generically nondegenerate. □

Remark 4.4.12 *Theorem 4.4.9 is a particular case of Theorem 4.4.10 as one notices that for the best rank-one approximation problem* $\det\,(L_1) = \cdots = \det\,(L_r) = \rho_{xx^\dagger}(P)$ *for all* $x \in \mathbb{C}^N$ *and all* $P \in \mathcal{M}$, *when the matrices* $L_1, \dots, L_r$ *are defined in* (4.110).

Chapter 5

Generalized Rayleigh-quotient on Lagrange-Grassmannians

In this chapter we are studying the optimization task of the generalized Rayleigh-quotient of a matrix A on the r-fold tensor product of Lagrange-Grassmannians. The optimization of the classical Rayleigh-quotient is closely related to determining solutions of algebraic Riccati equations with applications in linear optimal control such as Kalman filtering, spectral factorization, etc. (see [39, 63]). Most of the numerical algorithms developed to determine solutions of the algebraic Riccati equation achieve this by computing Lagrangian subspaces of a Hamiltonian matrix, see [33, 43, 52, 54, 67]. The results stated in this chapter hold for all Lagrange-Grassmannian types introduced in Chapter 3, i.e. classical $\mathrm{LG}_{\mathbb{R}}(n)$, $\mathrm{LG}_{\mathbb{C}}(n)$ and complex Lagrange-Grassmannian $\widetilde{\mathrm{LG}}(n)$. First we stress some of the properties of the classical Rayleigh-quotient on $\mathrm{LG}(n)$. In particular, we show that for a symmetric matrix A the optimization of the classical Rayleigh-quotient is equivalent to the optimization of the classical Rayleigh-quotient of the Hamiltonian part of A. We introduce the notions of decomposable symmetric Hamiltonian $A^{\mathfrak{h}}$ and decomposable skew-symmetric Hamiltonian $A^{\mathfrak{s}}$ matrices for a symmetric matrix A. In the case when $A = A^{\mathfrak{h}} + A^{\mathfrak{s}}$, we show that the optimization of ρ_A on $\mathrm{LG}^{\otimes}(\mathbf{n})$ is equivalent to the optimization of $\rho_{A^{\mathfrak{h}}}$ on $\mathrm{LG}^{\otimes}_{\mathbb{K}}(\mathbf{n})$. We derive explicit formulas for the gradient and the Hessian of GRQ on $\mathrm{LG}^{\otimes}(\mathbf{n})$ and characterize the critical points. Moreover, we prove that the critical points of ρ_A on $\mathrm{LG}^{\otimes}(\mathbf{n})$ are nondegenerate for A from an open and dense subset of the space of decomposable symmetric Hamiltonian matrices.

5.1 The generalized Rayleigh-quotient on Lagrange-Grassmannians

We recall some of the previous notations that we will use here as well

$$N := 2n_1 \cdots 2n_r, \quad \mathbf{n} := (n_1, \ldots, n_r), \quad (\mathbf{n}, \mathbf{2n}) := \Big((n_1, 2n_1), \ldots, (n_r, 2n_r)\Big). \tag{5.1}$$

Let $\mathbb{K}$ denote the field of real or complex numbers and let $\mathbb{K}^{2n_j}$ be the standard symplectic space with the symplectic form $J_j = \begin{bmatrix} 0 & I_{n_j} \\ -I_{n_j} & 0 \end{bmatrix}$, for $j = 1, \ldots, r$.

For a matrix $A \in \mathfrak{sym}_N$ when $\mathbb{K} = \mathbb{R}$ (resp. $A \in \mathfrak{her}_N$ when $\mathbb{K} = \mathbb{C}$), we address the optimization task

$$\max_{\mathbf{P} \in \mathrm{LG}^{\otimes}_{\mathbb{K}}(\mathbf{n})} \operatorname{tr}(A\mathbf{P}) \tag{5.2}$$

on the r-fold tensor product of Lagrange-Grassmannians $\mathrm{LG}^{\otimes}_{\mathbb{K}}(\mathbf{n})$ which can be equivalently formulated as the optimization task

$$\max_{(P_1,\ldots,P_r) \in \mathrm{LG}^{\times}_{\mathbb{K}}(\mathbf{n})} \operatorname{tr}\Big(A(P_1 \otimes \cdots \otimes P_r)\Big) \tag{5.3}$$

on the r-fold direct product of Lagrange-Grassmannians $\mathrm{LG}^{\times}_{\mathbb{K}}(\mathbf{n})$. This is a restriction of the optimization task (4.5) discussed in the previous chapter. The map

$$\rho_A : \mathrm{LG}^{\times}_{\mathbb{K}}(\mathbf{n}) \to \mathbb{R}, \quad (P_1, \ldots, P_r) \mapsto \operatorname{tr}\big(A(P_1 \otimes \cdots \otimes P_r)\big), \tag{5.4}$$

is called the *generalized Rayleigh-quotient* of $A \in \mathfrak{sym}_N$ (resp. $A \in \mathfrak{her}_N$) on $\mathrm{LG}^{\times}_{\mathbb{K}}(\mathbf{n})$.

If we have only one Lagrange-Grassmannian $\mathrm{LG}_{\mathbb{K}}(n)$, i.e. $r = 1$, then the optimization task (5.3) determines Lagrangian invariant subspaces of A. This follows directly from the critical point condition (5.35), i.e. $AP = PA$, and the fact that P is a projection operator associated to the Lagrangian subspace $\operatorname{Im} P$. Computing Lagrangian invariant subspaces of Hamiltonian matrices is a classical assignment arising in many applications such as Kalman filtering, spectral factorization, etc. (see [39, 63]). Explicitly, the symmetric (resp. Hermitian) solutions $X \in \mathbb{K}^{n\times n}$ of the algebraic Riccati equation

$$RX + XR - XSX + S = 0, \tag{5.5}$$

define Lagrangian invariant subspaces $W = \operatorname{span} \begin{bmatrix} I_n \\ X \end{bmatrix}$ of the symmetric (resp. Hermitian) Hamiltonian matrix

$$A = \begin{bmatrix} R & S \\ S & -R \end{bmatrix} \in \mathbb{K}^{2n\times 2n}, \tag{5.6}$$

where $R,\ S \in \mathfrak{sym}_n$ are known. Furthermore, the projection operator P associated to W is given by

$$P = \begin{bmatrix} I_n \\ X \end{bmatrix} (I_n + X^2)^{-1} \begin{bmatrix} I_n & X \end{bmatrix}. \tag{5.7}$$

We refer to the literature for a thorough discussion in this direction, [48].

We say that a matrix $A \in \mathbb{K}^{2n\times 2n}$ is *skew-Hamiltonian* if

$$JA = -(JA)^\top, \tag{5.8}$$

where J is the standard symplectic form on $\mathbb{K}^{2n}$. Every matrix $A \in \mathbb{K}^{2n\times 2n}$ can be decomposed in a Hamiltonian and a skew-Hamiltonian part

$$A = A_{\mathfrak{h}} + A_{\mathfrak{s}}, \quad A_{\mathfrak{h}} = \frac{A + JA^\top J}{2}, \; A_{\mathfrak{s}} = \frac{A - JA^\top J}{2}. \tag{5.9}$$

Recall that a matrix $A \in \mathbb{C}^{2n\times 2n}$ is called *complex Hamiltonian* if

$$JA = (JA)^\dagger.$$

Moreover, we say that a matrix $A \in \mathbb{C}^{2n\times 2n}$ is *complex skew-Hamiltonian* if

$$JA = -(JA)^\dagger.$$

Then, any matrix $A \in \mathbb{C}^{2n\times 2n}$ can be given as

$$A = A_{\mathfrak{h}} + A_{\mathfrak{s}}, \quad A_{\mathfrak{h}} = \frac{A + JA^\dagger J}{2}, \; A_{\mathfrak{s}} = \frac{A - JA^\dagger J}{2}. \tag{5.10}$$

For the optimization of the classical Rayleigh-quotient of a symmetric matrix on the Lagrange-Grassmannian only the Hamiltonian part of the matrix counts, as we show in the next lemma.

Lemma 5.1.1 *Let $A \in \mathfrak{sym}_{2n}$ (resp. $A \in \mathfrak{her}_{2n}$). Then the maximization problem*

$$\max_{P\in \mathrm{LG}_{\mathbb{K}}(n)} \operatorname{tr}(AP) \tag{5.11}$$

is equivalent to the following one

$$\max_{P\in \mathrm{LG}_{\mathbb{K}}(n)} \operatorname{tr}\Big(A_{\mathfrak{h}}P\Big), \tag{5.12}$$

where $A_{\mathfrak{h}} = \dfrac{A + JA^\top J}{2}$ is the Hamiltonian part of A.

Proof. Every matrix $A \in \mathfrak{sym}_{2n}$ can be written as a sum of its Hamiltonian part and its skew-Hamiltonian part, which are both symmetric, i.e. $A = A_{\mathfrak{h}} + A_{\mathfrak{s}}$. For every $P \in \mathrm{LG}_{\mathbb{K}}(n)$, a straight-forward computation shows that

$$JPJ = P - I_{2n}.$$

Then

$$\begin{aligned} \operatorname{tr}(A_{\mathfrak{h}}P) &= \tfrac{1}{2}\operatorname{tr}((JAJ + A)P) = \tfrac{1}{2}\operatorname{tr}(AJPJ) + \tfrac{1}{2}\operatorname{tr}(AP) \\ &= \operatorname{tr}(AP) - \tfrac{1}{2}\operatorname{tr}(A). \end{aligned}$$

Since $\operatorname{tr}(A)$ is a constant, the conclusion follows. For $A \in \mathfrak{her}_{2n}$ and $\mathrm{LG}_{\mathbb{C}}(n)$, we use the fact that

$$JPJ = P^\top - I_{2n}$$

is true for any $P \in \mathrm{LG}_{\mathbb{C}}(n)$, and thus, the conclusion follows. $\square$

From the proof of the above Lemma it follows that for all $P \in \mathrm{LG}_{\mathbb{K}}(n)$ one has

$$\mathrm{tr}(A_{\mathfrak{s}}P) = \frac{1}{2}\,\mathrm{tr}(A),$$

where $A_{\mathfrak{s}}$ is the skew-Hamiltonian part of $A \in \mathfrak{sym}_{2n}$ (resp. $A \in \mathfrak{her}_{2n}$).

An identical result is true also for the optimization of the Rayleigh-quotient of a Hermitian matrix on the complex Lagrange-Grassmannian, i.e.

$$\max_{P \in \widehat{\mathrm{LG}}(n)} \mathrm{tr}(AP) \quad \equiv \quad \max_{P \in \widehat{\mathrm{LG}}(n)} \mathrm{tr}\big(A_{\mathfrak{h}}P\big),$$

where $A_{\mathfrak{h}} = \dfrac{A + JAJ}{2}$ is the complex Hamiltonian part of A. In this situation, the critical points of the Rayleigh-quotient are complex Lagrangian invariant subspaces of the complex Hamiltonian part $A_{\mathfrak{h}}$ of $A \in \mathfrak{her}_{2n}$.

5.2 Decomposable Hamiltonian matrices

In this section we introduce a special class of matrices that we will call decomposable Hamiltonian matrices. Let $J_1, \dots, J_r$ denote the standard symplectic forms on $\mathbb{K}^{2n_1}, \dots, \mathbb{K}^{2n_r}$ respectively. We know from Section 3.4 that for r odd, the skew-symmetric nondegenerate matrix $J_1 \otimes \cdots \otimes J_r$ defines a symplectic bilinear form on $\mathbb{K}^N \cong \mathbb{K}^{2n_1} \otimes \cdots \otimes \mathbb{K}^{2n_r}$ as

$$\omega : \mathbb{K}^N \times \mathbb{K}^N \to \mathbb{R}, \quad \omega(x, y) = x^\top (J_1 \otimes \cdots \otimes J_r) y. \tag{5.13}$$

The notions of Hamiltonian or complex Hamiltonian we consider with respect to ω, i.e.

$$\begin{aligned} A(J_1 \otimes \cdots \otimes J_r) &= -(J_1 \otimes \cdots \otimes J_r)A^\top \quad \text{(Hamiltonian)} \\ A(J_1 \otimes \cdots \otimes J_r) &= -(J_1 \otimes \cdots \otimes J_r)A^\dagger \quad \text{(complex Hamiltonian)} \end{aligned} \tag{5.14}$$

Definition 5.2.1 *Let $n_1, \dots, n_r \in \mathbb{N}$ be given and $N = 2n_1 2n_2 \cdots 2n_r$. A matrix $A \in \mathbb{K}^{N \times N}$ is called* decomposable Hamiltonian *if it can be obtained as a summation of Kronecker products of Hamiltonian matrices, i.e.*

$$A = \sum_{k=1}^{L} X_{1k} \otimes \cdots \otimes X_{rk}, \tag{5.15}$$

where $X_{jk} \in \mathfrak{sp}(2n_j, \mathbb{K})$, for $j = 1, \dots, r$. The space of all decomposable Hamiltonian matrices is the tensor product space

$$\mathfrak{sp}^{\otimes}(\mathbf{2n}, \mathbb{K}) := \mathfrak{sp}(2n_1, \mathbb{K}) \otimes \cdots \otimes \mathfrak{sp}(2n_r, \mathbb{K}). \tag{5.16}$$

Similar to decomposable Hamiltonian matrices, the decomposable skew-Hamiltonian *matrices are matrices $B \in \mathbb{K}^{N\times N}$ which can be obtained as a sum of Kronecker products of skew-Hamiltonian matrices, i.e.*

$$B = \sum_{k=1}^{L} Y_{1k} \otimes \cdots \otimes Y_{rk}, \tag{5.17}$$

where $Y_{jk} \in \mathbb{K}^{2n_j\times 2n_j}$ skew-Hamiltonian, for $j = 1, \dots, r$.

By counting dimensions

$$\dim_{\mathbb{K}}\ \mathfrak{sp}(N,\mathbb{K}) = \frac{N}{2}(N+1) > \prod_{j=1}^{r} n_j(2n_j+1) = \dim_{\mathbb{K}}\ \mathfrak{sp}^{\otimes}(\mathbf{2n},\mathbb{K}),$$

it is clear that not every Hamiltonian matrix can be a decomposable Hamiltonian matrix. As well we can argue that not all skew-Hamiltonian matrices are decomposable skew-Hamiltonian matrices.
We define a finite iterative process to determine the decomposable Hamiltonian and skew-Hamiltonian parts of a matrix $A \in \mathbb{K}^{N\times N}$:

$$\begin{aligned}
A_1 &= \frac{(J_1 \otimes I_{2n_2} \cdots \otimes I_{2n_r})A^\top (J_1 \otimes I_{2n_2} \cdots \otimes I_{2n_r}) + A}{2} \\
A_2 &= \frac{(I_{2n_1} \otimes J_2 \cdots \otimes I_{2n_r})A_1^\top (I_{2n_1} \otimes J_2 \cdots \otimes I_{2n_r}) + A_1}{2} \\
&\vdots \\
A_r &= \frac{(I_{2n_1} \otimes I_{2n_2} \cdots \otimes J_r)A_{r-1}^\top (I_{2n_1} \otimes I_{2n_2} \cdots \otimes J_r) + A_{r-1}}{2},
\end{aligned} \tag{5.18}$$

and

$$\begin{aligned}
B_1 &= \frac{A - (J_1 \otimes I_{2n_2} \cdots \otimes I_{2n_r})A^\top (J_1 \otimes I_{2n_2} \cdots \otimes I_{2n_r})}{2} \\
B_2 &= \frac{B_1 - (I_{2n_1} \otimes J_2 \cdots \otimes I_{2n_r})B_1^\top (I_{2n_1} \otimes J_2 \cdots \otimes I_{2n_r})}{2} \\
&\vdots \\
B_r &= \frac{B_{r-1} - (I_{2n_1} \otimes I_{2n_2} \cdots \otimes J_r)B_{r-1}^\top (I_{2n_1} \otimes I_{2n_2} \cdots \otimes J_r)}{2},
\end{aligned} \tag{5.19}$$

In conclusion, A_r is the decomposable Hamiltonian part of A and B_r is the decomposable skew-Hamiltonian part of A.

In contrast to the classical symplectic spaces, every complex Hamiltonian matrix $A \in \mathbb{C}^{N\times N}$ is also a *complex decomposable Hamiltonian* matrix, i.e.,

$$A = \sum_{k=1}^{K} X_{1k} \otimes \cdots \otimes X_{rk}, \tag{5.20}$$

where $X_{jk} \in \widehat{\mathfrak{sp}}(2n_j, \mathbb{C})$, for $j = 1, \ldots, r$. This is straight-forward by counting dimensions

$$\dim\Big(\widehat{\mathfrak{sp}}(2n_1, \mathbb{C}) \otimes \cdots \otimes \widehat{\mathfrak{sp}}(2n_r, \mathbb{C})\Big) = \prod_{j=1}^{r} 4n_j^2 = N^2 = \dim(\widehat{\mathfrak{sp}}(N, \mathbb{C})).$$

Let $\widehat{\mathfrak{hsp}}(2n, \mathbb{C})$ denote the vector space of complex Hamiltonian matrices which are also Hermitian, i.e.

$$\widehat{\mathfrak{hsp}}(2n, \mathbb{C}) := \widehat{\mathfrak{sp}}(2n, \mathbb{C}) \cap \mathfrak{her}_{2n}. \tag{5.21}$$

Definition 5.2.2 *Let r be an odd number, $n_1, \ldots, n_r \in \mathbb{N}$ be given and denote $N = 2^r n_1 n_2 \cdots n_r$. A matrix $A \in \widehat{\mathfrak{hsp}}(N, \mathbb{C})$ is called* decomposable Hermitian Hamiltonian *if it can be obtained as a sum of Kronecker products of Hermitian Hamiltonian matrices*

$$A = \sum_{k=1}^{K} X_{1k} \otimes \cdots \otimes X_{rk}, \tag{5.22}$$

where $X_{jk} \in \widehat{\mathfrak{hsp}}(2n_j, \mathbb{C})$, for $j = 1, \ldots, r$. A matrix A is called decomposable Hermitian skew-Hamiltonian *if it can be written as*

$$A = \sum_{k=1}^{K} Y_{1k} \otimes \cdots \otimes Y_{rk}, \tag{5.23}$$

where $Y_{jk} \in \mathfrak{her}_{2n_j}$ skew-Hamiltonian, for $j = 1, \ldots, r$.

The set of all decomposable Hermitian Hamiltonian matrices is by definition the tensor product space

$$\mathfrak{hsp}^{\otimes}(\mathbf{2n}) := \widehat{\mathfrak{hsp}}(2n_1, \mathbb{C}) \otimes \cdots \otimes \widehat{\mathfrak{hsp}}(2n_r, \mathbb{C}). \tag{5.24}$$

Unlike to the general case of complex Hamiltonian matrices, not every Hermitian Hamiltonian matrix is a decomposable Hermitian Hamiltonian matrix as it can be seen by counting the dimension of the corresponding vector spaces

$$\dim\Big(\mathfrak{hsp}^{\otimes}(\mathbf{2n}) \quad = \prod_{j=1}^{r} \dim\Big(\widehat{\mathfrak{hsp}}(2n_j, \mathbb{C})\Big) = \prod_{j=1}^{r} 2n_j^2 \neq \tfrac{N^2}{2} = \dim \widehat{\mathfrak{hsp}}(N, \mathbb{C}).$$

Now, we can formulate a similar result to Lemma 5.1.1 for the optimization of the generalized Rayleigh-quotient of a symmetric (or Hermitian) matrix on $\mathrm{LG}_{\mathbb{K}}^{\times}(\mathbf{n})$. First, notice that any matrix $A \in \mathfrak{sym}_N$ can be written as

$$A = A^{\mathfrak{h}} + A^{\mathfrak{s}} + R, \tag{5.25}$$

where $A^{\mathfrak{h}}, A^{\mathfrak{s}}$ are the decomposable Hamiltonian and decomposable skew-Hamiltonian parts of A and R is in the orthogonal complement of the space spanned by the space of decomposable Hamiltonian and decomposable skew-Hamiltonian matrices.

Proposition 5.2.3 *Let $A \in \mathfrak{sym}_N$ (resp. $A \in \mathfrak{her}_N$) be of the form*

$$A = A^{\mathfrak{h}} + A^{\mathfrak{s}}, \tag{5.26}$$

where $A^{\mathfrak{h}}$, $A^{\mathfrak{s}} \in \mathfrak{sym}_N$ (resp. $A^{\mathfrak{h}}$, $A^{\mathfrak{s}} \in \mathfrak{her}_N$) are the decomposable Hamiltonian and the decomposable skew-Hamiltonian parts of A. Then, the optimization problem

$$\max_{(P_1,\dots,P_r)\in \mathrm{LG}_{\mathbb{K}}^{\times}(\mathbf{n})} \operatorname{tr}\Big(A(P_1 \otimes \cdots \otimes P_r)\Big) \tag{5.27}$$

is equivalent to

$$\max_{(P_1,\dots,P_r)\in \mathrm{LG}_{\mathbb{K}}^{\times}(\mathbf{n})} \operatorname{tr}\Big(A^{\mathfrak{h}}(P_1 \otimes \cdots \otimes P_r)\Big). \tag{5.28}$$

Proof. The result follows from the fact that $\operatorname{tr}(XP) = \frac{1}{2}\operatorname{tr}(X)$ for all $X \in \mathfrak{sym}_{2n}$ (resp. $X \in \mathfrak{her}_{2n}$) skew-Hamiltonian and all $P \in \mathrm{LG}_n$. $\square$

5.3 Riemannian gradient and Hessian of the generalized Raleigh-quotient

In the sequel, we derive expressions for the gradient and the Hessian of the generalized Rayleigh-quotient of a symmetric or Hermitian matrix on $\mathrm{LG}_{\mathbb{K}}^{\times}(\mathbf{n})$ and $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$. As in (5.1), we have $N := 2n_1 \cdots 2n_r$, $\mathbf{n} = (n_1,\dots,n_r)$. Previously, we have seen that the optimization of the classical Rayleigh-quotient (RQ) of a symmetric or Hermitian matrix A on $\mathrm{LG}_{\mathbb{K}}(n)$ or $\widehat{\mathrm{LG}}(n)$ can be reduced to the optimization of the RQ of the Hamiltonian or complex Hamiltonian part of the matrix. Since not every symmetric/Hermitian matrix can be decomposed only in a decomposable Hamiltonian and a decomposable skew-Hamiltonian part, we will analyze the optimization of the generalized Rayleigh-quotient (GRQ) for a symmetric/Hermitian matrix in general. We will discuss only the optimization of GRQ on $\mathrm{LG}_{\mathbb{C}}^{\times}(\mathbf{n})$ (the results are easily translated for $\mathrm{LG}_{\mathbb{R}}^{\times}(\mathbf{n})$) and on $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$.

Let ρ_A be the generalized Rayleigh-quotient of $A \in \mathfrak{sym}_N$ on $\mathrm{LG}_{\mathbb{C}}^{\times}(\mathbf{n})$, and $\widehat{\rho}_A$ be the generalized Rayleigh-quotient of $A \in \mathfrak{her}_N$ on $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$. Using the submanifold structure of $\mathrm{LG}_{\mathbb{C}}^{\times}(\mathbf{n})$ and $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$, we derive the formulas for the gradient and the Hessian of ρ_A and $\widehat{\rho}_A$ by projecting the analog objects from $\mathrm{Gr}_{\mathbb{C}}^{\times}(\mathbf{n},\mathbf{2n})$. For simplicity and since there is no danger of confusion, we will refer to $\mathrm{LG}_{\mathbb{C}}^{\times}(\mathbf{n})$ by $\mathrm{LG}^{\times}(\mathbf{n})$.

Let $P := (P_1,\dots,P_r) \in \mathrm{LG}^{\times}(\mathbf{n})$, then the map

$$\begin{aligned} \pi_P = (\pi_{P_1},\dots,\pi_{P_r}) &: \mathrm{T}_P\mathrm{Gr}_{\mathbb{C}}^{\times}(\mathbf{n},\mathbf{2n}) \to \mathrm{T}_P\mathrm{Gr}_{\mathbb{C}}^{\times}(\mathbf{n},\mathbf{2n}), \\ (\xi_1,\dots,\xi_r) &\mapsto \left(\frac{J_1\xi_1^{\top}J_1 + \xi_1}{2},\dots,\frac{J_r\xi_r^{\top}J_r + \xi_r}{2}\right) \end{aligned} \tag{5.29}$$

is the orthogonal projector onto $\mathrm{T}_P\mathrm{LG}^\times(\mathbf{n})$, and

$$\widehat{\pi}_P = (\widehat{\pi}_{P_1}, \ldots, \widehat{\pi}_{P_r}) : \mathrm{T}_P\mathrm{Gr}^\times_{\mathbb{C}}(\mathbf{n}, \mathbf{2n}) \to \mathrm{T}_P\mathrm{Gr}^\times_{\mathbb{C}}(\mathbf{n}, \mathbf{2n}),$$
$$(\xi_1, \ldots, \xi_r) \mapsto \left(\frac{J_1\xi_1J_1 + \xi_1}{2}, \ldots, \frac{J_r\xi_rJ_r + \xi_r}{2}\right) \tag{5.30}$$

is the orthogonal projector onto $\mathrm{T}_P\widehat{\mathrm{LG}}^\times(\mathbf{n})$, $P \in \widehat{\mathrm{LG}}^\times(\mathbf{n})$. For $A \in \mathbb{C}^{N\times N}$, recall from Section 4.1.1 the multilinear maps $\Psi_{A,j} : \mathfrak{her}_{2n_1} \times \cdots \times \mathfrak{her}_{2n_r} \to \mathbb{C}^{2n_j\times 2n_j}$ defined by

$$\mathrm{tr}(A^\dagger(X_1 \otimes \cdots \otimes X_jZ \otimes \cdots \otimes X_r) = \mathrm{tr}(\Psi_{A,j}(X_1, \ldots, X_r)^\dagger Z), \tag{5.31}$$

for all $Z \in \mathbb{C}^{n_j\times n_j}$ and for all $j = 1, \ldots, r$. We know that if $A \in \mathfrak{her}_N$ and $(P_1, \ldots, P_r) \in \mathrm{LG}^\times(\mathbf{n})$, the matrix $\Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r)$ is also Hermitian. However, if A is Hamiltonian, it does not mean that $\Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r)$ is Hamiltonian as well. This is true if $A \in \mathfrak{her}$ is decomposable Hamiltonian, as the next lemma will show.

Lemma 5.3.1 *Let $A \in \mathfrak{her}_N$ be decomposable Hamiltonian matrix and $(P_1, \ldots, P_r)$ an element in $\mathrm{LG}^\times(\mathbf{n})$. Then, $\Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r) \in \mathbb{C}^{2n_j\times 2n_j}$ is Hermitian and Hamiltonian for all $j = 1, \ldots, r$.*

Proof. Without loss of generality, we can assume that $A \in \mathfrak{sp}^\otimes(\mathbf{2n}, \mathbb{C})$ is of the form $A_1 \otimes \cdots \otimes A_r$ with $A_j \in \mathfrak{sp}(2n_j, \mathbb{C}) \cap \mathfrak{her}_{2n_j}$, for $j = 1, \ldots, r$. Then,

$$\mathrm{tr}(A(P_1 \otimes \cdots \otimes P_r)) = \prod_{j=1}^{r} \mathrm{tr}(A_jP_j)$$

and hence,

$$\Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r) = A_j \prod_{i=1,\, i\neq j}^{r} \mathrm{tr}(A_iP_i) \in \mathfrak{sp}(2n_j, \mathbb{C}).$$

□

Moreover, when the point $P \in \mathrm{LG}^\times(\mathbf{n})$ is understood from the context, we use the following shortcut

$$\widehat{A_j} := \Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r), \tag{5.32}$$

for $j = 1, \ldots, r$. Now, the gradient of the generalized Rayleigh-quotient of A on $\mathrm{LG}^\times(\mathbf{n})$ or $\widehat{\mathrm{LG}}^\times(\mathbf{n})$ can be explicitly given.

Theorem 5.3.2 *Let $A \in \mathfrak{her}_N$ and $P := (P_1, \ldots, P_r)$ an element in $\mathrm{LG}^\times(\mathbf{n})$ (resp. in $\widehat{\mathrm{LG}}^\times(\mathbf{n})$). Then, one has:*

(i) The gradient of ρ_A at $P \in \mathrm{LG}^\times(\mathbf{n})$ (resp. $\widehat{\rho}_A$ at $P \in \widehat{\mathrm{LG}}^\times(\mathbf{n})$) with respect to the Riemannian metric induced by (3.19) is

$$\mathrm{grad}\ \rho_A(P) = \pi_P\left(\mathrm{ad}^2_{P_1}\widehat{A}_1, \ldots, \mathrm{ad}^2_{P_r}\widehat{A}_r\right), \tag{5.33}$$

$$\left(\textit{resp.}\ \operatorname{grad} \widehat{\rho}_A(P) = \widehat{\pi}_P \left(\operatorname{ad}^2_{P_1} \widehat{A}_1, \ldots, \operatorname{ad}^2_{P_r} \widehat{A}_r \right) \right). \tag{5.34}$$

(ii) The critical points of ρ_A on $\mathrm{LG}^\times(\mathbf{n})$ (resp. $\widehat{\rho_A}$ on $\widehat{\mathrm{LG}}^\times(\mathbf{n})$) are characterized by

$$\left[P_j, \frac{J_j \widehat{A}_j^\top J_j + \widehat{A}_j}{2} \right] = 0, \tag{5.35}$$

$$\left(\textit{resp.}\ \left[P_j, \frac{J_j \widehat{A}_j J_j + \widehat{A}_j}{2} \right] = 0, \right), \tag{5.36}$$

for all $j = 1, \ldots, r$, i.e. P_j is a self-adjoint projector associated to an invariant Lagrangian subspace of the Hamiltonian part of $\widehat{A}_j$ (resp. P_j is the self-adjoint projector associated to an invariant complex Lagrangian subspace of the complex Hamiltonian part of $\widehat{A}_j$).

Proof. The conclusion (5.33) and (5.34) follow immediately from Theorem 4.1.7 with the projection operators (5.29) and (5.30). Conclusions (5.35) and (5.36) can be drawn from Theorem 4.1.7 and the equalities

$$J(\operatorname{ad}^2_P X)^\top J = \operatorname{ad}^2_P J X^\top J \quad \text{and} \quad J(\operatorname{ad}^2_P X) J = \operatorname{ad}^2_P J X J,$$

for $P \in \mathrm{LG}(n)$ and $P \in \widehat{LG}(n)$, respectively. Here $X \in \mathfrak{her}_{2n}$ and J the skew-symmetric matrix defined by (3.26). $\square$

Next, we formulate a necessary and sufficient critical point condition.

Corollary 5.3.3 *Let $A \in \mathfrak{her}_N$, $P := (P_1, \ldots, P_r) \in \mathrm{LG}^\times(\mathbf{n})$ and chose $\Theta_j \in \mathrm{Sp}(n_j)$ such that $\overline{\Theta}_j P_j \Theta_j^\top = \Pi_j$, where Π_j is the standard projector of $\mathbb{C}^{2n_j}$, for $j = 1, \ldots, r$. We write*

$$\overline{\Theta_j} \widehat{A}_j \Theta_j^\top = \begin{bmatrix} \Psi'_j & \Psi'''_j \\ \Psi'''^\top_j & \Psi''_j \end{bmatrix}, \tag{5.37}$$

with Ψ'_j, $\Psi''_j, \Psi'''_j \in \mathfrak{her}_{n_j}$ and $\Psi'''_j \in \mathbb{C}^{n_j \times n_j}$, for $j = 1, \ldots, r$. Then, P is a critical point of ρ_A if and only if

$$\Psi'''_j = -\Psi'''^\top_j, \tag{5.38}$$

for all $j = 1, \ldots, r$. Moreover, if A is decomposable Hamiltonian matrix, then

$$\Psi''_j = -\Psi'_j \quad \textit{and} \quad \Psi'''_j = (\Psi'''_j)^\top,$$

and hence, in the critical points

$$\Psi'''_j = 0,$$

for all $j = 1, \ldots, r$. A similar statement holds for $P \in \widehat{\mathrm{LG}}^\times(\mathbf{n})$ with transpose replaced by conjugate transpose.

Proof. Let $P := (P_1, \ldots, P_r) \in \mathrm{LG}^{\times}(\mathbf{n})$ be a critical point for ρ_A given in the form $\overline{\Theta_j} P_j \Theta_j^{\top} = \Pi_j$, with $\Theta_j \in \mathrm{Sp}(n_j)$ for $j = 1, \ldots, r$. Since (5.35) holds in the critical point, we have that

$$\overline{\Theta_j}\pi_{P_j}(\widehat{A}_j)\Theta_j^{\top} = \begin{bmatrix} X_j & 0 \\ 0 & -X_j \end{bmatrix},$$

with $X_j \in \mathfrak{her}_{n_j}$, for all $j = 1, \ldots, r$. Moreover,

$$\begin{aligned} \overline{\Theta_j}\pi_{P_j}(\widehat{A}_j)\Theta_j^{\top} &= \overline{\Theta_j} J_j \widehat{A}_j J_j \Theta_j^{\top} + \overline{\Theta_j}\widehat{A}_j\Theta_j^{\top} \\ &= J_j \overline{\Theta_j}\widehat{A}_j\Theta_j^{\top} J_j + \overline{\Theta_j}\widehat{A}_j\Theta_j^{\top} \\ &= \pi_{P_j}\left(\overline{\Theta_j}\widehat{A}_j\Theta_j^{\top}\right), \end{aligned}$$

for all $j = 1, \ldots, r$ and hence, $X_j = \Psi_j' - \overline{\Psi_j''}$ and $\Psi_j''' + (\Psi_j''')^{\top} = 0$, for Ψ_j', Ψ_j'', Ψ_j''' from the block–structure (5.37), for all $j = 1, \ldots, r$. $\square$

Let $\widetilde{\rho}_A : \mathrm{Gr}_{\mathbb{K}}^{\times}(\mathbf{n}, \mathbf{2n}) \to \mathbb{R}$ denote the extension of ρ_A and $\widehat{\rho}_A$ to $\mathrm{Gr}_{\mathbb{R}}^{\times}(\mathbf{n}, \mathbf{2n})$ and $\mathrm{Gr}_{\mathbb{C}}^{\times}(\mathbf{n}, \mathbf{2n})$, respectively. The Hessian of ρ_A at $P \in \mathrm{LG}^{\times}(\mathbf{n})$ (resp. $\widehat{\rho}_A$ at $P \in \widehat{\mathrm{LG}}^{\times}(\mathbf{n})$) is given as follows.

Theorem 5.3.4 *Let $A \in \mathfrak{her}_N$ and $P := (P_1, \ldots, P_r) \in \mathrm{LG}^{\times}(\mathbf{n})$ (resp. $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$). Then, the Riemannian Hessian of ρ_A (resp. $\widehat{\rho}_A$) at P is the self-adjoint operator*

$$\mathbf{H}_{\rho_A}(P) : \mathrm{T}_P\mathrm{LG}^{\times}(\mathbf{n}) \to \mathrm{T}_P\mathrm{LG}^{\times}(\mathbf{n}), \tag{5.39}$$

defined by

$$\mathbf{H}_{\rho_A}(P) = \pi_P \circ \mathbf{H}_{\widetilde{\rho}_A}(P) \circ i, \tag{5.40}$$

where

$$i : \mathrm{T}_P\mathrm{LG}^{\times}(\mathbf{n}) \to \mathrm{T}_P\mathrm{Gr}^{\times}(\mathbf{n}, \mathbf{2n}), \quad \xi \mapsto \xi$$

is the inclusion map. If $P \in \widehat{\mathrm{LG}}^{\times}(\mathbf{n})$, then there is a corresponding result obtained by replacing π_P in (5.40) by $\widehat{\pi}_P$.

Proof. We can argue as in the proof of Theorem 4.1.11. The result follows from the fact that the geodesics on $\mathrm{LG}^{\times}(\mathbf{n})$ and on $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$ are restrictions of the geodesics on $\mathrm{Gr}^{\times}(\mathbf{n}, \mathbf{2n})$. $\square$

For completeness, we give next the explicit formula for the Hessian of ρ_A at $P \in \mathrm{LG}^{\times}(\mathbf{n})$.

Corollary 5.3.5 *With the same hypothesis as in Theorem 5.3.4, the Hessian $\mathbf{H}_{\rho_A}(P) = (\mathbf{H}_{\rho_A}^1, \ldots, \mathbf{H}_{\rho_A}^r)$ is expressed as*

$$\mathbf{H}_{\rho_A}^j(\xi) := -\mathrm{ad}_{P_j}\mathrm{ad}_{(\widehat{A}_j)_{\mathfrak{h}}}\xi_j + \sum_{k=1, k\neq j}^{r} \mathrm{ad}_{P_j}^2\left(\Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, \xi_k, \ldots, P_r)\right)_{\mathfrak{h}}, \tag{5.41}$$

for all $\xi := (\xi_1, \dots, \xi_r) \in \mathrm{T}_P\mathrm{LG}^\times(\mathbf{n})$ *and* $j = 1, \dots, r$*, where* $\widehat{A}_j$ *is given in* (5.32) *and* $(X)_\mathfrak{h}$ *is the Hamiltonian part of* X*. For* $P \in \widehat{\mathrm{LG}}^\times(\mathbf{n})$ *there is a corresponging result obtained by the Hamiltonian part by the complex Hamiltonian part.*

Proof. Let $P := (P_1, \dots, P_r) \in \mathrm{LG}^\times(\mathbf{n})$ and $\widetilde{\rho}_A$ be the extension of ρ_A to the manifold $\mathrm{Gr}^\times(\mathbf{n}, \mathbf{2n})$. Then, from (5.40) one obtains

$$\mathbf{H}^j_{\rho_A}(\xi) = \pi_{P_j}(\mathbf{H}^j_{\widetilde{\rho}_A}(\xi)) \tag{5.42}$$

for all $\xi \in \mathrm{T}_P\mathrm{LG}^\times(\mathbf{n})$ and $j = 1, \dots, r$. Moreover, formula (5.41) follows from (4.20) with the specifications

$$\begin{aligned} J_j[P_j, X]J_j &= J_jP_jXJ_j - J_jXP_jJ_j \\ &= -J_jP_jJ_jJ_jXJ_j + J_jXJ_jJ_jP_jJ_j \\ &= -(P_j - I_{n_j})J_jXJ_j + J_jXJ_j(P_j - I_{n_j}) \\ &= [P_j, J_jXJ_j], \end{aligned}$$

$$\begin{aligned} \pi_{P_j}(\mathrm{ad}_{P_j}\mathrm{ad}_{\widehat{A}_j}\xi_j) &= \frac{J_j[P_j, [\widehat{A}_j, \xi_j]]J_j + [P_j, [\widehat{A}_j, \xi_j]]}{2} \\ &= \frac{[P_j, J_j[\widehat{A}_j, \xi_j]J_j] + [P_j, [\widehat{A}_j, \xi_j]]}{2} \\ &= \frac{[P_j, J_j[\widehat{A}_j, \xi_j]J_j + [\widehat{A}_j, \xi_j]]}{2} \\ &= \mathrm{ad}_{P_j}\mathrm{ad}_{(\widehat{A}_j)_\mathfrak{h}}\xi_j \end{aligned}$$

and

$$\pi_{P_j}(\mathrm{ad}^2_{P_j}Y) = \mathrm{ad}^2_{P_j}(Y)_\mathfrak{h},$$

for all $\xi \in \mathrm{T}_P\mathrm{LG}^\times(\mathbf{n})$ and all $X,\ Y \in \mathfrak{her}_{2n_j}$. □

Remark 5.3.6 *For decomposable Hamiltonian matrices* A*, the given expressions of the gradient and Hessian of* ρ_A *do no longer need to project the matrices* $\widehat{A}_j$ *and* $\Psi_{A,j}(P_1, \dots, I_{n_j}, \dots, \xi_k, \dots, P_r)$ *onto their Hamiltonian parts, since they are already Hamiltonian.*

Further, we give necessary conditions such that the Hessian of ρ_A on $\mathrm{LG}^{\times r}(\mathbf{n})$ and on $\widehat{\mathrm{LG}}^{\times r}(\mathbf{n})$ is nondegenerate in local extrema.

Theorem 5.3.7 *Let* $A \in \mathfrak{her}_N$ *be decomposable Hamiltonian and* $P \in \mathrm{LG}^\times(\mathbf{n})$ *(resp.* $P \in \widehat{\mathrm{LG}}^\times(\mathbf{n})$*) a local maximizer or a local minimizer of* ρ_A *(resp.* $\widehat{\rho}_A$*). If* $\mathbf{H}_{\rho_A}(P)$ *(resp.* $\mathbf{H}_{\widehat{\rho}_A}(P)$*) is nondegenerate, then, for* $j = 1, \dots, r$ *the matrix* Ψ'_j *cannot have as eigenvalues both* $\lambda_j \in \mathbb{R}$ *and* $-\lambda_j \in \mathbb{R}$*. Here* Ψ'_j *is defined by* (5.37)*.*

Proof. Assume that $P \in \mathrm{LG}^{\times}(\mathbf{n})$ is a local maximizer of ρ_A. If $\mathbf{H}_{\rho_A}(P)$ is nondegenerate, it means that $\mathbf{H}_{\rho_A}(P)$ is negative definite. By restricting the tangent vectors $(\xi_1,\dots,\xi_r) \in \mathrm{T}_P\mathrm{LG}^{\times}(\mathbf{n})$ to the vectors of the form $(0,\dots,\xi_j,\dots,0)$, it holds

$$\mathbf{H}^j_{\rho_A}(0,\dots,\xi_j,\dots,0) \neq 0, \tag{5.43}$$

for all $0 \neq \xi_j \in \mathrm{T}_{P_j}\mathrm{LG}_{n_j}$, for all $j = 1,\dots,r$. Hence, $\mathrm{ad}_{P_j}\mathrm{ad}_{\widehat{A}_j}\xi_j \neq 0$, for all $0 \neq \xi_j \in \mathrm{T}_{P_j}\mathrm{LG}_{n_j}$, for all $j = 1,\dots,r$. Moreover, by the representation (5.37) of $\overline{\Theta_j}\widehat{A}_j\Theta_j^{\top}$ in a critical point, we obtain

$$\Psi'_j Z_j + Z_j \Psi'_j \neq 0, \tag{5.44}$$

for all $0 \neq Z_j \in \mathfrak{her}_{n_j}$, for all $j = 1,\dots,r$. This means that λ and $-\lambda$ are not both eigenvalues of Ψ'_j, for $j = 1,\dots,r$. $\square$

As a consequence of Theorem 4.3.2, we can conclude that the critical points of the generalized Rayleigh-quotient on $\mathrm{LG}^{\times}(\mathbf{n})$ and on $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$ are nondegenerate for a generic choice of the parameter $A \in \mathfrak{her}_N \cap \mathfrak{sp}^{\otimes}(\mathbf{2n},\mathbb{C})$, where $N := 2n_1 2n_2 \cdots 2n_r$.

Corollary 5.3.8 *For a generic $A \in \mathfrak{her}_N \cap \mathfrak{sp}^{\otimes}(\mathbf{2n},\mathbb{C})$, the critical points of ρ_A on $\mathrm{LG}^{\times}(\mathbf{n})$ are nondegenerate.*

Proof. Let V denote the vector space $\mathfrak{her}_N \cap \mathfrak{sp}^{\otimes}(\mathbf{2n},\mathbb{C})$ and $\mathcal{M} := \mathrm{LG}^{\times}(\mathbf{n})$. Moreover define $F : V \times \mathcal{M} \to \mathrm{T}\mathcal{M}$ as $F(A,P) = \mathrm{grad}\ \rho_A(P)$. We want to show that $F \pitchfork S$, where S is the zero section in $\mathrm{T}\mathcal{M}$, i.e.,

$$\mathrm{Im}\,\mathrm{T}_{(A,P)}F + \mathrm{T}_{F(A,P)}S = \mathrm{T}_{F(A,P)}(\mathrm{T}_P\mathcal{M}),$$

for all $(A,P) \in F^{-1}(S)$. It is enough to prove that $F_P \pitchfork \{0\}$ for all $P \in \mathcal{M}$, where $F_P : V \to \mathrm{T}\mathcal{M},\ F_P(A) = \mathrm{grad}\ \rho_A(P)$, for all $A \in F_P^{-1}(\{0\})$. Hence, we will prove that

$$\mathrm{Im}\,\mathrm{T}_A F_P = \mathrm{T}_P\mathcal{M}, \tag{5.45}$$

for all $A \in F_P^{-1}(\{0\})$ and all $P \in \mathcal{M}$. Since

$$\mathrm{grad}_{\rho\ (\cdot)}(P) : V \to \mathrm{T}_P\mathcal{M},\ X \mapsto \mathrm{grad}\ \rho_X(P)$$

is linear, it follows that

$$\mathrm{T}_A F_P(X) = \mathrm{grad}\ \rho_X(P)$$

and, hence,

$$\mathrm{Im}\,\mathrm{T}_A F_P = \mathrm{Im}\,\mathrm{grad}\ \rho_{(\cdot)}(P).$$

Let $\xi := (\xi_1,\dots,\xi_r) \in (\mathrm{Im}\,\mathrm{grad}\ \rho_{(\cdot)}(P))^{\perp}$, i.e.

$$\langle \mathrm{grad}\ \rho_X(P), \xi\rangle = \mathrm{tr}\Big(X(\sum_{j=1}^{r} P_1 \otimes \cdots \otimes \xi_j \otimes \cdots \otimes P_r)\Big) = 0, \tag{5.46}$$

for all $X \in \mathfrak{sp}^{\otimes}(\mathbf{2n}, \mathbb{R})$ and hence also for X of the form $X = X_1 \otimes \cdots \otimes X_r$ with $X_j \in \mathfrak{sp}(2n_j, \mathbb{C})$. Nw, (5.46) becomes

$$\sum_{j=1}^{r} \operatorname{tr}(X_1 P_1) \cdots \operatorname{tr}(X_j \xi_j) \cdots \operatorname{tr}(X_r P_r) = 0,$$

for all $X_j \in \mathfrak{sp}(2n_j, \mathbb{C})$ and $j = 1, \ldots, r$.
Since $\xi_j \in \mathfrak{sp}(2n_j, \mathbb{C})$, by taking $X_j = \xi_j$ and recalling that $\operatorname{tr}(P_j \xi_j) = 0$, it follows that $\|\xi_j\| = 0$ and hence $\xi_j = 0$, for $j = 1, \ldots, r$. thus, we have proven (5.45). $\square$

From the proof, we obtain that the set of parameters $A \in \mathfrak{her}_N \cap \mathfrak{sp}^{\otimes}(\mathbf{2n}, \mathbb{C})$ for which the critical points of ρ_A are nondegenerate is open and dense and not only residual. A similar result holds also for $\widehat{\rho}_A$ on $\widehat{\mathrm{LG}}^{\times}(\mathbf{n})$.

Chapter 6

Riemannian numerical algorithms

Several problems in Linear Algebra (eigenvalue computation, numerical ranges, low rank matrix approximation), Control Theory (Riccati equation), Robotics (grasping problem), Quantum Control (entanglement measure), Computer Vision (camera estimation, face recognition), etc., can be formulated as optimization tasks with the set of constraints carrying the structure of a Riemannian manifold, see [3, 34]. In this way, the constraint optimization problems can be solved by techniques from Riemannian optimization as unconstrained ones. By replacing the classical objects from numerical optimization (directional derivatives, line search) with their Riemannian counterparts (geodesics, Levi-Civita connection, parallel transport), one obtains efficient methods which perform on the smallest possible parameter space.

In this chapter, we develop two intrinsic methods for the optimization of the generalized Rayleigh-quotient on the $r-$fold direct product of Grassmannians and on the $r-$fold direct product of Lagrangian Grassmannian manifolds: a Newton-like method and a conjugate gradient method. We give suggestions for the implementation of the two methods mentioned as well as a convergence proof in the case of Newton method. Numerical experiments at the end of the chapter confirm the efficiency of the developed algorithms. The above-mentioned algorithms are given in a generalized form with the scope to determine the zeros of a smooth vector field on $\mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ or on $\mathrm{LG}^{\times}(\mathbf{m})$, which does not have to be the gradient vector field of some real-valued function.

6.1 Newton-like method

The intrinsic approach to develop a Newton algorithm for the optimization of a smooth real-valued function on a Riemannian manifold is described by means of the Levi-Civita connection taking iteration steps along geodesics. In particular, if $f : \mathcal{M} \to \mathbb{R}$ is a smooth function and $P^* \in \mathcal{M}$ is a nondegenerated critical point, then the Newton iteration reads as

$$P^{\text{new}} = \exp_P\Big(-(\mathbf{H}_f(P))^{-1}(\text{grad } f(P))\Big), \tag{6.1}$$

where $P \in \mathcal{M}$ is in the neighborhood of P^*. We do not enter into more details, instead we refer the reader to the literature, e.g. [19, 23, 68] and the references therein. As solutions of second-order differential equations, geodesics are sometimes difficult to determine. Even when one knows the geodesic, it is computationally expensive to determine the corresponding Riemannian exponential map. Thus, we are interested in a more general approach due to Shub [65] and further extended [3, 32], which introduces the Newton iteration via local coordinates. More precisely, we follow the ideas introduced by Helmke, Hüper and Trumpf in [32] and use two locally smooth families of parametrizations $\{\mu_P\}_{P\in\mathcal{M}}$ and $\{\nu_P\}_{P\in\mathcal{M}}$ around P^*. Then, the *Newton-like iteration* reads as follows

$$P^{new} = \nu_P\Big(-(\mathbf{H}_{f\circ\mu_P}(0))^{-1}\mathrm{grad}\ f(P)\Big), \tag{6.2}$$

where P is in the neighborhood of P^* and $\mathbf{H}_{f\circ\mu_P^{\exp}}(P)$ is the Hessian of the function $f\circ\mu_P : \mathrm{T}_P\mathcal{M} \to \mathbb{R}$ at 0. In this way, one can develop families of numerical methods with quadratic convergence as the intrinsic Newton method.

Further, let $\mathcal{M}$ denote either one of the manifolds $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ or $\mathrm{LG}^\times(\mathbf{m})$, with the specification that in the case of the Lagrangian Grassmann manifold, n_j is even and $m_j = n_j/2$ for all $j = 1,\dots,r$. As before $N = n_1n_2\cdots n_r$, and $(\mathbf{m},\mathbf{n})$ and $(\mathbf{m})$ denote the multi-index

$$(\mathbf{m},\mathbf{n}) := \Big((m_1,n_1),\dots,(m_r,n_r)\Big), \quad \text{resp. } (\mathbf{m}) := (m_1,\dots,m_r).$$

In view of applications, we will use Riemannian normal coordinates and QR-coordinates as parametrizations around $P \in \mathcal{M}$. These notions of coordinates around a point $P := (P_1,\dots,P_r) \in \mathcal{M}$ can be immediately generalized from one Grassmannian (resp. Lagrange-Grassmann manifold) to the $r-$fold direct product of Grassmannians (resp. direct product of Lagrange-Grassmann manifolds) as follows: *Riemannian normal coordinates* are given by the Riemannian exponential map

$$\mu_P^{\exp} : \mathrm{T}_P\mathcal{M} \to \mathcal{M}, \quad \xi \mapsto (e^{[\xi_1,P_1]}P_1e^{-[\xi_1,P_1]},\dots,e^{[\xi_r,P_r]}P_re^{-[\xi_r,P_r]}), \tag{6.3}$$

while *QR-type coordinates* are defined by the QR-approximation of the matrix exponential, i.e.

$$\mu_P^{\mathrm{QR}} : \mathrm{T}_P\mathcal{M} \to \mathcal{M}, \quad \xi \mapsto ([X_1]_Q\ P_1\ [X_1]_Q^\dagger,\dots,[X_r]_Q\ P_r\ [X_r]_Q^\dagger), \tag{6.4}$$

where $[X_j]_Q$ is the $Q-$factor from the unique QR decomposition of $X_j := I + [\xi_j, P_j]$.

In what follows, we will give the Newton-like method to determine the stationary points of the generalized Rayleigh-quotient of a Hermitian matrix $A \in \mathfrak{her}_N$

$$\rho_A : \mathcal{M} \to \mathbb{R}, \quad \rho_A(P) = \mathrm{tr}\Big(A(P_1\otimes\cdots\otimes P_r)\Big). \tag{6.5}$$

6.1.1 Newton-like algorithm on $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$

The Newton-like iteration (6.2) for the generalized Rayleigh-quotient of a Hermitian matrix A is stated as

$$P^{\text{new}} = \mu_P^{QR}(\xi), \tag{6.6}$$

where $\xi \in \mathrm{T}_P\mathcal{M}$ is the solution of the equation

$$\mathbf{H}_{\rho_A}(P)\xi = -\mathrm{grad}\ \rho_A(P). \tag{6.7}$$

For an explicit expression of the Newton direction ξ, recall first the multilinear maps $\Psi_{A,j} : \mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r} \to \mathbb{C}^{n_j \times n_j}$ defined by

$$\mathrm{tr}\Big(A^\dagger(X_1 \otimes \cdots \otimes X_j Z \otimes \cdots X_r)\Big) = \mathrm{tr}\Big(\Psi_{A,j}(X_1, \ldots, I_{n_j}, \ldots, X_r)^\dagger Z\Big),$$

for all $Z \in \mathbb{C}^{n_j \times n_j}$ and $j = 1, \ldots, r$. To simplify the exposure, denote

$$\widehat{A}_j := \Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r) \tag{6.8}$$

and remind that $\widehat{A}_j$ is Hermitian for all $P \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ and $j = 1, \ldots, r$.
Replacing the objects in (6.7) by their explicit form computed in Chapter 3, we get the following:

$$-\mathrm{ad}_{P_j}\mathrm{ad}_{\widehat{A}_j}\xi_j + \sum_{k=1,k\neq j}^{r} \mathrm{ad}_{P_j}^2 \Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, \xi_k, \ldots, P_r) = -\mathrm{ad}_{P_j}^2 \widehat{A}_j, \tag{6.9}$$

for all $j = 1, \ldots, r$. Solving the system (6.9) in the embedding space $\mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r}$ requires $2(n_1^2 + \cdots + n_r^2)$ parameters. By exploiting the particular structure of the tangent vectors

$$\xi_j = \Theta_j^\dagger \zeta_j \Theta_j = \Theta_j^\dagger \begin{bmatrix} 0 & Z_j \\ Z_j^\dagger & 0 \end{bmatrix} \Theta_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}, \tag{6.10}$$

where $\Theta_j \in \mathrm{SU}_{n_j}$ and $Z_j \in \mathbb{C}^{m_j \times (n_j - m_j)}$, the number of parameters is reduced to the dimension of the tangent space $\mathrm{T}_P\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$, i.e. $2(m_1(n_1 - m_1) + \cdots + m_r(n_r - m_r))$.
Multiplying the system (6.9) from the left with Θ_j and from the right with $\Theta_j^\dagger$, we reduce it to a system in $Z_j \in \mathbb{C}^{m_j \times (n_j - m_j)}$, i.e.

$$\Psi_j' Z_j - Z_j \Psi_j'' - \sum_{k=1,k\neq j}^{r} \Phi_j(Z_k) = \Psi_j'''. \tag{6.11}$$

An explicit form for Ψ_j', Ψ_j'', Ψ_j''' and $\Phi_j(Z_k)$ is given in what follows.

Let

$$\Theta_j = \begin{bmatrix} U_j \\ V_j \end{bmatrix} \in \mathrm{SU}_{n_j}, \tag{6.12}$$

where U_j and V_j are $m_j \times n_j$ and $(n_j - m_j) \times n_j$ matrices, respectively. Then,

$$\Psi_j' = U_j \widehat{A}_j U_j^\dagger, \quad \Psi_j'' = V_j \widehat{A}_j V_j^\dagger, \quad \Psi_j''' = U_j \widehat{A}_j V_j^\dagger. \tag{6.13}$$

For expressing $\Phi_j(Z_k)$ with $j < k$, we introduce the multilinear operators

$$\Psi_{A,j,k} : \mathfrak{her}_{n_1} \times \cdots \times \mathfrak{her}_{n_r} \to \mathbb{C}^{n_j \cdot n_k \times n_j \cdot n_k}$$

defined in a similar way as $\Psi_{A,j}$ by

$$\operatorname{tr}\Big(A(X_1 \otimes \cdots \otimes X_j S \otimes \cdots \otimes X_k T \otimes \cdots \otimes X_r)\Big) = \operatorname{tr}\Big(\Psi_{A,j,k}(X_1, \ldots, X_r)(S \otimes T)\Big), \tag{6.14}$$

for all $S \in \mathbb{C}^{n_j \times n_j}$ and $T \in \mathbb{C}^{n_k \times n_k}$.
For convenience, we will use the following shortcut

$$\widehat{A}_{jk} := \Psi_{A,j,k}(P_1, \ldots, I_{n_j}, \ldots, I_{n_k}, \ldots, P_r) \tag{6.15}$$

and can argue similarly to $\widehat{A}_j$ that $\widehat{A}_{jk}$ is a Hermitian matrix of size $n_j n_k \times n_j n_k$. Furthermore, we partition the matrix $\widehat{A}_{jk}$ into block form

$$\widehat{A}_{jk} = \left[\begin{array}{c|c|c|c} \widehat{\mathbf{a}}_{11} & \widehat{\mathbf{a}}_{12} & \cdots & \widehat{\mathbf{a}}_{1n_j} \\ \hline \widehat{\mathbf{a}}_{12}^\dagger & \widehat{\mathbf{a}}_{22} & \cdots & \widehat{\mathbf{a}}_{2n_j} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \widehat{\mathbf{a}}_{1n_j}^\dagger & \widehat{\mathbf{a}}_{2n_j}^\dagger & \cdots & \widehat{\mathbf{a}}_{n_j n_j} \end{array}\right], \tag{6.16}$$

where each $\widehat{\mathbf{a}}_{st}$ is an $n_k \times n_k$ matrix.
Then, the linear map $\Phi_j : \mathbb{C}^{m_k \times (n_k - m_k)} \to \mathbb{C}^{m_j \times (n_j - m_j)}$ is given by

$$Z_k \mapsto \Phi_j(Z_k) = U_j \Big[\operatorname{tr}(U_k \widehat{\mathbf{a}}_{st} V_k^\dagger Z_k^\dagger + Z_k V_k \widehat{\mathbf{a}}_{st}^\dagger U_k^\dagger) \Big]_{s,t=1}^{n_j} V_j^\dagger. \tag{6.17}$$

The squared norm of the gradient of ρ_A at P can be expressed as

$$\|\text{grad } \rho_A(P)\|^2 = 2 \sum_{j=1}^{r} \|\Psi_j'''\|^2 \tag{6.18}$$

and hence, the complete Newton-like algorithm for the optimization of ρ_A on the Riemannian manifold $\mathrm{Gr}^\times(\mathbf{m}, \mathbf{n})$ is given in Table 6.1.

The Newton-like algorithm can be generalized to determine the zeros of the vector field

$$\begin{aligned} \mathcal{X}_A : \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n}) &\to \mathrm{TGr}^{\times}(\mathbf{m},\mathbf{n}), \\ (P_1,\dots,P_r) &\mapsto ([\Omega_1,P_1]+[P_1,[P_1,S_1]],\dots,[\Omega_r,P_r]+[P_r,[P_r,S_r]]), \end{aligned} \tag{6.19}$$

where $A \in \mathbb{C}^{N\times N}$ is not necessarily Hermitian. Here, S_j and Ω_j denote the Hermitian and respectively the skew-Hermitian part of $\widehat{A}_j$, for $j = 1,\dots,r$. In this case, the Newton-like update is given as

$$P^{\mathrm{new}} = \mu_P^{\mathrm{QR}}(\xi) \quad \text{with} \quad \boldsymbol{\nabla}_\xi \mathcal{X}_A(P) = -\mathcal{X}_A(P).$$

From the formula (2.3) which defines the Levi-Civita connection on a submanifold, the Newton equation is writen in an implicit form as

$$\mathrm{ad}_{P_j}^2 \mathrm{ad}_{\Omega_j}\xi_j - \mathrm{ad}_{P_j}\mathrm{ad}_{S_j}\xi_j + \sum_{k=1,k\neq j}^{r} \mathrm{ad}_{P_j}^2 \widehat{S}_{jk} = \mathrm{ad}_{P_j}\Omega_j - \mathrm{ad}_{P_j}^2 S_j,$$

where $\widehat{S}_{jk}$ is the Hermitian part of $\Psi_{A,j}(P_1,\dots,I_{n_j},\dots,\xi_k,\dots,P_r)$. Following the exact steps as in the case of the Newton-like method for the optimization of the generalized Rayleigh-quotient, one gets

$$\Psi_j' Z_j - Z_j \Psi_j'' - \sum_{k=1,k\neq j}^{r} \frac{\Phi_j(Z_k) + \Lambda_j(Z_k)^\dagger}{2} = -\Psi_j''''$$

where $\Phi_j(Z_k)$ and $\Psi_j', \Psi_j'', \Psi_j''', \Psi_j''''$ are determined as in (6.17) and

$$\Theta_j^\dagger \widehat{A}_j \Theta_j = \begin{bmatrix} \Psi_j' & \Psi_j''' \\ & \\ \Psi_j'''' & \Psi_j'' \end{bmatrix}. \tag{6.20}$$

Moreover, $\Lambda_j(Z_k)$ is given similar to $\Phi_j(Z_k)$ by

$$Z_k \mapsto \Phi_j(Z_k) = V_j \left[\mathrm{tr}(U_k \widehat{\mathbf{a}}_{st} V_k^\dagger Z_k^\dagger + Z_k V_k \widehat{\mathbf{a}}_{st}^\dagger U_k^\dagger) \right]_{s,t=1}^{n_j} U_j^\dagger.$$

ALGORITHM 1. **N-like algorithm for the optimization of ρ_A on $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$**

Step 1. Starting point: Given $P = (P_1, \dots, P_r) \in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ choose

$$\Theta_j = \begin{bmatrix} U_j \\ \\ V_j \end{bmatrix} \in \mathrm{SU}_{n_j},\ U_j U_j^\dagger = I_{m_j},\ V_j V_j^\dagger = I_{n_j - m_j},$$

such that $P_j = \Theta_j^\dagger \Pi_j \Theta_j$, for $j = 1, \dots, r$.

Step 2. Stopping criterion: If $\|\mathrm{grad}_{\rho_A}(P)\|/\rho_A(P) < \mathrm{err}$ then STOP.

Step 3. Newton direction: Set

$$\widehat{A}_j := \Psi_{A,j}(P_1, \dots, I_{n_j}, \dots, P_r)$$

and compute $\Psi_j',\ \Psi_j'',\ \Psi_j'''$ as in (6.13), for $j = 1, \dots, r$. Set

$$\widehat{A}_{jk} := \Psi_{A,j,k}(P_1, \dots, I_{n_j}, \dots, I_{n_k}, \dots, P_r)$$

and compute $\Phi_j(Z_k)$ as in (6.17), for $j,\ k = 1, \dots, r$, with $j < k$. With $\Phi_k(Z_j) = \Phi_j(Z_k)^\dagger$ solve the Newton equation

$$\Psi_j' Z_j - Z_j \Psi_j'' - \sum_{k=1, k\neq j}^{r} \Phi_j(Z_k) = \Psi_j''',$$

to obtain $Z_j \in \mathbb{C}^{m_j \times (n_j - m_j)}$, for $j = 1, \dots, r$.

Step 4. QR-updates:

$$\Theta_j^{new} = \begin{bmatrix} I_{m_j} & -Z_j \\ \\ Z_j^\dagger & I_{n_j - m_j} \end{bmatrix}_Q \Theta_j \quad \text{and} \quad P_j^{new} = \Theta_j^{new\dagger} \Pi_j \Theta_j^{new}, \tag{6.21}$$

for all $j = 1, \dots, r$. Here $[\ ]_Q$ refers to the Q part from the QR factorization.

Step 5. Set $P := P^{\mathrm{new}}$, $\Theta := \Theta^{\mathrm{new}}$ and return to Step 2.

Table 6.1:

6.1.2 Convergence proof and suggestions for implementation.

It was proved in [32] that under the assumption

$$D\mu_P(0) = D\nu_P(0), \tag{6.22}$$

on the two families of local parametrization μ_P and ν_P, the Newton-like iteration (6.2) converges quadratically to a critical point of the cost function, when it starts in a sufficiently small neighborhood of this critical point. Since in the case when $\mu_P = \mu_P^{\exp}$ and $\nu_P = \mu_P^{QR}$ the condition (6.22) holds (we refer to [32] for the proof), we can state the following convergence result and mention the literature [32, 40] for the proof.

Theorem 6.1.1 *Let $A \in \mathfrak{her}_N$ and $P^* \in \mathcal{M}$ be a nondegenerate critical point of the generalized Rayleigh-quotient ρ_A, then the sequence generated by the N-like algorithm (Table 6.1) converges locally quadratically to P^*.*

Keeping in mind the objects and their definition from the previous section, we will give some suggestions regarding the computational aspects of the Newton-like algorithm. The convergence of the Newton-like method (6.2) is not guaranteed for arbitrary starting points and even in the case of convergence the limiting point need not be a local maximizer or local minimizer of the cost function. How could one overcome this, in particular in the case of maximizing ρ_A, where there is no closed form characterization of the stationary points of ρ_A available? A classical approach in the literature combines Newton steps with gradient steps, see [25]. More precisely, one starts in an arbitrary point and verifies if the Newton direction is ascending, else takes the gradient as the new direction. Furthermore, one can make an iterative line-search in the ascending direction to guarantee convergence to a local maximizer. Our numerical experiments have proved that this is a lengthy procedure and not recommended for large size problems. Instead of the gradient direction, we propose an alternating least square method, i.e. at each step we perform the following optimization task to obtain a new update $(P_1, \dots, P_r) \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$

$$\max_{P_j \in \mathrm{Gr}_{m_j, n_j}} \operatorname{tr}(\widehat{A}_j P_j), \tag{6.23}$$

for all $j = 1, \dots, r$. This procedure is a generalization of the well-known Power method [27] use to determine the biggest eigenvalue of a matrix, and is immediately solved by a truncated singular value decomposition (SVD) of $\widehat{A}_j$. It was proved by Golub and Zuhang in [79] that for the best rank-one tensor approximation problem, the HOOI method converges globally with at most linear rate to a critical point of ρ_A. In general, being an alternating least square method, it converges locally, see [6, 58, 62].

Further, for given $A \in \mathfrak{her}_N$ and fixed $P \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ we suggest methods that efficiently compute $\widehat{A}_j$ and $\widehat{A}_{jk}$ given by (6.8) and (6.15) respectively. In the general case, the computation of $\widehat{A}_j$ and $\widehat{A}_{jk}$ is performed according to formula (4.9), but this can be notably simplified when we tackle the specified applications of the generalized

Rayleigh-quotient on $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$, i.e. best low-rank tensor approximation and subspace clustering problem.

Best low-rank tensor approximation. In the first case, the matrix $A \in \mathfrak{her}_N$ is of rank -1, i.e. $A = vv^{\dagger}$, with $v = \mathrm{vec}(\mathcal{A})$ and $\mathcal{A} \in \mathbb{C}^{n_1 \times \cdots \times n_r}$. From (4.25) and the definition of matrix unfolding along the $j-$th direction, it follows that

$$(\mathcal{A} \times_1 U_1 \times_2 \cdots \times_r U_r)_{(j)} = U_j(\mathcal{A} \times_1 U_1 \times_2 \cdots \times_j I_{n_j} \times_{j+1} \cdots \times_r U_r)_{(j)}, \tag{6.24}$$

where U_j are given in (6.12), for $j = 1, \ldots, r$.
Recall that $A_{(j)}$ refers to the matrix unfolding along direction j of a tensor $\mathcal{A}$.
The Frobenius norm of a tensor (4.29) has the obvious property

$$\|\mathcal{A}\|^2 = \|A_{(1)}\|^2 = \cdots = \|A_{(r)}\|^2. \tag{6.25}$$

By denoting $\mathcal{B} := \mathcal{A} \times_1 U_1 \times_2 \cdots \times_j I_{n_j} \times_{j+1} \cdots \times_r U_r$, from (6.24) and (6.25) it holds

$$\rho_A(P) = \|\mathcal{A} \times_1 U_1 \times_2 \cdots \times_r U_r\|^2 = \|\mathcal{B} \times_j U_j\|^2 = \|U_j B_{(j)}\|^2 = \mathrm{tr}\Big(B_{(j)} \cdot B_{(j)}^{\dagger} P_j\Big),$$

for $j = 1, \ldots, r$. Since $\rho_A(P) = \mathrm{tr}(\widehat{A}_j P_j)$ it follows that we can take

$$\widehat{A}_j = B_{(j)} \cdot B_{(j)}^{\dagger} \in \mathfrak{her}_{n_j}, \tag{6.26}$$

for $j = 1, \ldots, r$.

Similar, by denoting $\mathcal{C} := \mathcal{A} \times_1 U_1 \times_2 \cdots \times_j I_{n_j} \times_{j+1} \cdots \times_k I_{n_k} \times_{k+1} \cdots \times_r U_r$ and recalling that $C_{(j,k)}$ stands for the $(j,k)-$th mode matrix of $\mathcal{C}$,

$$\widehat{A}_{jk} = C_{(j,k)} \cdot C_{(j,k)}^{\dagger} \in \mathfrak{her}_{n_j n_k}, \tag{6.27}$$

for $j, k = 1, \ldots, r$ with $j \neq k$.

Remark 6.1.2 *In a more general framework, if $A \in \mathfrak{her}_N$ is semipositive definite of a certain rank K considerably smaller than N, one can use iteratively the same procedure as in the case of rank-1 matrices to compute $\widehat{A}_j$ and $\widehat{A}_{jk}$. Explicitly, there exists a matrix $X \in \mathbb{C}^{N \times K}$ such that $A = XX^{\dagger}$. If x_k denote the columns of X, then $A = x_1 x_1^{\dagger} + \cdots + x_K x_K^{\dagger}$ and hence $\rho_A(P) = \rho_{x_1 x_1^{\dagger}}(P) + \cdots + \rho_{x_K x_K^{\dagger}}(P)$.*

Subspace clustering. For the problem of recovering subspaces from data points $x_l \in \mathbb{C}^n$ with given $L \in \mathbb{N}$, the matrix A has the form

$$A = \sum_{l=1}^{L} \underbrace{x_l x_l^{\dagger} \otimes \cdots \otimes x_l x_l^{\dagger}}_{r \text{ times}}.$$

As in the previous application, one doesn't work directly with the matrix A, but with the data points x_l. In particular, from the properties of the *trace* function and matrix Kronecker product one has

$$\widehat{A}_j = \sum_{l=1}^{L} \Big(\prod_{\substack{i=1\\ i\neq j}}^{r} \|P_i x_l\|^2 \Big) x_l x_l^\dagger \tag{6.28}$$

and

$$\widehat{A}_{jk} = \sum_{l=1}^{L} \Big(\prod_{\substack{i=1\\ i\neq j, i\neq k}}^{r} \|P_i x_l\|^2 \Big) x_l x_l^\dagger \otimes x_l x_l^\dagger. \tag{6.29}$$

Low cost QR-update. The computation of geodesics on matrix manifolds usually requires the matrix exponential map, which is in general an expensive procedure of order $O(n^3)$. Yet, for the particular case of the Grassmann manifold $\mathrm{Gr}_{m,n}$, Gallivan et.al. [24] have developed an efficient method to compute the matrix exponential, reducing the complexity order to $O(nm^2)$ $(m < n)$. Our approach, however, is based on a first order approximation of the matrix exponential $\mathrm{e}^{[\zeta,\Pi]}$ followed by a QR-decomposition to preserve orthogonality/unitarity. Let X denote the matrix $I_n + [\zeta, \Pi]$, i.e.

$$X := \begin{bmatrix} I_m & -Z \\ Z^\dagger & I_{n-m} \end{bmatrix} \tag{6.30}$$

and $X = (X)_Q(X)_R$ be the $QR-$decomposition of X and $Z = U\Sigma V^\dagger$ with $U \in \mathrm{SU}_m$, $V \in \mathbb{C}^{(n-m)\times m}$, $V^\dagger V = I_m$ and $\Sigma \in \mathbb{C}^{m\times m}$ diagonal, the singular value decomposition of Z. Since

$$\begin{aligned} XX^\dagger = X^\dagger X &= (X)_Q(X)_R(X)_R^\dagger(X)_Q^\dagger = (X)_R^\dagger(X)_R \\ &= \begin{bmatrix} I_m + ZZ^\dagger & 0 \\ 0 & I_{n-m} + Z^\dagger Z \end{bmatrix}, \end{aligned}$$

by the singular value decomposition of Z it follows that

$$XX^\dagger = W \begin{bmatrix} I_m + \Sigma\Sigma^\dagger & 0 & 0 \\ 0 & I_m + \Sigma^\dagger\Sigma & 0 \\ 0 & 0 & I_{n-2m} \end{bmatrix} W^\dagger,$$

where

$$W := \begin{bmatrix} U^\dagger & 0 & 0 \\ 0 & V & V' \end{bmatrix} \in \mathrm{SU}_n, \quad D := \sqrt{I_m + \Sigma^\dagger\Sigma}, \tag{6.31}$$

and $[V \ \ V'] \in \mathrm{SU}_{n-m}$ is an unitary completion of V. Hence, $(X)_Q$ is explicitly given by

$$(X)_Q = W \begin{bmatrix} D^{-1} & \Sigma D^{-1} & 0 \\ -\Sigma^\dagger D^{-1} & D^{-1} & 0 \\ 0 & 0 & I_{n-2m} \end{bmatrix} W^\dagger. \tag{6.32}$$

The computational complexity of this QR-factorization is of order $O((n-m)m^2)$. To solve the system (6.11), one can rewrite it as a linear equation on $\mathbb{R}^d$ (d is the dimension of $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$) using matrix Kronecker products and vec−operations, then solve this by any linear equation solver.

6.1.3 Newton-like algorithm on $\mathrm{LG}^\times(\mathbf{m})$

We give the Newton-like iteration for the optimization of the generalized Rayleigh-quotient of a matrix $A \in \mathfrak{her}_N$ on the direct product of complex Lagrangian Grassmannian manifolds $\mathrm{LG}^\times(\mathbf{m})$ and we mention that it can be similarly given for the classical cases of direct product of Lagrangian Grassmannian manifolds. As a submanifold of $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$, the Newton-like algorithm on $\mathrm{LG}^\times(\mathbf{m})$ is a simplified version of the Newton-like algorithm on $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$.

Let $A \in \mathfrak{her}_N$ and let $P^* \in \mathrm{LG}^\times(\mathbf{m})$ be a nondegenerated stationary point of ρ_A on $\mathrm{LG}^\times(\mathbf{m})$. For $P \in \mathrm{LG}^\times(\mathbf{m})$ in the neighborhood of P^*, the Newton-like iteration is given as

$$P^{\mathrm{new}} = \mu_P^{\mathrm{QR}}\Big(-(\pi_P \mathbf{H}_{\rho_A}(P))^{-1}\pi_P \mathrm{grad}\ \rho_A(P)\Big), \tag{6.33}$$

where grad $\rho_A(P)$ and $\mathbf{H}_{\rho_A}(P))$ are the gradient and respective the Hessian of ρ_A at P on $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ and

$$\begin{aligned} \pi_P : \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n}) &\to \mathrm{T}_P\mathrm{Gr}^\times(\mathbf{m},\mathbf{n}), \\ (\xi_1,\dots,\xi_r) &\mapsto \left(\frac{\xi_1 + J_1\xi_1J_1}{2},\dots,\frac{\xi_r + J_r\xi_rJ_r}{2}\right) \end{aligned} \tag{6.34}$$

is the orthogonal projector onto $\mathrm{T}_P\mathrm{LG}^\times(\mathbf{m})$. Recall from Chapter 4 that J_j is the standard symplectic form on $\mathbb{C}^{n_j}$, for $j = 1,\dots,r$. As before $\widehat{A}_j$ is the shortcut for $\Psi_{A,j}(P_1,\dots,I_{n_j},\dots,P_r)$ and $X_\mathfrak{h}$ is the Hamiltonian part of $X \in \mathfrak{her}_{n_j}$, i.e. $X_\mathfrak{h} := X + J_jXJ_j)/2$. Then, the Newton equation

$$\pi_P(\mathbf{H}_{\rho_A}(P))\xi = -\pi_P \mathrm{grad}\ \rho_A(P)$$

becomes

$$-\mathrm{ad}_{P_j}\mathrm{ad}_{(\widehat{A}_j)_\mathfrak{h}}\xi_j + \sum_{k=1,k\neq j}^{r} \mathrm{ad}_{P_j}^2\Big(\Psi_{A,j}(P_1,\dots,X_j,\dots,\xi_k,\dots,P_r)\Big)_\mathfrak{h} = -\mathrm{ad}_{P_j}^2(\widehat{A}_j)_\mathfrak{h},$$

which reduces to the task of determining Hermitian solutions $Z_j \in \mathfrak{her}_{m_j}$ of

$$(\Psi_j' - \Psi_j'')Z_j - Z_j(\Psi_j'' - \Psi_j') - \sum_{k=1,k\neq j}^{r} \Phi_j(Z_k) + (\Phi_j(Z_k))^\dagger = \Psi_j''' + (\Psi_j''')^\dagger, \qquad (6.35)$$

for $j = 1, \ldots, r$. The terms $\Psi_j', \Psi_j'', \Psi_j'''$ and $\Phi_j(Z_k)$ are computed by (6.13) and (6.17) respectively. The Newton-like algorithm for the optimization of ρ_A on the Riemannian manifold $\mathrm{LG}^\times(\mathbf{m})$ is given in Table 6.2.
For the generalized Lagrange-Grassmannian LG_n, a computational cheap $QR-$update is given by

$$\begin{bmatrix} I_m & -Z \\ Z & I_m \end{bmatrix}_Q = W \begin{bmatrix} D^{-1} & \Sigma D^{-1} \\ -\Sigma^\dagger D^{-1} & D^{-1} \end{bmatrix} W^\dagger, \qquad (6.36)$$

where

$$W := \begin{bmatrix} U^\dagger & 0 \\ 0 & U \end{bmatrix} \in \mathrm{SU}_n, \quad D := \sqrt{I_m + \Sigma^2}, \qquad (6.37)$$

and $U \in \mathrm{SU}_m$ gives the singular value decomposition $Z = U\Sigma U^\dagger$ with $\Sigma \in \mathbb{R}^{m\times m}$ diagonal. In this case, the numerical complexity is given only by the singular value decomposition of Z and hence is of order $O(m^3)$, where $m = n/2$.

The convergence of the sequence generated by the Newton-like iteration (6.33) converges quadratically to a critical point of ρ_A when starting in the neighborhood of that point. For the proof, see the literature [32]. The question is how to get in the neighborhood of a critical point of ρ_A? One idea would be to combine the Newton-like direction with a steepest ascent direction, or one could try an alternating least square approach as in the case of the Grassmann manifold. However, for the optimization (6.23) on the Lagrange-Grassmannian manifold, presently there is no closed form solution for the maximizers.

ALGORITHM 1'. **N-like algorithm for the optimization of ρ_A on $\mathrm{LG}^{\times}(\mathbf{m})$**

Step 1. Starting point: Given $P = (P_1, \ldots, P_r) \in \mathrm{LG}^{\times}(\mathbf{m})$ choose

$$\Theta_j = \begin{bmatrix} U_j \\ \\ V_j \end{bmatrix} \in \widehat{\mathrm{Sp}}(m_j),\ U_j U_j^{\dagger} = I_{m_j},\ V_j V_j^{\dagger} = I_{n_j - m_j},$$

such that $P_j = \Theta_j^{\dagger} \Pi_j \Theta_j$, for $j = 1, \ldots, r$.

Step 2. Stopping criterion: $\|\mathrm{grad}_{\rho_A}(P)\| / \rho_A(P) < \varepsilon$.

Step 3. Newton direction: Set

$$\widehat{A}_j := \Psi_{A,j}(P_1, \ldots, I_{n_j}, \ldots, P_r), \quad (\widehat{A}_j)_{\mathfrak{h}} := J_j \widehat{A}_j J_j + \widehat{A}_j \tag{6.38}$$

and compute Ψ_j', Ψ_j'', Ψ_j''' as in (6.13), for $j = 1, \ldots, r$.
Set

$$\widehat{A}_{jk} := \Psi_{A,j,k}(P_1, \ldots, I_{n_j}, \ldots, I_{n_k}, \ldots, P_r)$$

and compute $\Phi_j(Z_k)$ as in (6.16) and (6.17), for $j,\ k = 1, \ldots, r$, with $j < k$. Solve the Newton equation

$$(\Psi_j' - \Psi_j'')Z_j - Z_j(\Psi_j'' - \Psi_j') - \sum_{k=1, k\neq j}^{r} \Phi_j(Z_k) + (\Phi_j(Z_k))^{\dagger} = \Psi_j''' + (\Psi_j''')^{\dagger}, \tag{6.39}$$

to obtain $Z_j \in \mathfrak{her}_{n_j}$, for $j = 1, \ldots, r$.

Step 4. QR-updates:

$$\Theta_j^{\mathrm{new}} = \Theta_j \begin{bmatrix} I_{m_j} & -Z_j \\ \\ Z_j & I_{n_j - m_j} \end{bmatrix}_Q \quad \text{and} \quad P_j^{\mathrm{new}} = \Theta_j \Pi_j \Theta_j^{\mathrm{new}\dagger} \tag{6.40}$$

for all $j = 1, \ldots, r$. Here $[\,]_Q$ refers to the Q part from the QR factorization.

Step 5. Set $P := P^{\mathrm{new}}$, $\Theta := \Theta^{\mathrm{new}}$ and go to Step 2.

Table 6.2:

6.2 Riemannian conjugate gradient algorithm

The quadratic convergence of the Newton-like algorithm has the drawback of high computational complexity. Solving the Newton equation (6.11) (resp. (6.35)) yields a cost per iteration of order $O(d^3)$, where d is the dimension of $\mathrm{Gr}^\times(\mathbf{m},\mathbf{n})$ (resp. $\mathrm{LG}^\times(\mathbf{m})$). In what follows, we offer as an alternative to reduce the computational costs of the Newton-like algorithm by a conjugate gradient method. The linear conjugate gradient (LCG) method is used for solving large systems of linear equations with a symmetric positive definite matrix, which is achieved by iteratively minimizing a convex quadratic function $x^\dagger Ax$. The initial direction d_0 is chosen as the steepest descent and every forthcoming direction d_j is required to be conjugate to all the previous ones, i.e. $d_j^\dagger Ad_k = 0$, for all $k = 0, \cdots, j-1$, [21, 25, 64]. The exact maximum along a direction gives the next iterate. Hence, the optimal solution is found in at most n steps, where n is the dimension of the problem. Nonlinear conjugate gradient (NCG) methods use the same approach for general functions $f : \mathbb{R}^n \to \mathbb{R}$, not necessarily convex and quadratic.

The update rule reads as

$$x^{\text{new}} = x + \alpha d \quad \text{and} \quad d^{\text{new}} = -\nabla f(x^{\text{new}}) + \beta d, \tag{6.41}$$

where the step-size α is obtained by a line search in the direction d

$$\alpha = \arg\min_t f(x + td) \tag{6.42}$$

and β is given by one of the formulas:

$$\text{Fletcher} - \text{Reeves}: \quad \beta^{\text{FR}} := \frac{\nabla f(x^{\text{new}})^T \nabla f(x^{\text{new}})}{\nabla f(x)^T \nabla f(x)},$$

$$\text{Polak} - \text{Ribiere}: \quad \beta^{\text{PR}} := \frac{\nabla f(x^{\text{new}})^T (\nabla f(x^{\text{new}}) - \nabla f(x))}{\nabla f(x)^T \nabla f(x)},$$

$$\text{Hestenes} - \text{Stiefel}: \quad \beta^{\text{HS}} := \frac{\nabla f(x^{\text{new}})^T (\nabla f(x^{\text{new}}) - \nabla f(x))}{d^T (\nabla f(x^{\text{new}}) - \nabla f(x))},$$

or others.

Table 6.3: The Euclidean conjugate gradient method for the optimization of a function $f : \mathbb{R}^n \to \mathbb{R}$, not necessarily convex and quadratic.

We refer to [68] for the generalization of the NCG method to a Riemannian manifold. For the computation of the step-size along the geodesic in direction ξ, an exact line search — as in the classical case — is an extremely expensive procedure. Therefore, one commonly approximates (6.42) by an Armijo-rule, which ensures at least that the step length decreases the function sufficiently. We, however, have decided to compute the step-size by performing a one-dimensional Newton-step along the geodesic, since in the neighborhood of a critical point one Newton step can lead very close to the solution. Starting at a point $P \in \mathcal{M}$ and moving in the direction $\xi \in \mathrm{T}_P\mathcal{M}$ we arrive in $P^{\text{new}} \in \mathcal{M}$. To compute a new direction, a "transport" of the old direction ξ from $\mathrm{T}_P\mathcal{M}$ to the tangent space $\mathrm{T}_{P^{\text{new}}}\mathcal{M}$ is required. We denote this transport with $\tau\xi$ and we refer to (2.5) for an explicit formula.

In what follows, we describe the Riemannian conjugate gradient for the maximization of the generalized Rayleigh-quotient of a Hermitian matrix $A \in \mathfrak{her}_N$ on $\mathcal{M} := \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ or $\mathcal{M} := \mathrm{LG}^{\times}(\mathbf{m})$.

Choose $P \in \mathcal{M}$ in a neighborhood of P^*, denote $\xi := \text{grad } \rho_A(P)$ and perform the following update

$$P^{\text{new}} := \mu_P^{\text{QR}}(\alpha\xi). \tag{6.43}$$

The step-size α is computed as

$$\alpha = -\frac{(\rho_A \circ \gamma)'(0)}{(\rho_A \circ \gamma)''(0)}, \tag{6.44}$$

where $\gamma : I \to \mathcal{M}$ is the unique geodesic through P in direction ξ. The new direction $\xi^{\text{new}} \in \mathrm{T}_{P^{\text{new}}}\mathcal{M}$ is given by

$$\xi^{\text{new}} = \text{grad } \rho_A(P^{\text{new}}) + \beta\tau\xi \tag{6.45}$$

with

$$\begin{aligned} \text{Fletcher} - \text{Reeves}: \quad & \beta^{\text{FR}} := \frac{\langle \text{grad } \rho_A(P^{\text{new}}), \text{grad } \rho_A(P^{\text{new}}) \rangle}{\langle \text{grad } \rho_A(P), \text{grad } \rho_A(P) \rangle}, \\ \text{Polak} - \text{Ribiere}: \quad & \beta^{\text{PR}} := \frac{\langle \text{grad } \rho_A(P^{\text{new}}), \text{grad } \rho_A(P^{\text{new}}) - \tau\text{grad } \rho_A(P) \rangle}{\langle \text{grad } \rho_A(P), \text{grad } \rho_A(P) \rangle}, \\ \text{Hestenes} - \text{Stiefel}: \quad & \beta^{\text{HS}} := \frac{\langle \text{grad } \rho_A(P^{\text{new}}), \text{grad } \rho_A(P^{\text{new}}) - \tau\text{grad } \rho_A(P) \rangle}{\langle \text{grad } \rho_A(P), \text{grad } \rho_A(P^{\text{new}}) - \tau\text{grad } \rho_A(P) \rangle}. \end{aligned} \tag{6.46}$$

Table 6.4: Riemannian conjugate gradient algorithm for the maximization of ρ_A on $\mathcal{M} := \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ or $\mathcal{M} := \mathrm{LG}^{\times}(\mathbf{m})$.

The above algorithm can be generalized to the task of finding the zeros of a vector filed $\mathcal{X}$ on $\mathcal{M}$ as follows. The initial direction is chosen as $\xi = \mathcal{X}(P)$ and the step-size is computed acording the formula

$$\alpha = -\frac{\langle \mathcal{X}(P), \xi \rangle}{\langle \boldsymbol{\nabla}_\xi \mathcal{X}(P), \xi \rangle}.$$

Then, the new direction will be

$$\xi^{\text{new}} = \mathcal{X}(P^{\text{new}}) + \beta\tau\xi,$$

where $P^{\text{new}} = \mu_P^{\text{QR}}(\alpha\xi)$, β is given via one of the formulas in (6.46) with grad $\rho_A(P)$ replaced by $\mathcal{X}(P)$ and grad $\rho_A(P^{\text{new}})$ by $\mathcal{X}(P^{\text{new}})$.

Next, we give an explicit formula for the parallel transport of a tangent vector in our particular cases of $\mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ and $\mathrm{LG}^{\times}(\mathbf{m})$. Let $\Theta := (\Theta_1, \cdots, \Theta_r)$ be such that $\Theta_k^{\dagger} P_k \Theta_k = \Pi_k$. Furthermore, let $P^{\text{new}} := (P_1^{\text{new}}, \cdots, P_r^{\text{new}})$ denote the updated point in $\mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ via the QR-coordinates as in (6.21). According to (2.5) and (6.4), the parallel transport of ξ to $\mathrm{T}_{P^{\text{new}}}\mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ along the curve $\mu_P^{QR}(t\xi)$ is given by the formula

$$\xi_j \mapsto \Theta_j^{\dagger} \begin{bmatrix} I_{m_j} & -Z_j \\ Z_j^{\dagger} & I_{n_j - m_j} \end{bmatrix}_Q \Theta_j \xi_j \Theta_j^{\dagger} \begin{bmatrix} I_{m_j} & Z_j \\ -Z_j^{\dagger} & I_{n_j - m_j} \end{bmatrix}_Q \Theta_j, \tag{6.47}$$

with $Z_j \in \mathbb{C}^{m_j \times (n_j - m_j)}$ as in (6.10), for $j = 1, \cdots, r$. By parallel transporting $\xi_j \in \mathrm{T}_{P_j}\mathrm{Gr}_{m_j,n_j}$ to $\xi_j^{\text{new}} \in \mathrm{T}_{P_j^{\text{new}}}\mathrm{Gr}_{m_j,n_j}$, the pull-back of ξ_j and ξ_j^{new} to the tangent space of the standard projector does not change, i.e.

$$\xi_j^{\text{new}} = (\Theta_j^{\text{new}})^{\dagger} \begin{bmatrix} 0 & Z_j \\ Z_j^{\dagger} & 0 \end{bmatrix} \Theta_j^{\text{new}}. \tag{6.48}$$

The complete Riemannian conjugate gradient algorithm for the optimization of the generalized Rayleigh-quotient on $\mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$ and on $\mathrm{LG}^{\times}(\mathbf{m})$ is presented in Table 6.5 and Table 6.6 respectively.

It is recommended to reset the search direction to the steepest ascent direction after d iterations, i.e. $Z_k^{\text{new}} := g_k^{\text{new}},\ k = 1, \ldots, r$, where d refers to the dimension of the manifold.

The convergence properties of the NCG methods are in general difficult to analyze. Yet, under moderate supplementary assumptions on the cost function one can guarantee that the NCG converges to a stationary point [57]. It is expected that the proposed Riemannian conjugate gradient method has properties similar to those of the NCG.

ALGORITHM 2. **RCG algorithm on** $\mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$

Step 1. Starting point: Given $P=(P_1,\dots,P_r)\in \mathrm{Gr}^{\times}(\mathbf{m},\mathbf{n})$ choose

$$\Theta_j=\begin{bmatrix} U_j \\ \\ V_j \end{bmatrix}\in \mathrm{SU}_{n_j},\ U_jU_j^\dagger=I_{m_j},\ V_jV_j^\dagger=I_{n_j-m_j},$$

such that $P_j=\Theta_j^\dagger\Pi_j\Theta_j$, for $j=1,\dots,r$.
Initial direction: Set

$$\widehat{A}_j:=\Psi_{A,j}(P_1,\dots,I_{n_j},\dots,P_r),$$

compute $\Psi_j',\ \Psi_j'',\ \Psi_j'''$ as in (6.13) and take the steepest ascent direction

$$Z_j=g_j:=\Psi_j''',$$

for $j=1,\dots,r$. Denote $Z:=(Z_1,\dots,Z_r),\ g:=(g_1,\dots,g_r)$.
Step 2. Stopping criterion: $\|\mathrm{grad}_{\rho_A}(P)\|/\rho_A(P)<\varepsilon$.
Step 3. QR-updates:

$$\Theta_j^{\mathrm{new}}=\begin{bmatrix} \alpha I_{m_j} & -\alpha Z_j \\ \\ \alpha Z_j^\dagger & \alpha I_{n_j-m_j}\end{bmatrix}_Q \Theta_j,\quad P_j=\Theta_j^{\mathrm{new}\dagger}\Pi_j\Theta_j^{\mathrm{new}},$$

with the step-size given by $\alpha=-a/(b+c)$, where

$$a:=\sum_{j=1}^{r}\mathrm{tr}\Big(\Psi_j'''Z_j^\dagger\Big),\quad b:=\sum_{j=1}^{r}\mathrm{tr}\Big(\Psi_j'Z_jZ_j^\dagger-Z_j\Psi_j''Z_j^\dagger\Big),$$
$$c:=\sum_{j=1}^{r-1}\sum_{k=j+1}^{r}\rho_A(P_1,\dots,\xi_j,\dots,\xi_k,\dots,P_r),$$

for $j=1,\dots,r$. The tangent vectors ξ_j are given in (6.10).
Step 4. Set $P:=P^{\mathrm{new}}$ and $\Theta:=\Theta^{\mathrm{new}}$.
Step 5. New direction: Update $\Psi_j',\ \Psi_j'',\ \Psi_j'''$ as in (6.13) and compute the new direction

$$Z_j^{\mathrm{new}}=g_j^{\mathrm{new}}+\beta\, Z_j,\ g_j^{\mathrm{new}}:=\Psi_j''',$$

for $j=1,\dots,r$. Here, β is given by the Polak-Ribiere formula

$$\beta=\frac{\langle g^{\mathrm{new}},g^{\mathrm{new}}-g\rangle}{\langle g,g\rangle}$$

Step 6. Set $g:=g^{\mathrm{new}},\ Z:=Z^{\mathrm{new}}$ and go to Step 2.

Table 6.5:

ALGORITHM 2'. **RCG algorithm on $\mathrm{LG}^\times(\mathbf{m})$**

Step 1. Starting point: Given $P = (P_1, \dots, P_r) \in \mathrm{LG}^\times(\mathbf{m})$ choose

$$\Theta_j = \begin{bmatrix} U_j \\ \\ V_j \end{bmatrix} \in \widehat{\mathrm{Sp}}(n_j),\ U_j U_j^\dagger = I_{m_j},\ V_j V_j^\dagger = I_{n_j - m_j},$$

such that $P_j = \Theta_j^\dagger \Pi_j \Theta_j$, for $j = 1, \dots, r$.
Initial direction: Set

$$\widehat{A}_j := \Psi_{A,j}(P_1, \dots, I_{n_j}, \dots, P_r),$$

compute $\Psi_j',\ \Psi_j'',\ \Psi_j'''$ as in (6.13) and take the steepest ascent direction

$$Z_j = g_j := \Psi_j''' + \Psi_j'''^\dagger,$$

for $j = 1, \dots, r$. Denote $Z := (Z_1, \dots, Z_r),\ g := (g_1, \dots, g_r)$.
Step 2. Stopping criterion: $\|\mathrm{grad}_{\rho_A}(P)\| / \rho_A(P) < \varepsilon$.
Step 3. QR-updates:

$$\Theta_j^{\mathrm{new}} = \Theta_j \begin{bmatrix} \alpha I_{m_j} & \alpha Z_j \\ \\ -\alpha Z_j & \alpha I_{n_j - m_j} \end{bmatrix}_Q, \quad P_j = \Theta_j \Pi_j \Theta_j^{\mathrm{new}\dagger},$$

with the step-size given by $\alpha = -a/(b+c)$, where

$$a := \tfrac{1}{2} \sum_{j=1}^{r} \mathrm{tr}\Big((\Psi_j''' + \Psi_j'''^\dagger) Z_j^\dagger \Big), \quad b := \sum_{j=1}^{r} \mathrm{tr}\Big(Z_j (\Psi_j' - \Psi_j'') Z_j \Big),$$

$$c := \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} \rho_A(P_1, \dots, \xi_j, \dots, \xi_k, \dots, P_r),$$

for $j = 1, \dots, r$. The tangent vectors ξ_j are given in (6.10).
Step 4. Set $P := P^{\mathrm{new}}$ and $\Theta := \Theta^{\mathrm{new}}$.
Step 5. New direction: Update $\Psi_j',\ \Psi_j'',\ \Psi_j'''$ as in (6.13) and compute the new direction

$$Z_j^{\mathrm{new}} = g_j^{\mathrm{new}} + \beta\, Z_j,\ g_j^{\mathrm{new}} := (\Psi_j''' + \Psi_j'''^\dagger),$$

for $j = 1, \dots, r$. Here, β is given by the Polak-Ribiere formula

$$\beta = \frac{\langle g^{\mathrm{new}}, g^{\mathrm{new}} - g \rangle}{\langle g, g \rangle}$$

Step 6. Set $g := g^{\mathrm{new}},\ Z := Z^{\mathrm{new}}$ and go to Step 2.

Table 6.6:

6.3 Numerical experiments on $\mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$

In this section we run several numerical experiments suitable for the applications mentioned in Chapter 3.1, i.e. best rank approximation for tensors and subspace clustering, to test the Newton-like (N-like) and Riemannian conjugate gradient (RCG) algorithms. The algorithms were implemented in MATLAB on a personal notebook with 1.8 GHz Intel Core 2 Duo processor.

6.3.1 Best multilinear rank-$(m_1, \ldots, m_r)$ tensor approximation.

To test the performance of the N-like and RCG algorithms, we have considered several examples of tensors of order 3 and 4 with entries chosen from the standard normal distribution and we have estimated their best low-rank approximation. We have started with a truncated HOSVD ([50]) and run several HOOI iterates before starting the N-like or RCG iterations. Depending on the size of the tensor, the number of HOOI iterations necessary to reach the region of attraction of a stationary point $P^* \in \mathrm{Gr}^{\times}(\mathbf{m}, \mathbf{n})$, ranges from 10 to 100. As a stopping criterion we have chosen the relative norm of the gradient $\|\mathrm{grad}_{\rho_A}(P)\|/\rho_A(P)$ to be approximately 10^{-13} in machine precision, i.e. $\|\mathrm{grad}_{\rho_A}(P)\|/\rho_A(P) \approx 10^{-13}$.

Computational complexity. The computational complexity of the N-like method is determined by the computation of the Hessian and the solution of the Newton equation (6.39). Thus, for the best rank-(m, m, m) approximation of a $n \times n \times n$ tensor, the computation of the Hessian is dominated by tensor-matrix multiplications and is of order $O(n^3 m)$. Solving the Newton equation by Gaussian elimination gives a computational complexity of order $O(m^3(n-m)^3)$, i.e. the dimension of the manifold to the power of three. For the computational costs of the RCG method we have to take into discussion only tensor-matrix multiplications, which give a cost per RCG iteration of order $O(n^3 m)$.

Experimental results and previous work. The problem of best low-rank tensor approximation has enjoyed a lot of attention recently. Apart from the well known higher order orthogonal iterations – HOOI ([51]), various algorithms which exploit the manifold structure of the constraint set have been developed. We refer to [20, 41] for Newton methods, to [66] for quasi-Newton methods and to [42] for conjugate gradient and trust region methods on the Grassmann manifold. Similar to the Newton methods in [20, 41], our N-like method converges quadratically to a stationary point of the generalized Rayleigh-quotient when starting in its neighborhood.

We have compared our algorithms with the existing ones in the literature for several tensor instances: quasi-Newton with BFGS, Riemannian conjugate gradient method which uses the Armijo-rule for the computation of the step-size (CG-Armijo), and HOOI. The algorithms were run on the same platform, identically initialized by a truncated HOSVD ([50]) and having the same stopping criterion. For the BFGS quasi-Newton and limited memory quasi-Newton (L-BFGS) methods we have used the code available in [60].

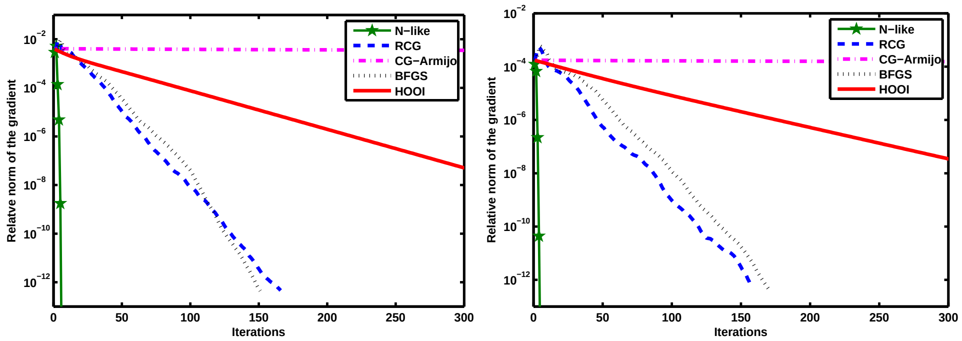
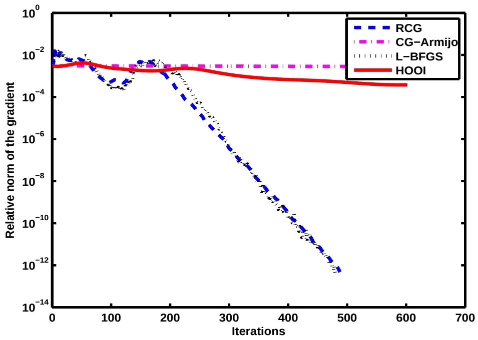


Figure 6.1: Convergence for multilinear rank tensor approximation: number of iterations versus the relative norm of the gradient $\|\mathrm{grad}_{\rho_A}(P^n)\|/\rho_A(P^n)$ at a logarithmic scale. Left: $100 \times 100 \times 100$ tensor approximated by a rank-$(5, 5, 5)$ tensor. Right: $100 \times 150 \times 200$ tensor approximated by a rank-$(15, 10, 5)$ tensor.

Fig. 6.1 shows convergence results for two large size tensors $100 \times 100 \times 100$ and $100 \times 150 \times 200$ approximated by rank-$(5, 5, 5)$ and rank-$(15, 10, 5)$ tensors, respectively. In Fig. 6.2 we plot the convergence behavior of the RCG method for the best rank-$(10, 10, 10)$ approximation of a $200 \times 200 \times 200$ tensor (left) and for the best rank-$(5, 5, 5, 5)$ approximation of a 4 order tensor $50 \times 50 \times 50 \times 50$. Due to the limited memory space allowance, we were not able to run the N-like and BFGS quasi-Newton algorithms for the example on the left. In this case it was still possible to run RCG, L-BFGS, CG-Armijo and HOOI. We did not run the N-like algorithm for the example on the right as well, but not because of memory limitation, but because of the huge number of HOOI iterations necessary to reach the area of attraction for the N-like iteration.

As the numerical experiments have shown, the N-like method has the advantage of fast convergence rate. However, for very large size problems, the N-like algorithm can not be applied, as mentioned before. Even in the cases when it is possible to apply N-like algorithm, it needs a large amount of time per iteration. As an example, for the best rank-$(10, 10, 10)$ of a $180 \times 180 \times 180$ tensor, one N-like iteration took 3 minutes. With the same problem are confronted algorithms which explicitly compute the Hessian and solve the Newton equation, such as [20, 41], but also the trust region method in [42], which approximately solves the Newton equation by a truncated conjugate gradient algorithm and does not compute explicitly the Hessian, but its action on a tangent vector. On the other hand, the low cost iterations of the RCG method makes it a good candidate to solve large size problems. The convergence rate is comparative to that of the BFGS quasi-Newton method in [60], at much lower computational costs. In the examples in which the tensor is a small perturbation of a low-rank tensor, our RCG algorithm manifests a quadratic convergence. For a general tensor, which is not a small perturbation of a low-rank tensor, the CG-Armijo and HOOI methods required an extremely high number of iterations to reach a stationary point.



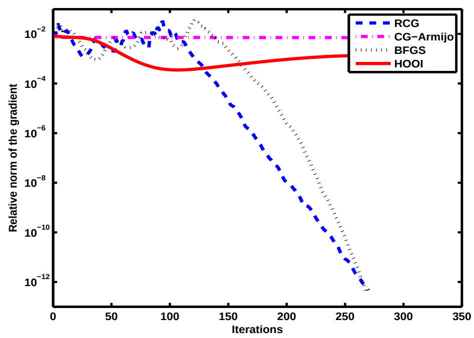


Figure 6.2: Convergence for multilinear rank tensor approximation: number of iterations versus the relative norm of the gradient $\|\text{grad}_{\rho_A}(P^n)\|/\rho_A(P^n)$ at a logarithmic scale. Left: $200 \times 200 \times 200$ tensor approximated by a rank-$(10, 10, 10)$ tensor. Right: $50 \times 50 \times 50 \times 50$ tensor approximated by a rank-$(5, 5, 5, 5)$ tensor.

In Table 6.7 we display the average CPU times (100 trials for each instance) necessary to compute a low rank best approximation for tensors of different sizes and orders by N-like, RCG, BFGS and L-BFGS quasi-Newton methods.

Table 6.7: Average CPU Time

Tensor size and rank	**N-like**	**RCG**	**BFGS**	**L-BFGS**
$50 \times 50 \times 50$, rank-$(7, 8, 5)$	2 s	6 s	24 s	13 s
$100 \times 100 \times 100$, rank-$(5, 5, 5)$	70 s	75 s	150 s	94 s
$100 \times 150 \times 200$, rank-$(15, 10, 5)$	1 min/it	9 min	25 min	15 min
$200 \times 200 \times 200$, rank-$(5, 5, 5)$	-	11 min	-	14 min
$50 \times 50 \times 50 \times 50$, rank-$(5, 5, 5, 5)$	2.5 s/it	9 min	11 min	-

Conclusions. As expected, there is no guarantee that the N-like and RCG iterations converge to a local maximizer of the generalized Rayleigh-quotient. However, in the examples shown in Fig.6.1 and Fig.6.2 the limiting points are local maximizers of the generalized Rayleigh-quotient. The RCG method has very cheap iterations as well as a good convergence rate. Our experiments exhibit shortest CPU time for the RCG method. In the implementation of the RCG, we have used the Polak-Ribiere strategy for the computation of a new direction, since it turned out to be the most efficient for these type of applications, see Figure 6.3.

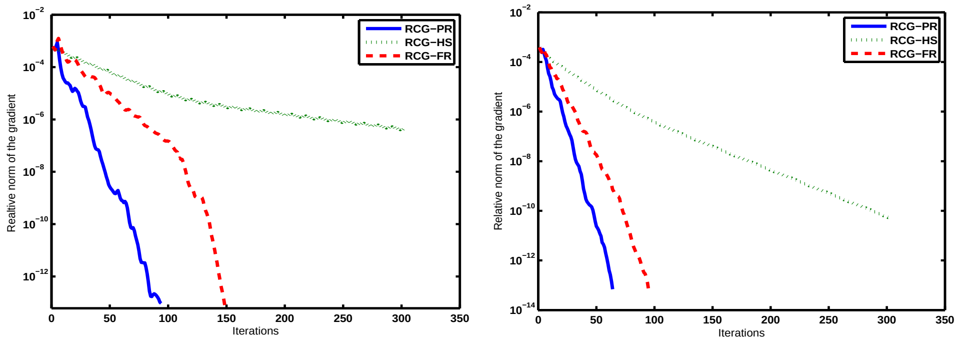


Figure 6.3: Comparison between the convergence speed of RCG method with Polak-Ribiere, Hestens-Stiefel and Fletcher-Reeves strategies in the case of the rank-$(5, 5, 5)$ approximation of randomly chosen tensors of size $10 \times 10 \times 10$ (left) and size $50 \times 50 \times 50$ (right).

6.3.2 Subspace clustering

The experimental setup consists in choosing r subspaces in $\mathbb{R}^3$ ($r = 2, 3$ and 4) and collections of 200 randomly chosen[1] points on each subspace. Then, the sample points are perturbed by adding zero-mean Gaussian noise with standard deviation varying from 0% to 5% in the different experiments. Now, the goal is to detect the exact subspaces or to approximate them as good as possible. For this purpose, we apply our N-like and RCG algorithms to solve the associated optimization task, cf. Section 3.1. The error between the exact subspaces and the estimated ones is measured as in [73], i.e.

$$\mathrm{err} := \frac{1}{r} \sum_{j=1}^{r} \arccos\left(\frac{1}{m_j^2} |\, \mathrm{tr}(P_j \tilde{P}_j)| \right), \tag{6.49}$$

where P_j is the orthogonal projector corresponding to the exact subspace and $\tilde{P}_j$ the orthogonal projector corresponding to the estimated one.

In the case of unperturbed data, we have shown that the global minimizer of ρ_A yields the exact subspaces, thus we expect that for noisy data the global minimizer still gives a good approximation. Since ρ_A has many local optima, for an arbitrary starting point our algorithms can converge to stationary points which lead to a significant error between the exact subspaces and their approximation. Thus, in what follows, we briefly describe a method (PDA, see below) for computing a suitable initial point which guarantees the convergence of our algorithms towards a good approximation of the exact subspaces in our numerical experiment:

The Polynomial Differential Algorithm (PDA) was proposed in [73]. It is a purely algebraic method for recovering a finite number of subspaces from a set of data points

[1] The points have been generated by fixing an orthogonal basis within the subspaces and choosing corresponding coordinates randomly with a uniform distribution over the interval $[-5, 5]$.

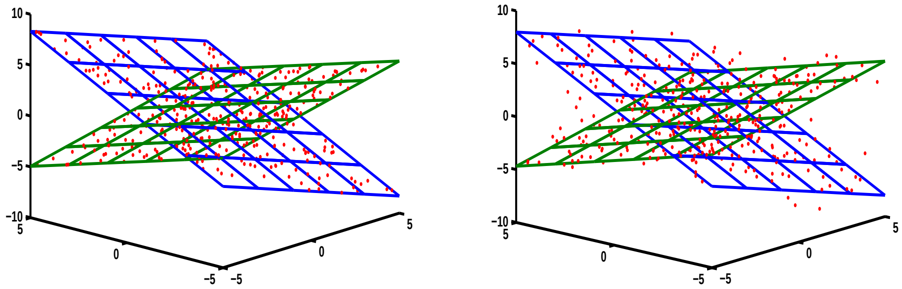


Figure 6.4: Left: Data points drawn from the union of two subspaces of dimension 2 (through the origin) of $\mathbb{R}^3$. Right: Data points from the left figure slightly perturbed by zero mean Gaussian noise with 5% standard deviation.

belonging to the union of these subspaces. From the data set finitely many homogeneous polynomials are computed such that their zero set coincides with the union of the sought subspaces. Then, an evaluation of their derivatives at given data points yields successively a basis of the orthogonal complement of subspaces one is interested in. For noisy data, a slightly modified version of PDA [73] yields an approximation of the unperturbed subspaces. This “first" approximation turned out to be a good starting point for our iterative algorithms which significantly improved the approximation quality.

For each noise level we perform 500 runs of the N-like and Local-CG algorithms for different data sets and compute the mean error between the exact subspaces and the computed approximations. As a preliminary step, we normalize all data points, such that no direction is favored.

In Fig. 6.4, 400 randomly chosen data points which lie exactly in the union of two 2-dimensional subspaces of $\mathbb{R}^3$ (left) and their perturbed[1] images (right) are depicted. Moreover, the two plots display the exact subspaces (left) as well as the ones computed by our N-like algorithm (right). The error between the exact subspaces and our approximation is ca. 2°, whereas the error for the PDA approximation is ca. 5°.

In Fig. 6.6, we plot the mean error (left) for different noise levels and different number of subspaces. We included also the mean error for the staring point of our algorithms, i.e. for the PDA approximation. On the right we demonstrate the fast convergence rate of the N-like and RCG algorithms for the case of 3 and, respectively, 4 subspaces.

Resume. Our numerical experiments have proven that (i) the minimization task proposed in Section 3 is capable to solve subspace detection problems and (ii) our numerical algorithms initialized with the PDA starting point yield an effective method

[1] Gaussian noise with 5% standard deviation

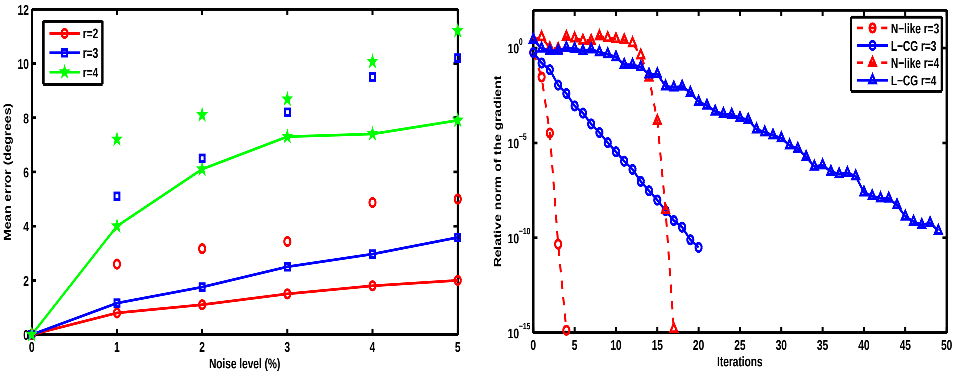


Figure 6.5: Left: The mean error for noise levels from 0% to 5% and different number of subspaces. The disconnected symbols refer to the initial error (PDA) and the corresponding continuous lines refer to the error estimated by our algorithms. Right: Convergence of N-like and RCG for subspace clustering: number of iterations versus the relative norm of the gradient $\|\mathrm{grad}_{\rho_A}(P^n)\|/\rho_A(P^n)$ at a logarithmic scale. Data points from 3 and resp. 4 subspaces perturbed with 5% Gaussian noise. Average CPU time: ca. 0.4 and ca. 2 seconds for the N-like and RCG algorithm, respectively (1.8 GHz Intel Core 2 Duo processor).

for computing a reliable approximation of the perturbed subspaces.

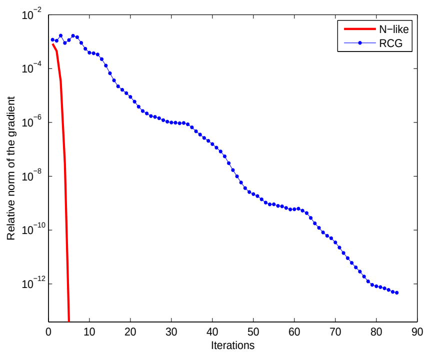


Figure 6.6: Convergence behavior for N-like and RCG methods for finding the zeros of the gradient vector field of ρ_A with $A \in \mathfrak{sym}_{1000}$ Hamiltonian.

6.4 Numerical experiments on $\mathrm{LG}^{\times}(\mathbf{m})$

To test the performance of the Newton-like and conjugate gradient methods on $\mathrm{LG}^{\times}(\mathbf{m})$, we have taken a small perturbation of a symmetric Hamiltonian matrix of the form

$A_1 \otimes A_2 \otimes A_3$ with $A_1,\ A_2,\ A_3 \in \mathfrak{sp}(10, \mathbb{R}) \cap \mathfrak{sym}_{10}$, i.e.

$$A = A_1 \otimes A_2 \otimes A_3 + E/\|E\|,$$

where $E \in \mathfrak{sym}_{1000}$ represents the noise with elements normally distributed in the interval $[0,\ 1]$. As a starting point we have chosen orthogonal projectors P_1, P_2, P_3 corresponding to Lagrangian invariant subspaces of A_1, A_2 and A_3 respectively. The convergence behavior of the Newton-like and conjugate gradient method are displayed in Fig. 6.6.

References

[1] R. Abraham and J. E. Marsden. *Foundations of Mechanics.* Benjamin/Cummings Publishing Co., Reading, MA, 1978.

[2] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Application.* Applied Mathematical Sciences, Vol. 75, Springer–Verlag, 2nd edition, New York, 1988.

[3] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, Princeton, 2007.

[4] V. I. Arnold. Characteristic class entering in quantization conditions. *Funktsional'nyi Analiz i Ego Prilozheniya*, 1 (1):1–14, 1967.

[5] V. I. Arnold. The complex Lagrangian Grassmannian. *Functional Anal. Appl., Brief Communications*, 34 (3):208–210, 2000.

[6] J. C. Bezdek and R. J. Hathaway. Convergence of alternating optimization. *Linear Algebra Appl.*, 370:125–132, 2003.

[7] R. Bhatia. Partial traces and entropy inequalities. *Neural Parallel Sci. Comput.*, 11:351–368, 2003.

[8] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press, 1957.

[9] P. S. Bradley and O. L. Mangasarian. k-plane clustering. *J. of Global Optimization*, 16(1):23–32, 2000.

[10] R. W. Brockett. Least squares matching problems. *Linear Algebra Appl.*, 122–124:761–777, 1989.

[11] J. D. Caroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, 35:283–319, 1970.

[12] P. Comon. Tensor decompositions: State of the art and applications. *Mathematics in Signal Processing*, V:1–24, 2002.

[13] O. Curtef, G. Dirr, and U. Helmke. Riemannian optimization on tensor products of Grassmann manifolds: Applications to generalized Rayleigh-quotients. *SIAM J. Matrix Anal. Appl.*, 33 Issue 1:210–234, 2012.

[14] J. Dieudonne. *Foundations of Modern Analysis, volume 3.* Academic Press, 1982.

[15] G. Dirr, U. Helmke, M. Kleinsteuber, and Th. Schulte-Herbrüggen. Relative C-numerical ranges for applications in quantum control and quantum information. *Linear and Multilinear Algebra*, 56:27–51, 2008.

[16] B. F. Doolin and C. F. Martin. *Introduction to Differential Geometry for Engineers*, volume 136 of *Monographs and Textbooks in Pure and Applied Mathematics.* Marcel Deckker Inc., New York, 1990.

[17] R. Duda, P. Hart, and D. Stork. *Pattern Classification.* Wiley, 2nd edition, New York, 2000.

[18] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–263, 1936.

[19] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20 (2):303–353, 1999.

[20] L. Elden and B. Savas. A Newton-Grassmann method for computing the best multilinear rank-(r_1, r_2, r_3) approximation of a tensor. *SIAM J. Matrix Anal. Appl.*, 31(2):248–271, 2009.

[21] R. Fletcher. *Practical Methods of Optimization, Vol. I.* second edition, John Wiley and Sons, Chichester, 1980.

[22] R. Fletcher. *Practical Methods of Optimization, Vol. II.* second edition, John Wiley and Sons, Chichester, 1981.

[23] D. Gabay. Minimizing a differentiable function over a differentiable manifold. *Journal of Optimization Theory and Applications*, 37(2):177–219, 1982.

[24] K. A. Gallivan, A. Srivastava, X. Liu, and P. Van Dooren. Efficient algorithms for inferences on Grassmann manifolds. *Proceedings of IEEE Conference on Statistical Signal Processing*, pages 315–318, 2003.

[25] C. Geiger and C. Kanzow. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben.* Springer–Verlag, Berlin, Heidelberg, 1999.

[26] C. Geiger and C. Kanzow. *Theorie und Numerik Restringierter Optimierungsaufgaben.* Springer–Verlag, Berlin, Heidelberg, New York, 2002.

[27] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, Maryland, 3rd Edition, 1996.

[28] W. Greub. *Multilinear Algebra.* Springer-Verlag, New York, 2nd Edition, 1978.

[29] V. Guillemin and A. Pollack. *Differential Topology.* Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

[30] R. A. Harshmann. Foundations of the PARAFAC procedure: Model and conditions for an explanatory multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[31] S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces.* Academic Press, New York, 1978.

[32] U. Helmke, K. Hüper, and J. Trumpf. Newton's method on Grassmann manifolds. *arXiv*, 0709.2205v:1–36, 2007.

[33] U. Helmke and M. Kleinsteuber. Riemannian Newton and conjugate gradient algorithm for computing Lagrangian invariant subspaces. *Mathematical Theory of Networks and Systems (MTNS) on CD-ROM*, 2008.

[34] U. Helmke and J. B. Moore. *Optimization and Dynamical Systems.* Springer-Verlag, London, 1996.

[35] J. Hilgert and K. H. Neeb. *Lie-Gruppen und Lie-Algebren.* Vieweg, Braunschweig, 1991.

[36] M. W. Hirsch. *Differential Topology.* Springer-Verlag, New York, 1976.

[37] J. Ho, M. H. Yang, J. Lim, K.C. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. *in IEEE Conf. on Computer Vision and Pattern Recognition*, 1:11–18, 2003.

[38] Wei Hong, John Wright, Kun Huang, and Yi Ma. Multi-scale hybrid linear models for lossy image representation. *IEEE Trans. on Image Processing*, 15:3655–3671, 2006.

[39] M. Horodecki. Entanglement measures. *Quantum information and computation*, 1 (1):3–26, 2001.

[40] K. Hüper and J. Trumpf. Newton-like methods for numerical optimization on manifolds. *Proceedings of Thirty-eighth Asilomar Conference on Signals, Systems and Computers*, pages 136–139, 2004.

[41] M. Ishteva, L. De Lathauwer, P. A. Absil, and S. Van Huffel. Differential-geometric Newton method for the best rank-(R_1, R_2, R_3) approximation of tensors. *Numerical Algorithms*, 51(2):179–194, 2009.

[42] M. Ishteva, L. De Lathauwer, P. A. Absil, and S. Van Huffel. On the best low multilinear rank approximation of higher-order tensors, based on trust region scheme. *Technical report*, pages ESAT–SISTA–09–142, 2009.

[43] D. L. Kleinman. On an iterative technique for Riccati equation computations. *IEEE Trans. Automatic Control*, 13:114–115, 1968.

[44] W. Klingenberg. *Riemannian Geometry.* Studies in Mathematics. Walter de Gruyter, Berlin, New York, 1995.

[45] Anthony W. Knapp. *Lie Groups Beyond an Introduction. 2nd ed.* Progress in Mathematics (Boston, Mass.) 140. Boston, MA: Birkhäuser. xviii, 812 p., 2002.

[46] E. Kofidis and P. A. Regalia. On the best rank-1 approximation of higher–order supersymmetric tensors. *SIAM J. Matrix Anal. Appl.*, 23 (3):863–884, 2002.

[47] T. G. Kolda. Orthogonal tensor decompositions. *SIAM J. Matrix Anal. Appl.*, 23, 2001.

[48] P. Lancaster and L. Rodman. *The Algebraic Riccati Equation.* Oxford University Press, Oxford, 1995.

[49] S. Lang. *Algebra (Graduate Texts in Mathematics).* Springer-Verlag, New York, 3rd Edition, 2002.

[50] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278, 2000.

[51] 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. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, 2000.

[52] A. Laub. A Schur method for solving algebraic Riccati equations. *IEEE Trans. Automatic and Control*, AC-24:913–921, 1979.

[53] D. Leibovici and R. Sabatier. A singular value decomposition of a k-way array for a principal component analysis of multiway data, PTA-k. *Lin. Alg. Appl.*, 269:307–329, 1998.

[54] V. Mehrmann. *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solutions.* Lecture notes in Control and Information Sciences 163, Springer-Verlag, Heidelberg, 1991.

[55] J. W. Milnor. *Topology from the Differentiable Viewpoint.* Princeton University Press, Princeton, New Jersey, 1997.

[56] Nielsen and Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, 2000.

[57] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer, New York, 2nd Edition, 2006.

[58] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[59] M. Plenoi and S. Virmani. An introduction to entanglement measures. *Quantum information and computation*, 7 (1):1–51, 2007.

[60] B. Savas. Algorithm package manual: Best low rank tensor approximation. 2008.

[61] B. Savas and L. H. Lim. Quasi-Newton methods on Grassmannians and multilinear approximations of tensors. *SIAM J. Sci. Comput.*, 32(6):3352–3393, 2010.

[62] S. Schechter. Iteration methods for nonlinear problems. *Trans. Amer. Math. Soc.*, 104:179–189, 1962.

[63] M. A. Shayman. Phase portrait of the matrix Riccati equation. *SIAM J. Optimization and Control*, 24:1–65, 1986.

[64] J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. *CMU School of Computer Science*.

[65] M. Shub. Some remarks on dynamical systems and numerical analysis. *L. Lara-Carrero and J. Lewowicz eds., Dynamical Systems and Partial Differential Equations, Proceedings of VII ELAM Equinoccio, Universidad Simon Bolivar, Caracas*, pages 69–92, 1986.

[66] V. De Silva and L. H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix. Anal. Appl.*, 30(3):1084–1127, 2008.

[67] V. Sima. *Algorithms for Linear-Quadratic Optimization*. Monogr. Textbooks Pure Appl. Math. 200, Marcel Dekker, New York, 1991.

[68] T. Smith. Optimization techniques on Riemannian manifolds. *Fields Institute Communications*, 3:113–146, 1994.

[69] M. Tipping and C. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.

[70] P. Tseng. Nearest q-flat to m points. *Journal of Optimization Theory and Applications*, 105(1):249–252, 2000.

[71] L. R. Tucker. Some mathematical notes of three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[72] C. Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds*. Kluwer Academic, Dordrecht, 1994.

[73] R. Vidal, Y. Ma, and J. Piazzi. Minimum effective dimension for mixtures of subspaces: a robust GPCA algorithm and its applications. *CVPR, vol. 2*, pages 631–638, 2004.

[74] R. Vidal, S. Soatto, Y. Ma, and S. Sastry. An algebraic geometric approach to the identification of a class of linear hybrid systems. *Conference on Decision and Control*, 1:167–172, 2003.

[75] R. Vidal, R. Tron, and R. Hartley. Multiframe motion segmentation with missing data using Power Factorization and GPCA. *International Journal of Computer Vision*, 79(1):85–105, 2008.

[76] T. Wei and P. Goldbart. Geometric measure of entanglement and applications to bipartite and multipartite quantum states. *Phys. Rev. A*, 68(042307):1–12, 2003.

[77] Y. C. Wong. Differential geometry of Grassmann manifolds. *Proc. Nat. Acad. Sci. USA*, 57:589–594, 1967.

[78] A. Yang, J. Wright, Y. Ma, and S. Sastry. Unsupervised segmentation of natural images via lossy data compression. *Computer Vision and Image Understanding*, 110(2):212–225, 2008.

[79] T. Zhang and G. H. Golub. Rank–one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.*, 23:534–550, 2001.