Lie Group Methods in Blind Signal Processing

Dariusz Mika [1],* and Jerzy Jozwik [2],*

[1]   Institute of Technical Sciences and Aviation, The State School of Higher Education in Chelm,
     22-100 Chelm, Poland
[2]   Faculty of Mechanical Engineering, Lublin University of Technology, 20-618 Lublin, Poland
*   Correspondence: dmika@pwsz.chelm.pl (D.M.); j.jozwik@pollub.pl (J.J.); Tel.: +48-501-401-421 (D.M.)

Abstract: This paper deals with the use of Lie group methods to solve optimization problems in blind signal processing (BSP), including Independent Component Analysis (ICA) and Independent Subspace Analysis (ISA). The paper presents the theoretical fundamentals of Lie groups and Lie algebra, the geometry of problems in BSP as well as the basic ideas of optimization techniques based on Lie groups. Optimization algorithms based on the properties of Lie groups are characterized by the fact that during optimization motion, they ensure permanent bonding with a search space. This property is extremely significant in terms of the stability and dynamics of optimization algorithms. The specific geometry of problems such as ICA and ISA along with the search space homogeneity enable the use of optimization techniques based on the properties of the Lie groups $\mathbf{O}(n)$ and $\mathbf{SO}(n)$. An interesting idea is that of optimization motion in one-parameter commutative subalgebras and toral subalgebras that ensure low computational complexity and high-speed algorithms.

Keywords: geometric optimization; Independent Component Analysis; independent subspace analysis; Lie groups; Lie algebra; toral subalgebra; sensors

1. Introduction

Blind signal processing (BSP) is currently one of the most attractive and fast-growing signal processing areas with solid theoretical foundations and many practical applications. BSP has become a vital research topic in many areas of application, particularly in biomedical engineering, medical imaging, speech and image recognition, communication systems, geophysics, economics, and data analysis. The term "blind processing" originates from the basic feature of these processing methods, i.e., the fact that there is no need to use any training data or a priori knowledge to obtain results. These methods include, among others, Independent Component Analysis (ICA), independent subspace analysis (ISA), sparse component analysis (SCA), nonnegative matrix factorization (NMF), singular value decomposition (SVD), principal component analysis (PCA) and minor component analysis (MCA) as well as the related eigenproblem and invariant subspace problem. Optimization problems of this kind often occur in the context of artificial neural networks, signal processing, pattern recognition, computer vision and numeric [1]. BSP is widely used in biomedical engineering, in technical diagnostics as well as in energy. The work [2] presents the use of SCA to analyze biomedical EEG and fMRI signals proving the effectiveness of this method in the detection of ocular artifacts. The use of SCA in technical diagnostics is presented in [3]. The three-dimensional geometric features-based SCA algorithm was used for compound faults diagnosis of roller bearing. A similar topic was discussed in [4] where NMF was used to extract error signals. The conducted experiments confirmed the effectiveness of these methods in extract the fault features and diagnosis for roller bearing. An interesting use of BSP techniques in energy issues is presented in [5]. Bayesian-optimized bidirectional Long Short-Term Memory (LSTM) method was used for energy disaggregation aiming to identify the individual contribution of appliances in the

aggregate electricity load. The use of machine learning techniques as k-means clustering and Support Vector Machine for low-complexity energy disaggregation is presented in [6].

This paper primarily focuses on ICA and ISA problems, which does not, however, limit the applicability of the described methods to other types of problems. The scope of this paper is mainly limited to presenting the geometry of ICA and ISA problems and the application of Lie groups and Lie algebra without providing specific algorithms.

Standard Independent Component Analysis (ICA) consists of the linear transformation of multidimensional data such that the transformed signal components are as much statistically independent as possible. The effectiveness of ICA depends on the correct choice of a cost function and an optimization strategy. Most numerical optimization techniques assume that the model's parameter space is a usual Euclidean space. In many cases, however, the parameter space has a non-linear structure with its unique non-Euclidean geometry. From a mathematical point of view, the space of search equipped with an inner product takes on the properties of Riemannian manifolds, often with desired algebraic properties [7].

The authors of works in this field take advantage of the specific internal geometry and algebraic properties of models such as the orthogonal group $\mathbf{O}(n)$ or the special orthogonal group $\mathbf{SO}(n)$. Apart from the general group properties, these groups also have the structure of a smooth differential manifold, and thus acquire the Lie group properties and the corresponding Lie algebra. The application of this convenient structure to ICA algorithms is described in [8,9].

From the point of view of standard optimization techniques, in issues of this type one deals with the so-called constrained optimization. The problem of constrained optimization occurs in many issues related to signal processing. In the case of ICA, optimization of this kind consists of looking for extrema of the cost function on the set of matrices satisfying the condition of orthonormal columns ($W^T W = I$). However, with standard constrained algorithms one operates in Euclidean space, so in each iterative step the matrix orthogonality is lost. To restore the orthogonality condition, it is necessary to perform orthogonalization in each iterative step (e.g., by the well-known Gram-Schmidt orthogonalization process), which, however, reduces the convergence rate of the algorithms. Other algorithms use the Lagrangian method of optimization by the addition to the cost function of the so-called penalty function to prevent an orthogonality deviation. However, such algorithms are characterized by a low convergence rate and poor quality of the achieved optimum.

If there is a limitation in the form of matrix orthogonality, one can use an alternative method that ensures "locked" with the hyper-surface of orthogonal matrices during optimization motion. This method uses the group structure of a set of orthogonal square matrices which, apart from the properties of a smooth differential manifold, provides the set with the properties of a special structure known as a Lie group.

2. Model Definition (ICA, ISA)

Standard independent components analysis (ICA) consists of estimating a sequence of p statistically independent components (ICs) $s_1, \ldots, s_p$ and the mixing matrix A of dimension $n \times p$ with only a sequence of n observed signals $x_1, \ldots, x_n$. Giving the source signals and observed signals in the form of the source vector $s = (s_1, \ldots, s_p)^T$ and the vector of observed (mixed) signals $x = (x_1, \ldots, x_n)^T$, where T stands for transposition, the standard linear ICA model takes the form (1):

$$x = As \tag{1}$$

This assumes that there is no additional noise signal in the observed signal (Figure 1). The ICA model thereby formulated is characterized by a scale and permutation ambiguity, i.e., it is possible to scale (multiply by a given constant) of any source signals s_i and at the same time to divide the i-th column a_i of the mixing matrix A by this constant, while the observed signal x remains unchanged. The same phenomenon will occur at random transposition of any rows of the source vector s (permutation of the

source vector s) and the same transposition of the columns of the mixing matrix A. It is customary to assume that the source signals have the unit variance ($C_s = \mathrm{E}(ss^T) = I$). In non-negative ICA, it is additionally assumed that the source signals s_i satisfy the condition $s_i \geq 0$ [10,11].
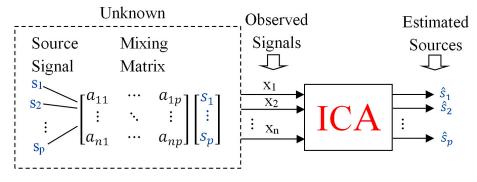

Figure 1. Schematic block scheme of Independent Component Analysis.

A solution for the ICA problem when $n = p$ consists of finding the demixing (filtration) matrix $Q^T = A^{-1}$, $Q \in \mathbf{Gl}(n)$ where the filtration matrix Q belongs to a general linear group $\mathbf{Gl}(n)$ of non-singular matrices $\det(Q) \neq 0$. Source signals are obtained via (2):

$$\hat{s} = Q^T x = Q^T A s \qquad (2)$$

where $\hat{s}$ is the estimator of a source vector s (it meets the statistical assumptions for a source signal).

To reduce the computation load in ICA, the pre-processing usually involves performing the whitening of the observed signal to obtain the signal $z = Bx = BAs$, where B is the whitening matrix, with unit variance and the decorrelation $C_z = \mathrm{E}(zz^T) = I$. Assuming that $C_s = I$, we get (3):

$$I = C_z = \mathrm{E}(zz^T) = BA\mathrm{E}(ss^T)(BA)^T = BA(BA)^T \qquad (3)$$

From this it follows that $(BA)^T = (BA)^{-1}$. Hence, the transformation from s to z takes place via an orthogonal matrix BA. Therefore, if $\hat{s} = W^T z = W^T BAs = Us$, then the matrix $U = W^T BA$ must be an orthogonal matrix (permutation matrix), and thus a new filtering matrix W (after whitening) must also satisfy the orthogonality condition. The whitening of the observed signal therefore simplifies the ICA problem from optimization on the general linear group $\mathbf{Gl}(n)$ (matrices Q only satisfying the invertibility condition $\det(Q) \neq 0$) to optimization on the special orthogonal group $\mathbf{SO}(n)$ (matrices W satisfying the orthogonality condition $W^T W = I$). Both groups are Lie groups at the same time.

Standard ICA is based on the assumption that $n = p$, i.e., the number of source signals s_i is known and equal to the number of observed signals x_i. ICA also yields interesting results in a more general case when the number of estimated source signals p is unknown. In this case, it can be $n \neq p$. When $n < p$, i.e., the number of observed signals is smaller than the number of source signals, the problem is known as over complete bases ICA, whereas when $n > p$ it is called under complete bases ICA. This kind of problem can be formally considered to be unconstrained optimization on the Stiefel manifold [12,13]. It is also possible to solve ICA problems for the case $p = 1$. This type of problem is often called Single Channel Source Separation [14,15].

Hyvarinen and Hoyer introduced independent subspace analysis (ISA) [16] by omitting the statistical independence condition between extracted source components. The source vector **s** is composed in d_k-tuple ($k = 1, \ldots, r$), where for a given tuple a statistical dependence between its source signals s_i is allowed, while signals belonging to different tuples are statistically independent. When using the whitening process, the ISA problem boils down to finding orthogonal matrices $W^T W = I$ as in standard ICA. However, due to the statistical relationship between the source signals in the tuple, ISA problem optimization cannot be performed on an ordinary Stiefel manifold. It is necessary to

introduce a different, more universal manifold allowing for additional symmetries. This manifold is known as a flag manifold.

Traditionally, the ICA model assumes the statistical independence of extracted source signals. It turns out, however, that there are reasons to replace the orthonormality condition with the condition of source signal normality [17]. A precise definition of the ICA problem consists of finding a linear non-orthogonal transformation (of the coordinate system) of multidimensional data such that the transformed data have minimal mutual information. Hyvarinen [18] demonstrated (in an ICA problem) the differences between the use of cost functions based on mutual information and those based on the so-called non-Gaussianity. Achieving maximum de-correlation by maximizing the sum of non-Gaussianity of independent components (ICs) is not necessarily related to the minimization of mutual information (MI). In addition, the orthonormality condition leads to a smaller subset of matrices, which simplifies the optimization process yet may reduce its quality. Orthonormality imposes a greater limitation on the degrees of freedom than normality. In standard ICA, the orthonormality condition of $n \times n$ filtering matrices reduces the number of the degrees of freedom to $(n-1)/2$, while the normality condition increases the number of free parameters to $n(n-1)$, which considerably improves the quality of obtained results. A problem of this type can be formally considered to be the unconstrained optimization of an oblique manifold [19,20].

3. Geometry of ICA, ISA and Other BSP Models

The manifolds frequently arise from BSP tasks for a general do not have the group properties. Nevertheless, they are homogenous spaces of the Lie groups. A homogenous space M is a manifold on which the Lie group G acts transitively [21]. This property is fundamental for the considered manifolds because it enables analyzing them as quotient spaces. As mentioned in Section 2, the optimization problem in standard ICA (ISA) boils down to optimizing on the general linear group $\mathbf{Gl}(n)$ (matrices Q only satisfy the invertibility condition $\det(Q) \neq 0$). The whitening of the observed signal simplifies the ICA problem to optimization on the special orthogonal group $\mathbf{SO}(n)$ (matrices W satisfy the orthogonality condition $W^T W = I_n$). In the case of an under complete problem $p < n$, i.e., when the number of extracted ICs is smaller than the number of observed signals, the set of filtering matrices can be treated as an orthogonal Stiefel manifold $\mathbf{St}(n,p)$ defined as the set of orthonormal matrices of dimension $n \times p$ with the form (4):

$$\mathbf{St}(n,p) = \left\{ W = (w_1, \ldots, w_p) \middle| W \in \mathbb{R}^{n \times p}, \ W^T W = I_p, \, rank(W) = p \right\} \tag{4}$$

which can be regarded as the quotient space arising from the orthogonal group.

Lie group $G = \mathbf{O}(n)$ acts transitively on the Stiefel manifold via (5):

$$\mathbf{O}(n) \times \mathbf{St}(n,p) \ni (Q, W) \rightarrow QW \in \mathbf{St}(n,p) \tag{5}$$

where $Q \in \mathbf{O}(n)$, $W \in \mathbf{St}(n,p)$. It is possible to demonstrate that for two given points W_1, $W_2 \in \mathbf{St}(n,p)$ there exists $Q \in \mathbf{O}(n)$ such that $W_2 = QW_1$. This means that starting from any point $W_0 \in \mathbf{St}(n,p)$ it is possible to reach any point $W \in \mathbf{St}(n,p)$ by the G action. Resorting to group theory terminology, one can say that the entire manifold $\mathbf{St}(n,p)$ is equivalent to the single orbit $G(W_0)$ of a given point W_0 where

$$G(W_0) = \left\{ W = QW_0 \middle| Q \in \mathbf{O}(n), \ W_0 \in \mathbf{St}(n,p) \right\} \tag{6}$$

The point W on the manifold $\mathbf{St}(n,p)$ can be expressed via a certain point Q on $\mathbf{O}(n)$. The mapping $\pi : Q \rightarrow W$ is surjective, i.e., many to one (projective mapping). Redundancy of this mapping is described by so-called the isotropy subgroup H of the point W_0. It is a set of matrices that do not change W_0

$$H = \{N \in \mathbf{O}(n) | N W_0 = W_0\} \tag{7}$$

The isotropy subgroup $H \in \mathbf{O}(n)$ of the group $\mathbf{O}(n)$ of the point $W_0 \in \mathbf{St}(n,p)$ has the form (8):

$$H = (W_0, W_{0\perp}) \begin{pmatrix} I_p & 0 \\ 0 & \mathbf{O}(n-p) \end{pmatrix} (W_0, W_{0\perp})^T \tag{8}$$

where $W_{0\perp} \in \mathbf{St}(n, n-p)$ is any $n \times (n-p)$ matrix that satisfies the condition $(W_0, W_{0\perp}) \in \mathbf{O}(n)$. It is easy to check that the isotropy condition of point W_0 is satisified (9):

$$H \cdot W_0 = (W_0, W_{0\perp}) \begin{pmatrix} I_p & 0 \\ 0 & \mathbf{O}(n-p) \end{pmatrix} (W_0, W_{0\perp})^T W_0 = W_0 \tag{9}$$

Choosing $W_0 = \begin{pmatrix} I_p \\ 0_{n-p,p} \end{pmatrix}$ the isotropy subgroup W_0 is a set $H = \begin{pmatrix} I_p & 0 \\ 0 & \mathbf{O}(n-p) \end{pmatrix}$.

In this shot two $n \times n$ orthogonal matrices represent the same point of the Stiefel manifold if their first p columns are identical or equivalently, if they are related by right multiplication of a matrix of the form $\begin{pmatrix} I_p & 0 \\ 0 & \mathbf{O}(n-p) \end{pmatrix}$ where $\mathbf{O}(n-p)$ is an orthogonal matrix group of dimension $(n-p) \times (n-p)$ [1]. From a mathematical point of view, we say that such representations are in an equivalence relation. All matrices in an equivalence relation form what is called the equivalence class $[W]$. Thus, the point on the Stiefel manifold is the equivalence class $[W]$ of $n \times n$ orthogonal matriceswith identical first p columns, while the Stiefel manifold is a quotient space of the form (10):

$$\mathbf{St}(n,p) \cong \mathbf{O}(n)/\mathbf{O}(n-p) \tag{10}$$

and specifically as $\mathbf{St}(n,p) \cong \mathbf{O}(n)/H$ where $H = \begin{pmatrix} I_p & 0 \\ 0 & \mathbf{O}(n-p) \end{pmatrix}$. However, H is isomorphic to $\mathbf{O}(n-p)$, i.e., $H \cong \mathbf{O}(n-p)$, therefore $\mathbf{St}(n,p) \cong \mathbf{O}(n)/\mathbf{O}(n-p)$.

There are many applications for the problem formulated as the finding of (zero) extreme of a given field defined in a non-Euclidean subspace of dimension p embedded in the Euclidean space $\mathbb{R}^n$. This non-Euclidean subspace is known as the Grassmann manifold $\mathbf{Gr}(n,p;\mathbb{R})$ [22,23]. Grassmann manifolds can be described as an equivalence class of $n \times p$ orthogonal matrices spanning the same p-dimensional subspace $[W] = \{W\mathbf{O}(p) \big| W \in \mathbf{St}(n,p)\}$. Therefore, from a theoretical point of view, the Grassmann manifold can be expressed as the quotient space $\mathbf{Gr}(n,p) \cong \mathbf{St}(n,p)/\mathbf{O}(p)$ and given that $\mathbf{St}(n,p) \cong \mathbf{O}(n)/\mathbf{O}(n-p)$, the Grassmann manifold $\mathbf{Gr}(n,p)$ can also be seen as the quotient space $\mathbf{O}(n)/\mathbf{O}(p) \times \mathbf{O}(n-p)$. In this case, the equivalence class $[W] = \{W \begin{pmatrix} \mathbf{O}_p & 0 \\ 0 & \mathbf{O}(n-p) \end{pmatrix} \big| W \in \mathbf{O}(n)\}$ is a set of square n-dimensional orthogonal matrices whose first p columns span the same p-dimensional subspace. Manifolds of this type are used, among others, in invariant subspace analysis, application-driven dimension reduction and subspace tracking [24,25].

When there is a need for a simultaneous (parallel) subspace extraction, as is the case in independent subspace analysis (ISA), one resorts to the concept of generalized flag manifold, which is a manifold consisting of orthogonal subspaces that constitutes a generalization of both Stiefel and Grassmann manifolds [26–28]. The generalized flag manifold $\mathbf{Fl}(n, d_1, \ldots, d_r; \mathbb{R})$ is defined as (11):

$$\mathbf{Fl}(n, d_1, \ldots, d_r; \mathbb{R}) = \left\{ W \big| W \in \mathbb{R}^{n \times p}, W^T W = I_p, \right\} \tag{11}$$

where the orthogonal matrix W takes the form (12):

$$W = [W_1, \ldots, W_r], W_i = [w_1^i, \ldots, w_{d_i}^i], \tag{12}$$

where $w_k^i \in \mathbb{R}^n$, $k = 1, \ldots, d_r$ for a specified $i = 1, \ldots, r$ is a set of orthogonal bases that span subspaces V_i. The subspaces V_i are orthogonal relative to each other and satisfy the condition (13):

$$V = V_1 \oplus V_2 \oplus \ldots \oplus V_r \subset \mathbb{R}^{n \times p} \tag{13}$$

Points on the flag manifold are a set of vector spaces V which can be decomposed as (13). If all $d_i (1 \leq i \leq r) = 1$, the manifold $\mathbf{Fl}(n, d_1, \ldots, d_r; \mathbb{R})$ is reduced to the Stiefel manifold $\mathbf{St}(n, p)$. If $r = 1$, it is reduced to the Grassmann manifold $\mathbf{Gr}(n, p)$. It is abbreviated as $\mathbf{Fl}(n, \mathbf{d})$ where $\mathbf{d} = (d_1, \ldots, d_r)$. The orthogonal group $\mathbf{O}(n)$ also acts transitively on the manifold $\mathbf{Fl}(n, \mathbf{d})$ via simple matrix multiplication (14):

$$\mathbf{O}(n) \times \mathbf{Fl}(n, \mathbf{d}) \ni (Q, W) \rightarrow QW \in \mathbf{Fl}(n, \mathbf{d}) \tag{14}$$

The isotropy subgroup $H \in \mathbf{O}(n)$ of the group $\mathbf{O}(n)$ of the point $W \in \mathbf{Fl}(n, \mathbf{d})$ has the form (15):

$$H = (W, W_\perp) diag(R_1, \ldots, R_r, R_{r+1})(W, W_\perp)^T \tag{15}$$

where $diag(R_1, \ldots, R_r, R_{r+1})$ is a block-diagonal matrix of the form $\begin{pmatrix} R_1 & & & 0 \\ & \ddots & & \vdots \\ \vdots & & R_r & \\ 0 & & & R_{r+1} \end{pmatrix}$, $R_k \in$
$\mathbf{O}(d_k)$, $(1 \leq k \leq r)$, $R_{r+1} \in \mathbf{O}(n - p)$, $W_\perp \in \mathbf{St}(n, n - p)$ is any $n \times (n - p)$ matrix that satisfies the condition $[W, W_\perp] \in \mathbf{O}(n)$. It is easy to check that the isotropy condition of point W is satisified (16):

$$\begin{aligned} H \cdot W &= [W](W, W_\perp) diag(R_1, \ldots, R_r, R_{r+1})(W, W_\perp)^T \cdot W = (WR, \ W_\perp R_{r+1})(W^T W_\perp^T) W \\ &= (WRW^T + W_\perp R_{r+1} W_\perp^T) W = WR = [W] \end{aligned} \tag{16}$$

where $R = diag(R_1, \ldots, R_r)$, $WR = [W] = diag(W_i R_i)$, $i = 1, \ldots, r$ is an equivalence class of the point on $\mathbf{Fl}(n, \mathbf{d})$. This means that any two matrices W_1 and W_2 satisfying the condition $W_2 = W_1 R = (W_1, \ldots, W_r) diag(R_1, \ldots, R_r) = diag(W_1 R_1, W_2 R_2, \ldots, W_r R_r)$ are identified with the very same point on the manifold $\mathbf{Fl}(n, \mathbf{d})$. Given the above,

$$\mathbf{Fl}(n, \mathbf{d}) \cong \mathbf{O}(n) / \mathbf{O}(d_1) \times \ldots \times \mathbf{O}(d_r) \times \mathbf{O}(n - p) \tag{17}$$

As it was already mentioned, the manifold $\mathbf{Fl}(n, \mathbf{d})$ is locally isomorphic to $\mathbf{St}(n, p)$ as a homogenous space when all $d_i (1 \leq i \leq r) = 1$ and to the manifold $\mathbf{Gr}(n, p)$ when $r = 1$.

In terms of optimization, the homogeneity of the considered differential manifolds enables the search (optimization motion) in the group $\mathbf{O}(n)$ or $\mathbf{SO}(n)$, and the use of optimization techniques that are well known and adapted to these types of groups. Section 4 presents the basic ideas of optimization methods used in $\mathbf{SO}(n)$ and the concept of toral subalgebra that is characteristic of problems of this type.

4. Lie Group Optimization Methods. One-Parameter Subalgebra and Toral Subalgebra

The idea of a standard optimization procedure based on the Lie groups consists of performing the optimization motion in the Lie algebra space and then mapping exp to find a solution in the Lie group (manifold). The optimization motion in the group $\mathbf{SO}(n)$ starting from the point (matrix) W_0 therefore consists of, first, the transition to the Lie algebra $\Omega = \log W \in \mathfrak{so}(n)$ via mapping inversely to the exponentiation $\log := \exp^{-1}$ of motionin the Lie algebra (performing an operation of addition (of matrices) in the abelian group) in order to obtain a new antisymmetric matrix $\Omega' \in \mathfrak{so}(n)$ and, finally, returning to the Lie group via exponential mapping $W' = \exp \Omega' \in \mathbf{SO}(n)$. A simple update method using line search procedure relies on finding the search direction in Lie algebra $\mathfrak{so}(n)$ calculating

the gradient of cost function J in Lie algebra space. This gradient must be skew-symmetric (see Appendix A) so (18) [9]:

$$\nabla_A J = (\nabla_w J)\, W^{\mathrm{T}} - W(\nabla_w J)^{\,\mathrm{T}} \tag{18}$$

Applying the steepest descent procedure with small constant update factor μ we start from $A = 0_n \in \mathfrak{so}(n)$, move to $B = -\mu\nabla_A J$, map to $R = \exp(B) \in \mathbf{SO}(n)$ and finally perform rotating (multiplicative) update $W_{k+1} = \exp(-\mu\nabla_A J)W_k$. This kind of optimization method is called a geodesic flow method [9].

At this point it is necessary to comment on motion in the Lie algebra. In our context, the addition of vectors in the Lie algebra $\mathfrak{so}(n)$ can only be useful if it is matched by multiplication in the Lie group $\mathbf{SO}(n)$. Then one can write (19):
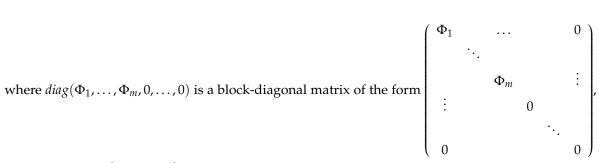
$$\exp(A)\exp(B) = \exp(B)\exp(A) = \exp(A + B) \tag{19}$$

As was already mentioned, this equation holds true only when the matrices A and B commutate $[A, B] = 0$. This condition is satisfied for all matrices with $\mathfrak{so}(2)$. When $n \geq 3$, this condition is not satisfied for all matrices in the algebra. When the matrices not commutate (non-abelian Lie algebra), Equation (19) is not satisfied and optimization motion in the Lie algebra (sum A+B) in a direction of e.g., the cost function gradient will not be reflected in the Lie group $\exp(A + B) \neq \exp(A)\exp(B)$. However, taking $\exp(A) = I_n$, which is tantamount to selecting an initial matrix $A_0 = 0_n$, this condition will always be satisfied. In this case, $[A_0, B] = 0$, and Equation (19) is satisfied too. This is tantamount to motion in the one-parameter Lie algebra. By selecting $A = t\Omega$ for a random antisymmetric matrix $\Omega \in \mathfrak{so}(n)$ and a scalar $\in \mathbb{R}$, all matrices of this form commutate with each other ($A = t_1\Omega$, $B = t_2\Omega$: $[A, B] = t_1\Omega t_2\Omega - t_2\Omega t_1\Omega = 0$). A set of such matrices $\mathfrak{so}_\Omega(n) = \left\{t\Omega \middle| \Omega \in \mathfrak{so}(n),\ t \in \mathbb{R}\right\}$ is in itself a Lie algebra known as a one-parameter subalgebra of the Lie algebra $\mathfrak{so}(n)$. The subalgebra $\mathfrak{so}_\Omega(n)$ is an abelian (commutative) algebra related to the one-parameter subgroup $R(t) = \exp(t\Omega)$. Optimization motion in the subalgebra $\mathfrak{so}_\Omega(n)$ is therefore an equivalent (generalization) to the idea of linear motion in Euclidean space. In this case, the optimization procedure consists of searching for a minimum of the cost function along the subalgebra $\mathfrak{so}_{\Omega_1}(n)$ (for a chosen search direction Ω_1), which corresponds to the search along the subgroup $R(t)$.

Having found the cost function minimum $(R(t)W_0)$, where W_0 is a starting point, a new direction of linear searches Ω_2 is selected, and the procedure is repeated until the desired convergence is achieved. Plumblay [8] proposed a modification of the standard procedure described above. This modification consists of moving the point of "origin" of the Lie algebra from a neutral element of the group to point W. Due to the group properties $\mathbf{SO}(n)$, it can be written that $W' = RW$ for some matrix $R \in \mathbf{SO}(n)$. Moving from the matrix $W = I_n W$ to $W' = RW$ is therefore equivalent to moving from the identity matrix I_n to the matrix R. This procedure consists of moving from the matrix $0_n = \log I_n \in \mathfrak{so}(n)$ to $\Omega = \log R \in \mathfrak{so}(n)$ in the Lie algebra and then returning to the group $\mathbf{SO}(n)$ via the exponential mapping $R = \exp\Omega$ and, finally, determining $W' = RW = (\exp\Omega)W \in \mathbf{SO}(n)$. This is equivalent to the concept of optimization motion in the one-parameter abelian subalgebra described above.

The above optimization procedures are computationally expensive due to the necessity of performing (computationally expensive) matrix exponentiation in every iterative step. The representation of antisymmetric matrices in the Jordan canonical form enables the decomposition of optimization movement in the group $\mathbf{SO}(n)$ to commutative rotations in orthogonal planes. Every antisymmetric matrix Ω can be presented in a block-diagonal form (for $2\,m \leq n$) (20):

$$\Omega = Q diag(\Phi_1, \ldots, \Phi_m, 0, \ldots, 0)Q^t \tag{20}$$

where $diag(\Phi_1, \ldots, \Phi_m, 0, \ldots, 0)$ is a block-diagonal matrix of the form
$$\begin{pmatrix} \Phi_1 & & \cdots & & 0 \\ & \ddots & & & \\ & & \Phi_m & & \vdots \\ \vdots & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{pmatrix},$$

$Q \in \mathbf{SO}(n)$, $\Phi_i = \begin{pmatrix} 0 & \varphi_i \\ -\varphi_i & 0 \end{pmatrix}$ denotes the 2×2 dimensional antisymmetric matrices [29]. This form is known as the Jordan canonical form. Since the relationship $\exp(Q^T \Omega Q) = Q^T \exp(\Omega) Q$ holds true, the matrix Ω can be decomposed into a sum of the form $\Omega = \Omega_1 + \ldots + \Omega_m$ where Ω_i is the matrix only containing the i-th Jordan matrix Φ_i and zeros beyond it (21):

$$\Omega = Q diag(\Phi_1, 0, \ldots, 0) Q^t + \ldots + Q diag(0, \ldots, \Phi_m, 0, \ldots, 0) Q^t \tag{21}$$

The exponentiation of thereby presented matrix Ω yields an orthogonal matrix W of the form (22):

$$W = \exp \Omega = Q diag(R_1, \ldots, R_m, 1, \ldots, 1) Q^t \tag{22}$$

where $R_i = \begin{pmatrix} cos\varphi_i & sin\varphi_i \\ -sin\varphi_i & cos\varphi_i \end{pmatrix}$ are the 2×2 dimensional rotation matrices. The matrix W can be decomposed into a product of the matrix $W = W_1 \ldots W_m$ where W_i has the form (23):

$$W_i = \exp \Omega_i = Q diag(1, \ldots, 1, R_i, 1, \ldots, 1) Q^t \tag{23}$$

One can notice that the exponentiation of the matrix Ω in the Jordan form is reduced to a simple and inexpensive calculation of the functions $sin\varphi_i$ and $cos\varphi_i$, which significantly increases the speed of optimization algorithms. The Jordan canonical form of antisymmetric matrix can be obtained via symmetric eigenvalue decomposition [29]. It can be observed that the antisymmetric matrix Ω commutates with the symmetric matrix $\Omega^2 = -\Omega^T \Omega$, which means that Ω and $\Omega^T \Omega$ have the same eigenvectors and eigenvalues. The eigenvalues $\Omega^T \Omega$ occur in pairs corresponding to individual Jordan matrices Φ_i.

This form can be visualized as compounding rotations (represented by W_i) in mutually orthogonal planes. In addition, the rotation matrices W_i commutate $[W_i, W_j] = 0$. The commutation property of the rotation matrix W_i provides the possibility of using the optimization procedure on $\mathbf{SO}(n)$, moving in the Lie algebra $\mathfrak{so}(n)$.

The case of $\mathbf{SO}(4)$ is interesting from a geometrical point of view. The Jordan canonical form of the antisymmetric matrix Ω contains two blocks (matrices) Φ_i:

$$\Omega = Q \begin{pmatrix} 0 & \varphi_1 & 0 & 0 \\ -\varphi_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \varphi_2 \\ 0 & 0 & -\varphi_2 & 0 \end{pmatrix} Q^t \tag{24}$$

Here, the orthogonal matrix W takes the form (25):

$$W = \exp \Omega = Q \begin{pmatrix} cos\varphi_1 & sin\varphi_1 & 0 & 0 \\ -sin\varphi_1 & cos\varphi_1 & 0 & 0 \\ 0 & 0 & cos\varphi_2 & sin\varphi_2 \\ 0 & 0 & -sin\varphi_2 & cos\varphi_2 \end{pmatrix} Q^t \tag{25}$$

A visual representation of this case shows rotations in two mutually orthogonal planes, which corresponds to toral geometry (Figure 2).

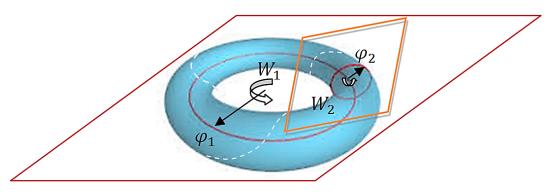**Figure 2.** Visual representation of the toral subalgebra $\mathfrak{t}(p)$ for $p = 2$. The angles φ_1, φ_2 and the matrices W_1 and W_2 are as in Equation (23). The broken line marks the search curve for the case $\frac{k_1}{k_2} = 3$.

From the point of view of optimization procedures, the rotation angles φ_1 and φ_2 should not be free parameters (independent of each other). For the procedure to make sense, the curve over which the search is carried out after a complete rotation (or its multiple) relative to one of the planes of rotation should return to the starting point on the toral surface (Figure 2). This is possible when for some t, the relationships $t\varphi_1 = 2k_1\pi$ and $t\varphi_2 = 2k_2\pi$ are satisfied for the integers k_1 and k_2. Therefore, the angle of rotation should be described by the relationship $\frac{\varphi_1}{\varphi_2} = k_1/k_2$ or $\varphi_1 = a\varphi_2$, where $a = k_1/k_2$ is a rational number. This concept is naturally transferred to a general case of $\mathbf{SO}(n)$ for $n > 4$. The Jordan canonical form represents optimization motion in one-parameter Lie subalgebra $R(t) = \exp(t\Omega) \in \mathfrak{so}_\Omega(n)$ as a rotation in p mutually orthogonal planes, and these rotations are commutative. The geometry of 2-dimensional torus for $\mathbf{SO}(4)$ can also be generalized to the geometry of the p-dimensional torus in $\mathbf{SO}(n)$ where $n = 2p$ for even n or $n = 2p + 1$ for odd n. This perception of motion in $\mathbf{SO}(n)$ leads to the concept of toral subalgebra $\mathfrak{t}(p) \subset \mathfrak{so}(n)$. If we consider a general case of motion on the surface of a p-dimensional torus where the angles of rotation φ_i are not interrelated by the above relationship, and individual independent planes of rotation are represented by a set p of commuting matrices Ω_i ($\Omega = \sum_i \Omega_i$). The motion (or rather rotation) on each of the independent planes of rotation can be expressed in the form of a parameterized curve $B_i = t_i\Omega_i$, or actually via its simple exponentiation $W_i = \exp(t_i\Omega_i)$. The set of independent parameters t_i that can be identified with the angles of rotation φ_i forms a coordinate system on the toral subalgebra $\mathfrak{t}(p)$. Compared to the original n–dimensional search space $\mathbf{SO}(n)$, the toral subalgebra $\mathfrak{t}(p)$ is, however, an abelian algebra, which means that motion in this search space is commutative. This ensures the possibility of motion in all directions specified by the coordinates t_i and their sum in the form $B = t_1\Omega_1 + \ldots + t_p\Omega_p$ will be reflected in the composition of rotations $W = \exp B = \exp(t_1\Omega_1) \cdot \ldots \cdot \exp(t_p\Omega_p)$. The optimization procedure based on such a concept consists of decomposing to canonical form a specific antisymmetric matrix $B \in \mathfrak{so}(n)$ (this can be, for example, a cost function gradient as in the method of steepest descent), and thereby formulating a toral subalgebra. Since the orthogonal matrix Q in the Jordan decomposition (24) is constant, the transition to a new point W in the search space is done by determining p values of the *sin* and *cos* functions corresponding to p planes of rotation. After finding in the subalgebra the point that minimizes the cost function, a new antisymmetric matrix $B' \in \mathfrak{so}(n)$ is calculated and again presented in the Jordan canonical form, which establishes a new toral subalgebra. The procedure is repeated until the set minimum cost function is reached. A different problem concerns the determination of the direction of search $B \in \mathfrak{so}(n)$ and the manner of search along the subalgebra. The selection of directions and the manner of searches depend on the adopted optimization procedure. It can be the steepest descent (SD) method and, in general, geodesic flow, Newton's method or conjugate gradients. This problem has been extensively studied in [12,22,30].

5. Experimental Results

To illustrate the presented optimization methods on Lie groups, we will first present a rather simply simulation experiment. The purpose of this example is to show how different algorithms work on optimization problems with unitarity constraint. To this end, let us consider the Lie group of complex numbers with the unit module **U**(1), which is isomorphic with the **SO**(2) group. The unitarity constraint of elements of this group is a unit circle on the complex plane. The cost function we will minimize will be $J(z) = |z + 0.3|^2$ with constraint $zz^* = 1$. For optimization, we will use five types of steepest descent (SD) algorithms:

(1)　algorithm SD unconstrained on the Euclidean space,
(2)　algorithm SD on the Euclidean space with constraint restoration,
(3)　algorithm SD on the Euclidean space with penalty function,
(4)　non-geodesic algorithm SD on Riemannian space,
(5)　geodesic algorithm SD on Riemannian space.

In Algorithm (1) update rule has the form $z_{k+1} = z_k - \mu(z_k + 0.3)$ where μ is the step size. The quantity $\frac{\partial J}{\partial z^*} = (z_k + 0.3)$ is the gradient (The gradient of the function defined in the complex space has the form [31]: $\frac{\partial J}{\partial z^*} = \frac{1}{2}\left(\frac{\partial J}{\partial \mathrm{Re}(z)} + \frac{\partial J}{\partial \mathrm{Im}(z)}\right)$ where $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ is respectively the real and imaginary part of a complex number z.) of the cost function J on the Euclidean space. Algorithm (2) uses the same update rule, but after each iteration the unitarity condition is restored in the form $z_{k+1} = \frac{z_{k+1}}{|z_{k+1}|}$. In Algorithm (3) we used the Lagrange multiplier method. The penalty function of the form $(|z_k|^2 - 1)^2$ weighted by a Lagrange parameter λ has been added to the initial cost function in order to penalize deviations from unitarity. In this case, update rule is $z_{k+1} = z_k - \mu[(z_k + 0.3) + \lambda z_k(|z_k|^2 - 1)^2]$. In the case of (4), the algorithm works in the Riemannian space (unit circle) determined by the condition $zz^* = 1$. At each point of z_k the algorithm determines the search direction tangent to the unit circle and after each iteration the obtained point is projected back to the unit circle. In this case, update rule has the form $z_{k+1} = \pi\left(z_k - \mu\left(\frac{\partial J}{\partial z^*} - \langle z, \frac{\partial J}{\partial z^*}\rangle z\right)\right) = \pi(z_k - \mu[z_k(1 - |z_k|^2) - 0.3(z_k^2 - 1)])$ where π is the projection operator on the unit circle. In Algorithm (5) we used a multiplicative optimization algorithm on the Lie group described in Section 4. In this case, the update rule has the form $z_{k+1} = \exp(0.6\,\mu i\mathrm{Im}(z_k))z_k$ where $\mathrm{Im}(z_k)$ is imaginary part of z_k. The starting point of each algorithm is $z_0 = \exp(i\pi/4)$. Figure 3 shows the results of the simulation.
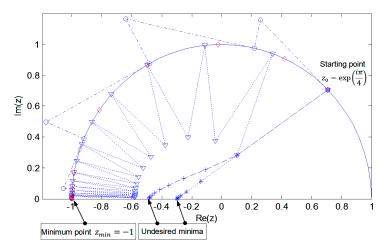

Figure 3. Comparison of SD algorithms for minimizing the cost function on group **U**(1). Methods in the Euclidean versus Riemannian space (Lie group methods). * algorithm SD unconstrained on the Euclidean space (1). ▽ algorithm SD on the Euclidean space with constraint restoration (2). + algorithm SD on the Euclidean space with penalty function (3). o non-geodesic algorithm SD on Riemannian space (4). ◇ geodesic algorithm SD on Riemannian space (5).

In the point $z_{(1)} = -0.3$ the cost function reaches its minimum $J(z_{(1)}) = 0$. However, this is the undesirable minimum determined in the Euclidean space by Algorithm (1) not taking into account the unitarity constraint (Figure 3). The minimum considering this constraint is at the point $z_{min} = -1$ and the value of the cost function reaches its minimum over the Riemannian space, i.e., on the unit circle $J(z_{min}) = 0.49$. Unconstrained SD Algorithm (1) and with penalty function (3) achieve undesirable minimums in points respectively from $z_{(1)} = -0.3$ and from $z_{(3)} = -0.5$ while Algorithms (2) (4) and (5) minimum appropriate $z_{min} = -1$. In the case of Algorithms (2) and (4), a characteristic "zig-zag" is associated with lowering the constraint surface and undesirable from the point of view of optimizing properties. Algorithm (4) determines the SD direction tangent to the constraint condition, thus leaving the unit circle in the optimization motion. The resulting point is again projected into a unit circle. Algorithm (5) using the multiplicative update rule on the Lie group (phase rotation) described in this article naturally ensures the condition of unitarity at each step. The optimization movement takes place at each step along the geodesic line. This simple one-dimensional example is only intended to present the idea of algorithms on Lie groups. The following is an example of using these methods on a real signal. As an example of the practical application of optimization methods on Lie groups, we will present a solution to the ICA problem. As the source signals, three speech recordings and a quasi-noisy signal (harmonic signal with high noise content) (Figure 4) with a length of 5000 samples (1.25 s) were used. The source signals were mixed using a four-by-four random mixing matrix. The four observed signals are shown in Figure 4. A **SO**(4) group optimization algorithm was used to implement ICA.
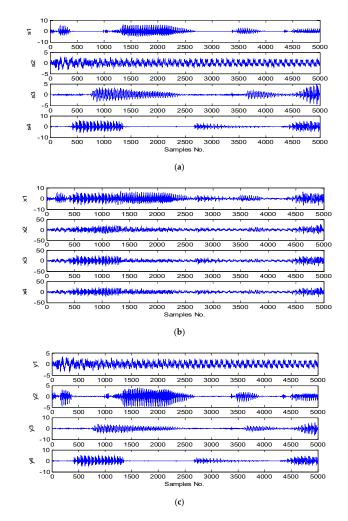

(a)

(b)

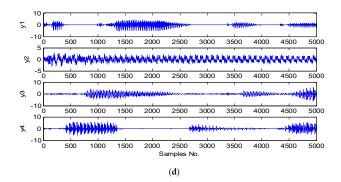

(c)

Figure 4. *Cont.*

(**d**)

Figure 4. Comparison of ICA results using INFOMAX algorithm and optimization on the **SO**(4) group, (**a**) source signals, (**b**) observed signals (mixed), (**c**) ICA results for the INFOMAX algorithm, (**d**) ICA results for the algorithm on the group **SO**(4).

For comparison, the INFOMAX algorithm in its original form was also used [32]. Based on visual inspection and listening to the separated components, it can be concluded that ICA results using the INFOMAX algorithm and optimization on the **SO**(4) group are good with scale and permutation accuracy. The INFOMAX algorithm with the assumed convergence criterion converges after about 30–40 steps, while the algorithm on the **SO**(4) group after about 20 steps. Figure 5 shows the sum of entropy values of separated components depending on the iteration number.
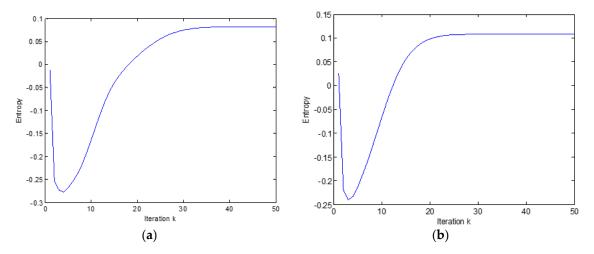

(**a**) (**b**)

Figure 5. Comparison of entropy sum value of received components, (**a**) INFOMAX algorithm, (**b**) optimization algorithm on a group **SO**(4).

The optimization algorithm on the **SO**(4) group converges to $E(Y) = 0.11$ (The entropy value was determined according to an approximate relationship [32]: $E(Y) = -\sum_i E\{\sum_n \tan h(y_i)\} + \log(\det(W)))$ while the INFOMAX algorithm to $E(Y) = 0.084$. Listening to the results and comparison with the sources confirms the better ICA separation results obtained by the **SO**(4) group optimization algorithm.

6. Conclusions

This paper described the application of the Lie group methods for blind signal processing, including ICA and ISA. Theoretical fundamentals of the Lie groups and the Lie algebra as well as the geometry of problems occurring in BSP and basic optimization techniques based on the use of Lie groups are presented. Owing to the specific geometry and algebraic properties of BSP problems, it is possible to use Lie group methods to solve these problems. The homogeneity of search space (parameters) in BSP problems enables the use of optimization techniques based on the Lie group methods for the groups **O**(n) and **SO**(n). It has been demonstrated that the one-parameter subalgebra $\mathfrak{so}_\Omega(n)$

ensures the convenient property of commutating search directions. In addition, the presentation of an antisymmetric matrix (search direction) in the Jordan canonical form establishes the toral subalgebra $\mathfrak{t}(p) \subset \mathfrak{so}(n)$, which—in terms of optimization algorithms—ensures low computational complexity and high process dynamics.

Appendix A Lie Group and Lie Algebra

A characteristic of Lie groups is optimization motion that always leaves the orthogonality condition intact (the system constantly remains on the constraint surface). Before providing a formal definition of Lie groups, we will give an example to illustrate this concept. Let us consider a set of complex numbers of modulus 1 in $r - \varphi$ notation: $G = \left\{ z \in \mathbb{C} \middle| |z| = \left| e^{i\varphi} \right| = 1,\ \varphi \in \mathbb{R} \right\}$. If one of these numbers $a = e^{i\varphi}$ is multiplied by a number $b = e^{i\omega}$, we get $ab = e^{i(\varphi + \omega)}$, which is also several modulus 1, i.e., it belongs to a set Z (Figure A1). On the other hand, if we want to reach any number $c = e^{i\tau}$ of modulus 1 starting from the number $a = e^{i\varphi}$, it is enough to multiply a by $b = e^{i\omega}$ such that $\omega = \tau - \varphi$.
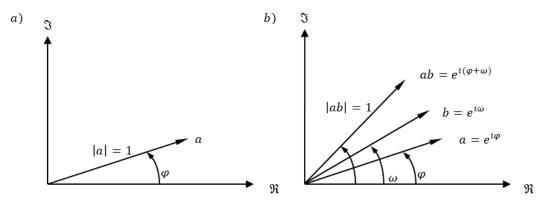

Figure A1. Complex value (**a**) a in $r - \varphi$ notation, and (**b**) product of a and b.

Thus, it can be observed that a group is formed by the set of complex numbers Z of modulus 1 combined with a multiplication operation. As a reminder, the group G is formed by a set of elements and a group operation $(\cdot)$ (colloquially known as multiplication) with the following properties:

1. Closure under group operation: if $a, b \in G$ then $a{\cdot}b \in G$
2. Associativity: $a{\cdot}(b{\cdot}c) = (a{\cdot}b){\cdot}c$
3. There exists a neutral element I and an inverse element $z^{-1} \in G$ for every element of the group, such that: $I{\cdot}z = z{\cdot}I = z.\ z{\cdot}z^{-1} = z^{-1}{\cdot}z = I$

It can easily be verified that the set Z satisfies these conditions with a group operation in the form of multiplication of complex numbers, the neutral element $I = 1 = e^{i0}$ and the inverse $z^{-1} = e^{-i\varphi}$ of any element of $z = e^{i\varphi}$. Moreover, since the multiplication operation in the set G is commutative, this set forms a commutative group, also known as an abelian group. In addition, the group G is smooth, i.e., differentiable. We can perform an infinitesimally small movement in the group by going from one number to another, these numbers differing from each other by an infinitesimally low value. In fact, the group G locally "looks" like a section of the real line $\mathbb{R}$. This means that the motion in the group G

can be described with one parameter, e.g., the angle φ, and thus this group with the coordinate system of φ creates a differential manifold. The group which is a smooth differential manifold is called the Lie group. The Lie group combines algebraic properties of the groups possessing the properties of a smooth differential manifold. As a result, we can, for example, consider a tangent space $T_x G$ of the Lie group G at point x. In a special case when $x = I$, the tangent space in an identity element of the group $T_I G$ additionally provided with an elementary operation known as the Lie bracket is called the Lie algebra $\mathfrak{g}$. The group operation (multiplication) by a constant element of the group Q creates a differentiable mapping of the group G to itself, i.e., a diffeomorphism known as left translation (A1):

$$L_Q : \ G \to G : \ W \to QW; Q, W \in G \tag{A1}$$

and right translation (A2):

$$R_Q : \ G \to G : \ W \to WQ; Q, W \in G \tag{A2}$$

and tangential mappings induced by them (A3) and (A4):

$$dL_Q : \ T_W G \to T_{QW} G : \ Z \to QZ; Z \in T_W G, \ Q \in G \tag{A3}$$

$$dR_Q : \ T_W G \to T_{WQ} G : \ Z \to ZQ; Z \in T_W G, \ Q \in G \tag{A4}$$

A characteristic of the Lie group is that each of its elements W can be "moved" by left or right translation to a "convenient" neighborhood of the identity element, and the same applies to every tangent space $T_W G$:

$$L_{W^{-1}}(W) = W^{-1}W = I \ : R_{W^{-1}}(W) = WW^{-1} = I \tag{A5}$$

$$dL_{W^{-1}}(T_W G) = \left\{ W^{-1}Z = X \in T_I G = \mathfrak{g} \right\} : dR_{W^{-1}}(T_W G) = \left\{ ZW^{-1} = X \in T_I G = \mathfrak{g} \right\} \tag{A6}$$

Equation (A6) can also be written in a reversed form (A7):

$$T_W G = dL_W(T_I G) = \{WX; \ X \in T_I G = \mathfrak{g}\} : \ T_W G = dR_W(T_I G) = \{WX; \ X \in T_I G = \mathfrak{g}\} \tag{A7}$$

From the above it can be seen that the tangent space $T_W G$ via the tangent mapping $dL_{W^{-1}}$ or $dR_{W^{-1}}$ is transferred to the Lie algebra $\mathfrak{g}$.

From the above one can draw an important conclusion. If the structure of a Lie algebra $\mathfrak{g}$ is known, it can be used to conveniently parameterize the neighbourhood of an identity element of the group G via the application of a suitable homeomorphism. Such homeomorphism is widely known as the exponential map or, in short, exponentiation, and is denoted as $\exp : \mathfrak{g} \to G$. It should be noted that "exp" is here only a symbolic denotation and even in the case of Lie matrix groups it does not necessarily mean a matrix exponentiation operation.

A set of 2-dimensional orthogonal matrices $W^T W = I_2$ forms a Lie group as is done by the set of complex numbers of modulus 1. It is also an abelian (commutative) group and has the same algebraic properties as the set of complex numbers of modulus 1. In general, square n-dimensional orthogonal matrices forming the group $\mathbf{O}(n)$ can be presented in a block-diagonal form with diagonal elements as 2×2 orthogonal matrices. This form is known as the Jordan canonical form of orthogonal matrix [23]. Understanding the "behaviour" of the Lie groups over the complex numbers of modulus 1 and thus 2×2 orthogonal matrices is therefore crucial for analysis of a general case. The group $\mathbf{O}(n)$ is also a Lie group composed of two disjoint parts (subsets): the orthogonal matrices with determinant 1 and the matrices with determinant -1. For example, for the group $\mathbf{O}(2)$, these two components can be shown in the form (A8):

$$W = \begin{pmatrix} cos\varphi & sin\varphi \\ -sin\varphi & cos\varphi \end{pmatrix} \qquad W' = \begin{pmatrix} cos\varphi & -sin\varphi \\ -sin\varphi & -cos\varphi \end{pmatrix} \tag{A8}$$

$$det W = 1 \qquad det W' = -1$$

for any real number φ that can be a Cartesian coordinate system identified with the angle of rotation. As one cannotice, both in the first and the second part, the transition from one element to another takes place (smoothly) by changing the parameter φ. The transition between the matrices of both parts is only possible via multiplying the matrices belonging to both parts, which is tantamount to permutation of the matrices. The transposition of the matrix columns or rows in any part will lead to a change in the sign of the determinant. This occurs as a transition between the parts. However, a smooth transition between the parts is not possible. Hence, these components (parts) are called disjoint. The subset of matrices with determinant 1 is called a special orthogonal group **SO**(n). It is a subgroup **O**(n) and constitutes an associative Lie group (it is possible to move between elements of the subgroup smoothly by performing multiplication in the subgroup).

In ICA problems, due to permutation ambiguity (Section 2), optima of the cost function must be present in both subgroups **O**(n). We will therefore limit our considerations to the (associative) subgroup **SO**(n) considerations, which will not diminish the general nature of the considerations. The multiplication of the matrices belonging to **SO**(2):

$$WV = \begin{pmatrix} cos\varphi & sin\varphi \\ -sin\varphi & cos\varphi \end{pmatrix} \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} = \begin{pmatrix} \cos(\varphi+\theta) & \sin(\varphi+\theta) \\ -\sin(\varphi+\theta) & \cos(\varphi+\theta) \end{pmatrix} \tag{A9}$$

This involves, as one can observe, adding the angles φ and θ. The above also demonstrates that the multiplication operation in the group **SO**(2) is commutative. The group **SO**(2) is also known as the rotation group because its operation over any vector can be associated with a rotation on the plane $\mathbb{R}^2$. Moreover, the group **SO**(2)—like the group of complex numbers of modulus 1 can be parameterized by means of one parameter—the angle $0 \leq \varphi < 2\pi$. Groups that "operate" in the same way (have the same algebraic properties) are called isomorphic. From the point of view of optimization techniques, the key question is how to determine a derivative in the Lie group G. Basing on the previously considered Lie group of complex numbers of modulus 1, let us consider the curve $z(t) = e^{i\varphi(t)}$ in the group G parameterized by t (the t parameter can be identified as time). Assuming that $\varphi(t) = \omega t$, the derivative of the curve $z(t)$ with respect to t has the form (A10):

$$\frac{dz}{dt} = \frac{d}{dt}\left(e^{i\varphi(t)}\right) = i\left(\frac{d\varphi}{dt}\right)e^{i\varphi(t)} = i\omega e^{i\varphi(t)} = i\omega z \tag{A10}$$

As one can observe, the derivative $z(t)$ is proportional to $iz = e^{i\frac{\pi}{2}}z = e^{i(\varphi(t)+\frac{\pi}{2})}$, which means that it is perpendicular to z. This result is analogous to the relationship between the vector of velocity and the radiusin circular motion. Assuming that ω is an angular velocity and z is a radius, the circular motion can be described in an identical manner. The velocity vector of length ωz is tangent to the trajectory, i.e., it is perpendicular to the radius of the circle. A similar relationship can be obtained by differentiating the group **SO**(2). Likewise, assuming that $\varphi = \omega t$ and differentiating $W = \begin{pmatrix} cos\varphi & sin\varphi \\ -sin\varphi & cos\varphi \end{pmatrix}$ relative to t, we get (A11):

$$\frac{dW}{dt} = \frac{d}{dt}\begin{pmatrix} cos\varphi & sin\varphi \\ -sin\varphi & cos\varphi \end{pmatrix} = \begin{pmatrix} -sin\varphi & cos\varphi \\ -cos\varphi & -sin\varphi \end{pmatrix}\omega = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}W = XW \tag{A11}$$

One can notice that similarly to the group of complex numbers of modulus 1, the derivative in the group **SO**(2) is proportional to an element of the group where the derivative $dW/dt \sim W$ is determined. It is easy to check that this derivative is tangent at point W to the constraint surface $tr\left(\left(\frac{dW}{dt}\right)^T W\right) = trX = 0$. This pattern is also true for the general case **SO**(n). Differential equations with the structure $\frac{dA}{dt} = B(t)A(t)$, where $A(t) \in G$ belongs to the Lie group and $B(t) \in \mathfrak{g}$ belongs to the Lie algebra, are known as differential equations of the Lie groups. As with the solution of the scalar differential equation $\frac{dz}{dt} = az \Longrightarrow z(t) = e^{at}$, the solution of the differential Equation (A11) is

a matrix $W(t) = \exp(tX)$, where $\exp(.)$ is a matrix exponentiation operation given by the general formula (Maclaurin series) (A12):

$$\exp(tX) = I + tX + \frac{t^2 X^2}{2!} + \ldots + \frac{t^k X^k}{k!} + \ldots, \tag{A12}$$

It can easily be verified that $W(t)$ satisfies Equation (A11):

$$\frac{dW}{dt} = \frac{d}{dt}\exp(tX) = X\exp(tX) = XW$$

The space tangent to **SO**(2) at point W is given by the matrices, where $X = -X^T = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}$ is an antisymmetric matrix. As one can observe, with any antisymmetric matrix $X \neq 0$, every element of the group **SO**(2) can be made precise by defining a single parameter t as (A13):

$$W(t) = \exp(tX) = \exp\begin{pmatrix} 0 & t\omega \\ -t\omega & 0 \end{pmatrix} = \exp(\Omega) \tag{A13}$$

where $\Omega = -\Omega^T = \begin{pmatrix} 0 & t\omega \\ -t\omega & 0 \end{pmatrix}$ is an antisymmetric matrix.

From the point of view of the optimization process, it is a very convenient property. Instead of searching in the space of orthogonal matrices, it is enough to search in the space of the angles φ or, alternatively, in the space of antisymmetric matrices Ω. The space tangent to **SO**(n) in the identity (neutral) element I_n is therefore determined by the set of antisymmetric matrices $\Omega = -\Omega^T$. According to previous general considerations, this space is called a Lie algebra $\mathfrak{so}(n)$ of the Lie group **SO**(n). Homeomorphism is determined by the matrix exponential operation, $\Omega \in \mathfrak{so}(n) \to \exp(\Omega) \in$ **SO**(n) (or $\exp : \mathfrak{so}(n) \to$ **SO**(n)). As mentioned above, from the point of view of the optimization process, it is convenient to navigate in the Lie algebra because it is a vector space, so the addition of elements and multiplication by scalars are allowed. The result of such operations will still belong to the Lie algebra. The idea presented above for the group **SO**(2) is generalized to **SO**(n). However, for $n \geq 3$, the matrix multiplication is not commutative: $\exp(A)\exp(B) \neq \exp(B)\exp(A)$. From the general dependence (A14):

$$\exp(\varepsilon A)\exp(\varepsilon B) - \exp(\varepsilon B)\exp(\varepsilon A) = \varepsilon^2[A, B] + O(\varepsilon^3) \tag{A14}$$

where $[\cdot, \cdot]$ is the matrix commutator or the Lie bracket defined as $[A, B] = AB - BA$, $O(\varepsilon^3)$ is a lower order ε^3 and $\varepsilon \ll 1$ is a scalar, it follows that the "non-commutativeness" of matrix multiplication in the group **SO**(n) is expressed with the commutator $[\cdot, \cdot]$ of the matrices belonging to the Lie algebra $\mathfrak{so}(n)$. The commutator of antisymmetric matrices is also an antisymmetric matrix $[A, B]^T = (AB - BA)^T = B^T A^T - A^T B^T = BA - AB = -[A, B]$, which means that $[A, B] \in \mathfrak{so}(n)$. The Lie algebra of antisymmetric matrices is therefore closed under addition, multiplication by scalars and the Lie bracket. In addition to this, the Lie bracket has the following properties (A15)–(A17):

$$[A, A] = 0 \tag{A15}$$

$$[A + B, C] = [A, C] + [B, C] \tag{A16}$$

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \tag{A17}$$

The first two properties result from the antisymmetric nature of the commutator, while the third property is called the Jacobi identity and is widely known in differential geometry. The multiplication of 2×2 antisymmetric matrices $A, B \in \mathfrak{so}(2)$ is commutative, therefore $[A, B] = 0$, and hence **SO**(2) is an abelian group. Elements of the group **SO**(3) can be equated with the rotations relative to the

coordinate system origin of the space $\mathbb{R}^3$. As we know, such rotations are not commutative. Therefore, the group **SO**(3) is non-abelian (non-commutative).

References

1. Fiori, S. Quasi-geodesic neural learning algorithms over the orthogonal group: A tutorial. *J. Mach. Learn. Res.* **2005**, *6*, 743–781.
2. Georgiev, P.; Theis, F.; Cichocki, A.; Bakardjian, H. Sparse component analysis: A new tool for data mining. In *Data Mining in Biomedicine*; Springer: Boston, MA, USA, 2007; pp. 91–116.
3. Hao, Y.; Song, L.; Cui, L.; Wang, H. A three-dimensional geometric features-based SCA algorithm for compound faults diagnosis. *Measurement* **2019**, *134*, 480–491. [CrossRef]
4. Hao, Y.; Song, L.; Cui, L.; Wang, H. Underdetermined source separation of bearing faults based on optimized intrinsic characteristic-scale decomposition and local non-negative matrix factorization. *IEEE Access* **2019**, *7*, 11427–11435. [CrossRef]
5. Kaselimi, M.; Doulamis, N.; Doulamis, A.; Voulodimos, A.; Protopapadakis, E. Bayesian-optimized Bidirectional LSTM Regression Model for Non-Intrusive Load Monitoring. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2747–2751.
6. Altrabalsi, H.; Stankovic, V.; Liao, J.; Stankovic, L. Low-complexity energy disaggregation using appliance load modelling. *Aims Energy* **2016**, *4*, 884–905. [CrossRef]
7. Smith, S.T. Optimization techniques on Riemannian manifolds. *Fields Inst. Commun.* **1994**, *3*, 113–135.
8. Plumbley, M.D. Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras. *Neurocomputing* **2005**, *67*, 161–197. [CrossRef]
9. Plumbley, M.D. Lie group methods for optimization with orthogonality constraints. In Proceedings of the International Conference on Independent Component Analysis and Signal Separation, Granada, Spain, 22–24 September 2004; pp. 1245–1252.
10. Plumbley, M.D. Algorithms for nonnegative independent component analysis. *IEEE Trans. Neural Netw.* **2003**, *14*, 534–543. [CrossRef] [PubMed]
11. Plumbley, M.D. Optimization using Fourier expansion over a geodesic for non-negative ICA. In Proceedings of the International Conference on Independent Component Analysis and Signal Separation, Granada, Spain, 22–24 September 2004; pp. 49–56.
12. Edelman, A.; Arias, T.A.; Smith, S.T. The geometry of algorithms with orthogonality constraints. *Siam J. Matrix Anal. Appl.* **1998**, *20*, 303–353. [CrossRef]
13. Birtea, P.; Casu, I.; Comanescu, D. Steepest descent algorithm on orthogonal Stiefel manifolds. *arXiv* **2017**, arXiv:1709.06295.
14. Mika, D.; Kleczkowski, P. ICA-based single channel audio separation: New bases and measures of distance. *Arch. Acoust.* **2011**, *36*, 311–331. [CrossRef]
15. Mika, D.; Kleczkowski, P. Automatic clustering of components for single channel ICA-based signal demixing. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Lisbon, Portugal, 13–16 June 2010; pp. 5350–5359.
16. Hyvärinen, A.; Hoyer, P. Emergence of phase-and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.* **2000**, *12*, 1705–1720. [CrossRef] [PubMed]
17. Selvana, S.E.; Amatob, U.; Qic, C.; Gallivanc, K.A.; Carforab, M.F.; Larobinad, M.; Alfanod, B. *Unconstrained Optimizers for ICA Learning on Oblique Manifold Using Parzen Density Estimation*; Tech. Rep. FSU11-05; Florida State University Department of Mathematics: Tallahassee, FL, USA, 2011.
18. Hyvarinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; John Wiley & Sons: New York, NY, USA, 2001.
19. Absil, P.-A.; Gallivan, K.A. Joint Diagonalization on the Oblique Manifold for Independent Component Analysis. In Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, France, 14–19 May 2006.
20. Selvan, S.E.; Amato, U.; Gallivan, K.A.; Qi, C.; Carfora, M.F.; Larobina, M.; Alfano, B. Descent algorithms on oblique manifold for source-adaptive ICA contrast. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 1930–1947. [CrossRef] [PubMed]

21. Nishimori, Y.; Akaho, S. Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold. *Neurocomputing* **2005**, *67*, 106–135. [CrossRef]

22. Absil, P.-A.; Mahony, R.; Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*; Princeton University Press: Princeton, NJ, USA, 2009.

23. Absil, P.-A.; Mahony, R.; Sepulchre, R. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.* **2004**, *80*, 199–220. [CrossRef]

24. Comon, P.; Golub, G.H. Tracking a few extreme singular values and vectors in signal processing. *Proc. IEEE* **1990**, *78*, 1327–1343. [CrossRef]

25. Demmel, J.W. Three methods for refining estimates of invariant subspaces. *Computing* **1987**, *38*, 43–57. [CrossRef]

26. Nishimori, Y.; Akaho, S.; Abdallah, S.; Plumbley, M.D. Flag manifolds for subspace ICA problems. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP'07, Honolulu, HI, USA, 15–20 April 2007; pp. 1417–1420.

27. Nishimori, Y.; Akaho, S.; Plumbley, M.D. Riemannian optimization method on the flag manifold for independent subspace analysis. In Proceedings of the International Conference on Independent Component Analysis and Signal Separation, Charleston, SC, USA, 5–8 March 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 295–302.

28. Nishimori, Y.; Akaho, S.; Plumbley, M.D. Natural conjugate gradient on complex flag manifolds for complex independent subspace analysis. In Proceedings of the International Conference on Artificial Neural Networks, Prague, Czech Republic, 3–6 September 2008; pp. 165–174.

29. Gallier, J.; Xu, D. Computing exponentials of skew-symmetric matrices and logarithms of orthogonal matrices. *Int. J. Robot. Autom.* **2003**, *18*, 10–20.

30. Pecora, A.; Maiolo, L.; Minotti, A.; De Francesco, R.; De Francesco, E.; Leccese, F.; Cagnetti, M.; Ferrone, A. Strain gauge sensors based on thermoplastic nanocomposite for monitoring inflatable structures. In Proceedings of the 2014 IEEE Metrology for Aerospace (MetroAeroSpace), Benevento, Italy, 29–30 May 2014; pp. 84–88.

31. Krantz, S.G. *Function Theory of Several Complex Variables*; American Mathematical Soc.: Providence, RI, USA, 2011.

32. Cichocki, A.; Amari, S. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*; John Wiley & Sons: New York, NY, USA, 2002.