Signal Processing 93 (2013) 459–467

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Averaging complex subspaces via a Karcher mean approach

K. Hüper [a], M. Kleinsteuber [b,*], H. Shen [b]

[a] Department of Mathematics, Julius-Maximilians-Universität Würzburg, 97074 Würzburg, Germany
[b] Department of Electrical Engineering and Information Technology, Technische Universität München, 80333 Munich, Germany

ARTICLE INFO

Article history:
Received 3 June 2011
Received in revised form
9 August 2012
Accepted 29 August 2012
Available online 7 September 2012

Keywords:
Conjugate gradient algorithm
Grassmannian
Complex projective space
Complex linear subspaces
Karcher mean

ABSTRACT

We propose a conjugate gradient type optimization technique for the computation of the Karcher mean on the set of complex linear subspaces of fixed dimension, modeled by the so-called Grassmannian. The identification of the Grassmannian with Hermitian projection matrices allows an accessible introduction of the geometric concepts required for an intrinsic conjugate gradient method. In particular, proper definitions of geodesics, parallel transport, and the Riemannian gradient of the Karcher mean function are presented. We provide an efficient step-size selection for the special case of one dimensional complex subspaces and illustrate how the method can be employed for blind identification via numerical experiments.

1. Introduction

In a wide range of signal processing applications and methods, subspaces of a fixed dimension play an important role. Signal and noise subspaces of covariance matrices are well studied objects in classical applications, such as subspace tracking [1] or direction of arrival estimation [2]. More recently, a significant amount of work is focussed on applying subspace based methods to image and video analysis [3], as well as to matrix completion problems [4]. One fundamental challenge amongst these works is the study of the statistical properties of distributions of subspaces. Specifically, in the present work, we are interested in computing the mean of a set of subspaces of equal dimension via averaging.

The averaging process, considered in this paper, employs the intrinsic geometric structure of the underlying set and is also known as the computation of the *Karcher mean* (in differential geometry, [5]), Fréchet mean or barycentre (statistics), geometric mean (linear algebra and matrix analysis), or center of mass (physics). General concepts of a geometric mean have been extensively studied from both theoretical and practical points of view. To mention just a few, they include probability theory and shape spaces [6,7], imaging [8], linear algebra and matrix analysis [9], interpolation [10], and convex and differential geometry [11,12].

An appropriate mathematical framework is given by the so-called Grassmannian, which assigns a differentiable manifold structure to the set of subspaces of equal dimension. Usually, this is achieved by identification with a matrix quotient space.[1] In this work, we do not follow such an approach. By following [14] instead, we identify the set of subspaces of equal dimension with a set of matrices. More precisely, we consider the set of Hermitian projectors of fixed rank, which inherits its differentiable structure from the surrounding vector space of Hermitian matrices. In contrast to [14], we consider the complex case here. The identification of the complex Grassmannian with Hermitian projection matrices allows an accessible

* Corresponding author.
E-mail addresses: hueper@mathematik.uni-wuerzburg.de (K. Hüper), kleinsteuber@tum.de (M. Kleinsteuber), hao.shen@tum.de (H. Shen).
URL: http://www.gol.ei.tum.de (M. Kleinsteuber).

0165-1684/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.08.020

[1] The set of m-dimensional subspaces of $\mathbb{C}^n$ is identified with $\mathbb{C}_*^{n\times m}/GL(m)$, cf. [13], $\mathbb{C}_*^{n\times m}$ is the set of full rank $(n\times m)$-matrices, and $GL(m)$ are the complex invertible $(m\times m)$-matrices. The equivalence relation is defined by $X\sim Y \Leftrightarrow X = gY$ for some $g\in GL(m)$.

introduction of the geometric concepts such as geodesics, parallel transport, and the Riemannian gradient of the Karcher mean function.

In general, computing the Karcher mean on a smooth manifold involves a process of optimization, which by its own is of both theoretical and practical interest. Various numerical methods have been developed on the Grassmannian, such as a direct method [15], gradient descent algorithms [13], Newton's method [14], and conjugate gradient methods [16,17].

In this work, we focus on the development of conjugate gradient methods. These methods have been proven to be efficient in many applications due to their trade-off between computational complexity and excellent convergence properties. In particular, we propose an efficient step-size selection for the interesting case where the Grassmannian is equal to the complex projective space. Moreover, we outline how the developed method can be employed for blind identification.

The paper is organized as follows. Section 2 recalls some basic concepts in differential geometry, which make the present work intuitive and self-contained. An abstract framework of conjugate gradient methods on smooth manifolds is given in Section 3. In Section 4, the geometry of the Grassmannian is presented, followed by a detailed analysis of the of the Karcher mean function in Section 5. A geometric CG algorithm is given in Section 6 for the computation of the Karcher mean on the Grassmannian in general, together with a particularly efficient step-size selection for the special case of the complex projective space. In Section 7, we outline how the proposed approach of averaging subspaces is evidenced to be useful in blind identification and a conclusion is drawn in Section 8.

2. Differential geometric concepts

In this section, we shortly recall and explain the differential geometric concepts that are needed for this work. We refer to [19] for a detailed insight into differential and Riemannian geometry and for the formal definitions of the mathematical objects, and to [13] for an introduction of the topic with a focus on matrix manifolds.

Strictly speaking, a *manifold* M is a topological space that can locally be continuously mapped to some linear space, where this map has a continuous inverse. These maps are called *charts*, and since charts are invertible, we can consider the change of two charts around any point in M as a local map from the linear space into itself. M is a *differentiable* or *smooth* manifold, if these maps are smooth for all points in M. Many data sets considered in signal processing are subsets of such a manifold. Important examples are matrix groups, the set of subspaces of fixed dimension, the set of matrices with orthonormal columns (so-called Stiefel manifold), the set of positive definite matrices, etc.

To every point x in the smooth manifold M one can assign a *tangent space*, consisting of all velocities of smooth curves in M that pass x. Formally, we define

$$T_xM := \{\dot{\alpha}_x(0) \,|\, \alpha(t) \subset M, \alpha_x(0) = x\}. \tag{1}$$

Intuitively, T_{xM} contains all possible directions in which one can tangentially pass through x. The elements of T_{xM} are called *tangent vectors* at x.

A *Riemannian manifold* M is a smooth manifold with a scalar product $g_x(\cdot,\cdot)$ assigned to each tangent space T_xM that varies smoothly with x, the so-called *Riemannian metric*. We drop the subscript x if it is clear from the context which tangent space g refers to. The corresponding norm will be denoted by $\|\cdot\|_g$. The Riemannian metric allows to measure the distance on the manifold. As a natural extension of a straight line in the Euclidean space, a *geodesic* is defined to be a smooth curve in M that connects two sufficiently close points with shortest length. The length of a smooth curve $\alpha : (a,b) \to M$ on a Riemannian manifold is defined as

$$L(\alpha) = \int_a^b \sqrt{g_{\alpha(t)}(\dot{\alpha}(t),\dot{\alpha}(t))}\, \mathrm{d}t. \tag{2}$$

In Euclidean space, two velocities at different locations are both vectors in this space. This allows to form linear combinations and scalar products of these vectors. In the manifold setting, however, this is not possible, since these velocities are elements in different (tangent) spaces. We hence need a way to identify tangent vectors at $x \in M$ with tangent vectors at $y \in M$ if $x \neq y$. To that end, we assume that there is a unique geodesic in M that connects x and y, say $\gamma(t)$, with $\gamma(0) = x$ and $\gamma(\tau) = y$, being possible if x,y are not too far apart. The *parallel transport* along $\gamma(t)$ admits one way of identifying T_xM with T_{yM}. A rigorous definition is beyond the scope of this work, but loosely speaking, the transportation is done in such a way that during the transportation process, there is no intrinsic rotation of the transported vector. In particular, this leaves the scalar product between the transported vector and the velocity of the curve invariant.

Certainly, such an identification of tangent vectors depends on the geodesic. Consider for example a sphere with two different geodesics connecting the south with the north pole (i.e. two meridians) that leave the south pole by an angle of $\pi/2$. Parallel transporting the same vector along both meridians from the south pole to the north will result in two antiparallel vectors at the north pole. Note that the identification of different tangent spaces via parallel transport along a geodesic is just *one* particular instance of a more general concept termed *vector transport* in [13].

In order to minimize a real valued function on M, we have to extend the notion of a gradient to the Riemannian manifold setting. To that end, recall that if $f : \mathbb{R}^n \to \mathbb{R}$ is smooth in x, there is a unique vector $G(x)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x+tH)\big|_{t=0} = G(x)^{\mathsf{T}} H =: \langle G(x), H \rangle_{\text{Euclid}} \quad \text{for all } H \in \mathbb{R}^n, \tag{3}$$

where $(\cdot)^{\mathsf{T}}$ denotes transpose. Typically, we write $\nabla f(x) := G(x)$ and call it the *gradient* of f at x. This coordinate free definition of a gradient can be straightforwardly adapted to the manifold case. Let

$$f : M \to \mathbb{R} \tag{4}$$