Cayley Transform on Stiefel manifolds[☆]

Enrique Macías-Virgós[a,*], María José Pereira-Sáez[b], Daniel Tanré[c]

[a] *Departamento de Matemáticas, Universidade de Santiago de Compostela, 15782 Spain;*
[b] *Facultade de Economía e Empresa, Universidade da Coruña, 15071 Spain;*
[c] *Département de Mathématiques, Université de Lille 1, 59655 Villeneuve d'Ascq Cedex, France*

Abstract

The Cayley transform for orthogonal groups is a well known construction with applications in real and complex analysis, linear algebra and computer science. In this work, we construct Cayley transforms on Stiefel manifolds. Applications to the Lusternik-Schnirelmann category and optimisation problems are presented.

Keywords: Stiefel manifold, Cayley transform, Lusternik-Schnirelmann category, Optimisation

2010 MSC: 53C30, 15A16, 55M30

1. Introduction

Denote by $\mathbb{K}$ the algebra of either the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$ or the quaternions $\mathbb{H}$.

Let $\mathrm{G}(n) = \mathrm{O}(n, \mathbb{K})$ be the Lie group of matrices $A \in \mathbb{K}^{n\times n}$ such that $AA^* = I_n$, where $A^* = \overline{A^t}$ is the conjugate transpose. Depending on $\mathbb{K}$ this group corresponds to the orthogonal group $\mathrm{O}(n)$, the unitary group $\mathrm{U}(n)$ or the symplectic group $\mathrm{Sp}(n)$.

[☆] This research was supported by Xunta de Galicia 2015-PG006, by ESF ACAT Grant 6442 and by the Labex "CEMPI" (ANR-11-LABX-0007-01) which gave us the opportunity to work together. The three authors are also partially supported by the MINECO and FEDER research project MTM2016-78647-P.

[*] Corresponding author

Email addresses: quique.macias@usc.es (Enrique Macías-Virgós), maria.jose.pereira@udc.es (María José Pereira-Sáez), daniel.tanre@univ-lille1.fr (Daniel Tanré)

Let $I \in \mathbb{K}^{n\times n}$ be the identity matrix. The classical Cayley transform [10, Lecture 3] $c_I \colon \Omega(I) \to \Omega(I)$ is defined by

$$c_I(X) = (I - X)(I + X)^{-1}$$

with

$$\Omega(I) = \{X \in \mathbb{K}^{n\times n} \mid (I + X)^{-1} \text{ exists}\}.$$

This map satisfies the equality $c_I^2 = \mathrm{id}$. Moreover c_I induces a diffeomorphism between the tangent space

$$T_I\mathrm{G}(n) = \{X \in \mathbb{K}^{n\times n} \mid X + X^* = 0\}$$

and $\Omega(I) \cap \mathrm{G}(n)$. This construction was generalized by A. Gómez-Tato and the first two authors [4] to any $A \in \mathrm{G}(n)$ as a map $c_A \colon \Omega(A) \to \Omega(A^*)$ defined by

$$c_A(X) = (I - A^*X)(A + X)^{-1} = c_I(A^*X)A^*, \tag{1}$$

with

$$\Omega(A) = \left\{X \in \mathbb{K}^{n\times n} \mid (A + X)^{-1} \text{ exists}\right\}.$$

In this case we have $c_A^{-1} = c_{A^*}$ and there is a diffeomorphism between the tangent space

$$T_A\mathrm{G}(n) = \{X \in \mathbb{K}^{n\times n} \mid A^*X + X^*A = 0\} \tag{2}$$

and $\Omega(A^*) \cap \mathrm{G}(n)$.

The Cayley transform for orthogonal groups is a well known construction with applications in real and complex analysis, linear algebra and computer science. In this work, we construct Cayley transforms on Stiefel manifolds. Applications to the Lusternik-Schnirelmann category and optimisation problems are presented.

We first specify some conventions and notations in use in this paper and state our main results.

Let $0 \leq k \leq n$. The compact Stiefel manifold $\mathrm{O}_{n,k}$ of orthonormal k-frames in $\mathbb{K}^n$ is the set of matrices $x \in \mathbb{K}^{n\times k}$ such that $x^*x = I_k$. This manifold appears also as the base space of the principal bundle

$$\mathrm{G}(n-k) \xrightarrow{\iota} \mathrm{G}(n) \xrightarrow{\rho} \mathrm{O}_{n,k},$$

where $\iota(B) = \begin{bmatrix} B & 0 \\ 0 & I_k \end{bmatrix}$ and ρ is the projection onto the last k columns. If $A \in \mathrm{G}(n)$ and $x = \rho(A)$, we denote by $\rho_{*A} \colon T_A\mathrm{G}(n) \to T_x\mathrm{O}_{n,k}$ the map induced between the tangent spaces.

The next statement contains the existence and the main properties of a Cayley transform in Stiefel manifolds.

Theorem 1.1. *Let $0 \leq k \leq n$ and $x = \begin{bmatrix} T \\ P \end{bmatrix} \in \mathrm{O}_{n,k}$ with $P \in \mathbb{K}^{k\times k}$. We choose $A = \begin{bmatrix} \alpha & T \\ \beta & P \end{bmatrix} \in \mathrm{G}(n)$. Then there exists a map $\gamma^A \colon T_x\mathrm{O}_{n,k} \to \mathrm{O}_{n,k}$, which is called a* Cayley transform for the Stiefel manifold*, such that $\gamma^A \circ \rho_{*A} = \rho \circ c_A$. Moreover we have the following properties:*

1. *The map γ^A is injective on the open subset*

$$\Gamma^x = \left\{ v = A \begin{bmatrix} X \\ Y \end{bmatrix} \in T_x\mathrm{O}_{n,k} \mid (\beta X + P)^{-1} \text{ exists} \right\}.$$

This subset Γ^x does not depend on the choice of A such that $\rho(A) = x$. Furthermore, if γ^A is injective on an open subset $U \subset T_x\mathrm{O}_{n,k}$ then we have $U \subset \Gamma^x$.

2. *The map γ^A induces a diffeomorphism between $\Gamma^x \subset T_x\mathrm{O}_{n,k}$ and the open subset*

$$\Omega^x = \left\{ \begin{bmatrix} \tau \\ \pi \end{bmatrix} \in \mathrm{O}_{n,k} \mid (\pi + P^*)^{-1} \text{ exists} \right\}.$$

An explicit formula for γ^A is given in Definition 2.3. Also, the expression of the inverse map $({\gamma^A}_{|\Gamma_x})^{-1}$ appears in Equations (13) and (14).

As we said before for the group $\mathrm{G}(n)$, the Cayley transform $c_A \colon T_A\mathrm{G}(n) \to \Omega(A^*) \cap \mathrm{G}(n)$ is a diffeomorphism. Therefore the Cayley open subset $\Omega(A^*) \cap \mathrm{G}(n)$ is contractible. This property cannot be extended as it stands to the case of a Stiefel manifold. However the image of the injectivity domain of a Cayley transform in $\mathrm{O}_{n,k}$ is contractible in $\mathrm{O}_{n,k}$.

Theorem 1.2. *For every $x \in \mathrm{O}_{n,k}$ the open subset Ω^x is contractible in $\mathrm{O}_{n,k}$.*

This property is a consequence of the existence of a local section (see Proposition 4.2) $s^A\colon \Omega^x \to \mathrm{G}(n)$ of the projection $\rho\colon \mathrm{G}(n) \to \mathrm{O}_{n,k}$ and the contractibility of the Cayley open subsets $\Omega(A^*) \cap \mathrm{G}(n)$.

The contents of the paper are as follows. Section 2 contains the construction of the Cayley transform $\gamma^A\colon T_x\mathrm{O}_{n,k} \to \mathrm{O}_{n,k}$. The study of the injectivity of its derivative is done in Section 3 and the proofs of Theorems 1.1 and 1.2 occupy Section 4. Finally, Section 5 is devoted to applications of this construction to the Lusternik-Schnirelmann category of quaternionic Stiefel manifolds and to optimisation problems on real Stiefel manifolds.

2. Construction

Let $\mathbb{K}^n$ be either the real vector space $\mathbb{R}^n$, the complex vector space $\mathbb{C}^n$ or the quaternionic vector space $\mathbb{H}^n$ (with the structure of a right $\mathbb{H}$-vector space) endowed with the inner product $\langle u, v\rangle = u^*v$. Let $0 \le k \le n$. The compact Stiefel manifold $\mathrm{O}_{n,k}$ of orthonormal k-frames in $\mathbb{K}^n$ is the set of matrices $x \in \mathbb{K}^{n\times k}$ such that $x^*x = I_k$. It is standard to denote $\mathrm{O}_{n,k}$ by $\mathrm{V}_{n,k}$ in the real case, $\mathrm{W}_{n,k}$ in the complex case and $\mathrm{X}_{n,k}$ in the quaternionic case.

Usually we shall write $x = \begin{bmatrix} T \\ P \end{bmatrix} \in \mathrm{O}_{n,k}$, with $T \in \mathbb{K}^{(n-k)\times k}$ and $P \in \mathbb{K}^{k\times k}$. The linear left action of $\mathrm{G}(n)$ on $\mathrm{O}_{n,k}$ is transitive and the isotropy group of $x_0 = \begin{bmatrix} 0 \\ I_k \end{bmatrix}$ is isomorphic to $\mathrm{G}(n-k)$. Therefore $\mathrm{O}_{n,k}$ is diffeomorphic to $\mathrm{G}(n)/\mathrm{G}(n-k)$ and we have the principal bundle $\mathrm{G}(n-k) \xrightarrow{\iota} \mathrm{G}(n) \xrightarrow{\rho} \mathrm{O}_{n,k}$.

Let $x \in \mathrm{O}_{n,k}$. We complete x to a matrix $A \in \mathrm{G}(n)$ such that $\rho(A) = x$. The tangent space $T_I\mathrm{G}(n)$ of the group $\mathrm{G}(n)$ at the identity is the set of skew-Hermitian (skew-symmetric in the real case) matrices, so the tangent space $T_A\mathrm{G}(n)$ at A given in (2) equals $A{\cdot}T_I\mathrm{G}(n)$. On the other hand, by differentiating the condition $x^*x = I_k$ we obtain that the tangent space to the Stiefel manifold at the point x is given by

$$T_x\mathrm{O}_{n,k} = \{v \in \mathbb{K}^{n\times k} \mid v^*x + x^*v = 0\} = A \cdot T_{x_0}\mathrm{O}_{n,k},$$

where $x_0 = \begin{bmatrix} 0 \\ I_k \end{bmatrix} = \rho(I_n)$. So each tangent vector $v \in T_x\mathrm{O}_{n,k}$ can be written as $v = A\begin{bmatrix} X \\ Y \end{bmatrix}$, with $X \in \mathbb{K}^{(n-k)\times k}$, $Y \in \mathbb{K}^{k\times k}$ and $Y + Y^* = 0$. Thanks to the principal bundle defining the Stiefel manifold, the tangent space $T_x\mathrm{O}_{n,k}$ can be identified with the orthogonal complement $(T_A\mathrm{G}(n-k))^\perp$ of the tangent space image of the inclusion of $\mathrm{G}(n-k)$ in $\mathrm{G}(n)$,

$$(T_A\mathrm{G}(n-k))^\perp = \big\{A\begin{bmatrix} 0 & X \\ -X^* & Y \end{bmatrix} \mid Y + Y^* = 0\big\} \cong T_x\mathrm{O}_{n,k},$$

where $\rho(A) = x$. With this identification, the tangent space

$$T_x\mathrm{O}_{n,k} = \big\{A\begin{bmatrix} X \\ Y \end{bmatrix} \mid Y + Y^* = 0\big\}$$

is considered as a subspace of $T_A\mathrm{G}(n)$ and we may apply the Cayley map c_A of $\mathrm{G}(n)$ on it. From Equation (1) we have

$$c_A = R_{A^*} \circ c_I \circ L_{A^*}, \tag{3}$$

where L_{A^*} and R_{A^*} denote as usual the left and right multiplications in a Lie group. Thus, we have first to determine c_I on the elements of $(T_I\mathrm{G}(n-k))^\perp$.

Lemma 2.1. *Let $X \in \mathbb{K}^{(n-k)\times k}$ and $Y \in \mathbb{K}^{k\times k}$ such that $Y + Y^* = 0$. Then the matrix $I_k + X^*X + Y$ is invertible.*

Proof. The hermitian part of $M = I_k + X^*X + Y$ is

$$H = I_k + X^*X + (Y + Y^*)/2 = I_k + X^*X,$$

which is a positive-definite matrix. Let $A = M - H$ be the skew-hermitian part. If there exists a non-zero vector $w \neq 0$ such that $Mw = 0$, then $Hw = -Aw$, so for the hermitian product $\langle u,v\rangle = u^*v$ we have

$$\begin{aligned}\langle w, Hw\rangle =&\langle w, -Aw\rangle = \langle A^*w, -w\rangle\\ =&\langle -Aw, -w\rangle = \langle Hw, -w\rangle = \langle w, -H^*w\rangle = -\langle w, Hw\rangle,\end{aligned}$$

which is a contradiction because $\langle w, Hw\rangle > 0$. □

We denote

$$b = (I_k + X^*X + Y)^{-1}. \tag{4}$$

Proposition 2.2. *Let* $M = \begin{bmatrix} 0 & X \\ -X^* & Y \end{bmatrix} \in T_I\mathrm{G}(n)$*, then we have*

$$c_I(M) = \begin{bmatrix} I_{n-k} - 2XbX^* & -2Xb \\ 2bX^* & -I_k + 2b \end{bmatrix}. \tag{5}$$

Proof. The skew-symmetric matrix $M = \begin{bmatrix} 0 & X \\ -X^* & Y \end{bmatrix}$ cannot have real eigenvalues, so $I_n + M = \begin{bmatrix} I_{n-k} & X \\ -X^* & I_k + Y \end{bmatrix}$ is invertible. From the following product,

$$\begin{bmatrix} I_{n-k} & 0 \\ X^* & I_k \end{bmatrix} \cdot (I_n + M) = \begin{bmatrix} I_{n-k} & X \\ 0 & X^*X + I_k + Y \end{bmatrix}, \tag{6}$$

we have

$$(I_n + M)^{-1} = \begin{bmatrix} I_{n-k} & X \\ 0 & b^{-1} \end{bmatrix}^{-1} \begin{bmatrix} I_{n-k} & 0 \\ X^* & I_k \end{bmatrix} = \begin{bmatrix} I_{n-k} & -Xb \\ 0 & b \end{bmatrix} \begin{bmatrix} I_{n-k} & 0 \\ X^* & I_k \end{bmatrix}.$$

By applying the definition of $c_I \colon T_I\mathrm{G}(n) \to \mathrm{G}(n)$, we get:

$$\begin{aligned} c_I(M) &= (I_n - M)\,(I_n + M)^{-1} \\ &= \begin{bmatrix} I_{n-k} & -X \\ X^* & I_k - Y \end{bmatrix} (I_n + M)^{-1} = \begin{bmatrix} I_{n-k} - 2XbX^* & -2Xb \\ 2bX^* & -I_k + 2b \end{bmatrix}. \end{aligned}$$

□

Remember that $A = \begin{bmatrix} \alpha & T \\ \beta & P \end{bmatrix} \in \mathrm{G}(n)$ is the matrix such that $\rho(A) = x \in \mathrm{O}_{n,k}$. A computation from (3) and (5) gives directly:

$$c_A\,(AM) = \begin{bmatrix} (I_{n-k} - 2XbX^*)\alpha^* - 2XbT^* & (I_{n-k} - 2XbX^*)\beta^* - 2XbP^* \\ 2bX^*\alpha^* + (-I_k + 2b)T^* & 2bX^*\beta^* + (-I_k + 2b)P^* \end{bmatrix}.$$

The Cayley transform is now obtained by projecting this expression on $\mathrm{O}_{n,k}$.

Definition 2.3. Let $x = \begin{bmatrix} T \\ P \end{bmatrix} \in \mathrm{O}_{n,k}$. We choose $A = \begin{bmatrix} \alpha & T \\ \beta & P \end{bmatrix} \in \mathrm{G}(n)$, and consider $v = A \begin{bmatrix} X \\ Y \end{bmatrix} \in T_x\mathrm{O}_{n,k}$. The *Cayley transform on the Stiefel manifold*, $\gamma^A \colon T_x\mathrm{O}_{n,k} \to \mathrm{O}_{n,k}$, is defined by

$$\begin{aligned} \gamma^A(v) &= \begin{bmatrix} (I_{n-k} - 2XbX^*)\beta^* - 2XbP^* \\ 2bX^*\beta^* + (-I_k + 2b)P^* \end{bmatrix} \\ &= 2 \begin{bmatrix} -Xb \\ b \end{bmatrix} (\beta X + P)^* + \begin{bmatrix} \beta^* \\ -P^* \end{bmatrix}, \end{aligned} \tag{7}$$

where b is given in Equation (4).

Remark 1. The map γ^A depends on the choice of A such that $\rho(A) = x$. With the previous notation, the elements of $\mathrm{G}(n)$ that are sent onto x are the matrices

$$A\bullet E := A \begin{bmatrix} E & 0 \\ 0 & I_k \end{bmatrix} = \begin{bmatrix} \alpha E & T \\ \beta E & P \end{bmatrix}$$

with $E \in \mathrm{G}(n-k)$. We observe

$$v = A \begin{bmatrix} X \\ Y \end{bmatrix} = (A\bullet E) \begin{bmatrix} E^*X \\ Y \end{bmatrix}$$

and

$$I_k + (E^*X)^*(E^*X) + Y = I_k + X^*X + Y.$$

Thus, in (7), if we replace X by E^*X, β by βE and keep unchanged b and Y, we get

$$\gamma^{A\bullet E}(v) = \begin{bmatrix} E^* & 0 \\ 0 & I_k \end{bmatrix} \gamma^A(v).$$

We end this section by noticing that the behavior of γ^A is different from that of the Cayley transform c_A in $\mathrm{G}(n)$. For instance, when $n-k \geq k$, if we choose $x = \begin{bmatrix} T \\ 0 \end{bmatrix}$ and $v = A \begin{bmatrix} 0 \\ Y \end{bmatrix}$, we have $\gamma^A(v) = \begin{bmatrix} \beta^* \\ 0 \end{bmatrix}$, which does not depend on Y. Thus γ^A is not injective on the tangent space $T_x\mathrm{O}_{n,k}$. We address the determination of a domain of injectivity for γ^A in Section 4 but, before that, we study the differential of γ^A.

3. Differential

The results of this section are used in the study of the domain of injectivity of the Cayley transform γ^A.

Let $x = \begin{bmatrix} T \\ P \end{bmatrix} \in \mathrm{O}_{n,k}$ and $A = \begin{bmatrix} \alpha & T \\ \beta & P \end{bmatrix} \in \mathrm{G}(n)$, as before. Let $v_0 = \begin{bmatrix} X \\ Y \end{bmatrix} \in T_{x_0}\mathrm{O}_{n,k}$. The differential of γ^A as a map $\gamma^A \colon T_x\mathrm{O}_{n,k} \to \mathbb{K}^{n\times k}$ at the point $v = Av_0 \in T_x\mathrm{O}_{n,k}$ is denoted by

$$(\gamma^A)_{*v} \colon T_vT_x\mathrm{O}_{n,k} \cong T_x\mathrm{O}_{n,k} \to \mathbb{K}^{n\times k}.$$

We compute $(\gamma^A)_{*v}(w)$ for any $w = A \begin{bmatrix} M \\ N \end{bmatrix} \in T_x\mathrm{O}_{n,k}$, that is, $N + N^* = 0$. Since, with the identification $T_x\mathrm{O}_{n,k} \cong (T_A\mathrm{G}(n-k))^\perp$, we have

$$\gamma^A = \rho \circ R_{A^*} \circ c_I \circ L_{A^*}, \tag{8}$$

the differential of γ^A is determined by that of c_I. Therefore, we first consider

$$(c_I)_{*v_0} \begin{bmatrix} M \\ N \end{bmatrix} = \frac{d}{dt}_{|t=0} c_I \left(\begin{bmatrix} X \\ Y \end{bmatrix} + t \begin{bmatrix} M \\ N \end{bmatrix} \right).$$

Let

$$b_t^{-1} := I_k + (X+tM)^*(X+tM) + Y + tN.$$

Its derivative $\frac{d}{dt}_{|t=0} b_t^{-1}$ is denoted ξ and equals

$$\xi := X^*M + M^*X + N. \tag{9}$$

Moreover $b_0^{-1} = b^{-1}$. From $b_t b_t^{-1} = I$, we deduce $b_0' = -b\xi b$. Then, a direct computation from Equation (5) gives

$$(c_I)_{*v_0} \begin{bmatrix} M \\ N \end{bmatrix} = \begin{bmatrix} -2MbX^* + 2Xb\xi bX^* - 2XbM^* & -2Mb + 2Xb\xi b \\ -2b\xi bX^* + 2bM^* & -2b\xi b \end{bmatrix}. \qquad (10)$$

Proposition 3.1. *With the previous notations, the differential of the Cayley transform of the Stiefel manifold* $\mathrm{O}_{n,k}$ *is given by*

$$(\gamma^A)_{*v}(w) = \begin{bmatrix} (-2MbX^* + 2Xb\xi bX^* - 2XbM^*)\beta^* + (-2Mb + 2Xb\xi b)P^* \\ (-2b\xi bX^* + 2bM^*)\beta^* - 2b\xi bP^* \end{bmatrix}. \qquad (11)$$

Proof. The equality (8) gives by the chain rule $(\gamma^A)_{*v}(w) = \rho\left((c_I)_{*v_0}(w_0) \cdot A^*\right)$, where $w_0 = \begin{bmatrix} M \\ N \end{bmatrix}$ and $A^* = \begin{bmatrix} \alpha^* & \beta^* \\ T^* & P^* \end{bmatrix}$, because the projection ρ and the translations R_{A^*} and L_{A^*} are linear maps. Then formula (10) gives the value (11). □

Proposition 3.2. *The differential* $(\gamma^A)_{*v}$ *is injective if and only if the matrix* $\beta X + P$ *is invertible.*

Proof. According to (11), the kernel of $(\gamma^A)_{*v}$ is the space of solutions $\begin{bmatrix} M \\ N \end{bmatrix}$ of the system

$$\begin{cases} \text{(i)} & 2Mb(\beta X + P)^* = 2Xb\xi b(\beta X + P)^* - 2XbM^*\beta^*, \\ \text{(ii)} & \xi b(\beta X + P)^* = M^*\beta^*, \end{cases}$$

where we have used that the matrix b is invertible. Then we get

$$Mb(X^*\beta^* + P^*) = 0,$$

so the first system is equivalent to

$$\begin{cases} \text{(iii)} & Mb(\beta X + P)^* = 0, \\ \text{(iv)} & \xi b(\beta X + P)^* = (\beta M)^*. \end{cases}$$

If we suppose the matrix $\beta X + P$ is invertible, then the equation (iii) gives $M = 0$ and the equation (iv) gives $\xi = 0$. Finally, from the definition of ξ in Equation (9) we have $M = N = 0$.

Conversely, we suppose the kernel of $\beta X + P$ is non-trivial and we look for an element in the kernel of $(\gamma^A)_{*v}$ of the particular type $M = 0$. In this case, the equation (iii) is trivially satisfied and the equation (iv) may be reduced to $Nb(\beta X + P)^* = 0$. We consider the singular value decomposition of $b(\beta X + P)^* \in \mathbb{K}^{k\times k}$ (for the quaternionic case see [15]):

$$b(\beta X + P)^* = \mu \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \nu^*, \quad \mu, \nu \in \mathrm{G}(k),$$

where Z_1 is a diagonal matrix with positive entries in the diagonal and $Z_2 = 0 \in \mathbb{K}^{r\times r}$. As $\beta X + P$ is not invertible, we have $r > 0$. The existence of solutions in the equation $Nb(\beta X + P)^* = 0$ is then equivalent to the existence of solutions in

$$N\mu \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} = 0 \iff \mu^* N\mu \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} = 0.$$

The fact that $r > 0$ allows the choice of a non-zero, skew-symmetric matrix $\mu^* N\mu$ satisfying the last equation. Thus $N \neq 0$ is skew-symmetric and $(\gamma^A)_{*v} \begin{bmatrix} 0 \\ N \end{bmatrix} = 0$. □

4. Properties

This section consists of the proof of Theorems 1.1 and 1.2. Recall the notations $x = \begin{bmatrix} T \\ P \end{bmatrix} \in \mathrm{O}_{n,k}$ with $T \in \mathbb{K}^{(n-k)\times k}$, $P \in \mathbb{K}^{k\times k}$ and the choice of $A = \begin{bmatrix} \alpha & T \\ \beta & P \end{bmatrix} \in \mathrm{G}(n)$.

Proof of Theorem 1.1.

1) First, we look at the independence of Γ^x on the choice of A. With the notations of Remark 1, any matrix projecting onto x can be written as $A\bullet E = \begin{bmatrix} \alpha E & T \\ \beta E & P \end{bmatrix}$ with $E \in \mathrm{G}(n-k)$. An element $v \in T_x\mathrm{O}_{n,k}$ may be expressed as

$$v = A \begin{bmatrix} X \\ Y \end{bmatrix} = (A\bullet E) \begin{bmatrix} E^* X \\ Y \end{bmatrix}.$$

The fact that Γ^x does not depend on the choice of A comes from $(\beta E)(E^*X) + P = \beta X + P$.

As for the injectivity, let $v_1 = A \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}$ and $v_2 = A \begin{bmatrix} X_2 \\ Y_2 \end{bmatrix}$ be two vectors of Γ^x. From Definition 2.3, the equality $\gamma^A(v_1) = \gamma^A(v_2)$ is equivalent to the system

$$\begin{cases} \text{(i)} & X_1 b_1 (\beta X_1 + P)^* = X_2 b_2 (\beta X_2 + P)^*, \\ \text{(ii)} & b_1 (\beta X_1 + P)^* = b_2 (\beta X_2 + P)^*, \end{cases}$$

from which we deduce

$$(X_1 - X_2) b_1 (\beta X_1 + P)^* = 0. \tag{12}$$

As $v_1 \in \Gamma^x$ means that $\beta X_1 + P$ is invertible, we get from (12) the equality $X_1 = X_2$. Then the equation (ii) implies $b_1 = b_2$ and, from (4), we deduce that $Y_1 = Y_2$ and the injectivity of γ^A on Γ^x.

Conversely, suppose γ^A invertible on an open subset U. This implies the injectivity of the differential $(\gamma^A)_{*v}$ for any $v \in U$ and Proposition 3.1 gives the inclusion $U \subset \Gamma^x$.

2) The values of τ and π such that $\gamma^A(v) = \begin{bmatrix} \tau \\ \pi \end{bmatrix}$ are given by (7). Let $\begin{bmatrix} \tau \\ \pi \end{bmatrix}$ such that $\pi + P^*$ is invertible. We are looking for matrices $X \in \mathbb{K}^{(n-k)\times k}$ and $Y \in \mathbb{K}^{k \times k}$, with Y skew-symmetric, such that $\beta X + P$ is invertible and the following system, which is equivalent to Definition 2.3, is satisfied:

$$\begin{cases} \text{(i)} & \tau - \beta^* = -2Xb(\beta X + P)^*, \\ \text{(ii)} & \pi + P^* = 2b(\beta X + P)^*. \end{cases}$$

In particular, we have from (ii) that the matrix $\pi + P^*$ is invertible if and only if $\beta X + P$ is so. From (a) we get the value of X,

$$X = -(\tau - \beta^*)(\pi + P^*)^{-1}. \tag{13}$$

Also from (ii) we obtain

$$b = \frac{1}{2}(\pi + P^*)\left[(\beta X + P)^*\right]^{-1}$$

and the expression of Y follows from the fact that Y is the skew-symmetric part of b^{-1}, that is,

$$2Y = b^{-1} - (b^{-1})^*. \tag{14}$$

Here we do not need to give an explicit expression. If we replace those values in (7), we get $\gamma^A(v) = \begin{bmatrix} \tau \\ \pi \end{bmatrix}$, so we have proved the existence of a right inverse to the map $\gamma^A \colon \Gamma^x \to \Omega^x$. Since γ^A is injective we obtain the desired result. □

Remark 2. For any Stiefel manifold it is possible to prove that the domain of injectivity Γ^x is not the whole vector space $T_x\mathrm{O}_{n,k}$.

Definition 4.1. For any $x = \begin{bmatrix} T \\ P \end{bmatrix} \in \mathrm{O}_{n,k}$, the open subset

$$\Omega^x = \left\{ \begin{bmatrix} \tau \\ \pi \end{bmatrix} \in \mathrm{O}_{n,k} \mid \pi + P^* \text{ invertible} \right\}$$

is called *a Cayley open subset of the Stiefel manifold.*

We continue with an explicit trivialization of the fibration ρ over each Cayley open set.

Proposition 4.2. *Let* $x = \begin{bmatrix} T \\ P \end{bmatrix} \in \mathrm{O}_{n,k}$ *and let* Ω^x *be the open subset given in Definition 4.1. Then the projection* $\rho \colon \mathrm{G}(n) \to \mathrm{O}_{n,k}$ *admits a local section* $s^A \colon \Omega^x \to \mathrm{G}(n)$.

Proof. With the identification $T_x\mathrm{O}_{n,k} \cong (T_A\mathrm{G}(n-k))^\perp$, and from the definition of γ^A we can write $\gamma^A = \rho \circ c_A$ on Γ^x. Moreover, we have proved in Theorem 1.1 that the restriction ${\gamma^A}_{|\Gamma^x} \colon \Gamma^x \to \Omega^x$ is a diffeomorphism whose inverse is denoted $({\gamma^A}_{|\Gamma^x})^{-1}$. We set

$$s^A = c_A \circ ({\gamma^A}_{|\Gamma^x})^{-1} \colon \Omega^x \to \mathrm{G}(n)$$

and verify

$$\rho \circ s^A = \rho \circ c_A \circ ({\gamma^A}_{|\Gamma^x})^{-1} = {\gamma^A}_{|\Gamma^x} \circ ({\gamma^A}_{|\Gamma^x})^{-1} = \mathrm{id}_{\Omega^x}.$$

Notice that $s^A(\Omega^x) \subset \Omega(A^*) \cap \mathrm{G}(n)$. □

An explicit formula for s^A could be obtained from those of c_A and $(\gamma^A{}_{|\Gamma^x})^{-1}$.

Proof of Theorem 1.2. We choose $A = \begin{bmatrix} \alpha & T \\ \beta & P \end{bmatrix} \in \mathrm{G}(n)$. Let $s^A \colon \Omega^x \to \mathrm{G}(n)$ be the local section of Proposition 4.2. With the notations of the statement, we consider the map $H \colon \Omega^x \times [0,1] \to \mathrm{O}_{n,k}$ defined by

$$H(y,t) = \rho(c_A(tc_{A^*}(s^A(y)))).$$

This map verifies $H(y,0) = \rho(c_A(0)) = \rho(A^*) = \begin{bmatrix} \beta^* \\ P^* \end{bmatrix} = \gamma^A(0)$ and $H(y,1) = \rho(s^A(y)) = y$. Therefore, it is a contraction of Ω^x on the point $\gamma^A(0)$. □

5. Some applications

5.1. L-S category of some quaternionic Stiefel manifolds

Let $\mathbb{K} = \mathbb{H}$ be the algebra of quaternions, $\mathrm{G}(n) = \mathrm{Sp}(n)$ the symplectic group and $\mathrm{O}_{n,k} = \mathrm{X}_{n,k}$ the quaternionic Stiefel manifold. With the notations of the proof of Theorem 1.2, we observe that, in general, the point $\gamma^A(0)$ does not belong to Ω^x. Therefore, our proof does not imply the contractibility of Ω^x but only its contractibility in $\mathrm{X}_{n,k}$, as stated. This property suffices for our first application.

Recall that the Lusternik-Schnirelmann category (henceforth LS-category) of a topological space X is the least integer $m \geq 0$ such that X admits a covering by $(m+1)$ open sets which are contractible in X, see [2] for more details. We denote it $\operatorname{cat} X$.

The LS-category has applications in a wide range of fields coming from dynamical systems to homotopy theory, but it has also proven to be difficult to determine. For instance, a longstanding problem is the computation of the LS-category of Lie groups. In the case of unitary groups, Singhof [12, 13] proved that $\operatorname{cat} \mathrm{U}(n) = n$ by using an argument based on the eigenvalues. Nevertheless, this method cannot be carried out for the quaternionic group $\mathrm{Sp}(n)$, see [7]. However, some results have been obtained for small n such as $\operatorname{cat} \mathrm{Sp}(2) = 3$

[11] and $\operatorname{cat}\operatorname{Sp}(3) = 5$ [3] and for the determination of some bounds such as $\operatorname{cat}\operatorname{Sp}(n) \geq n+2$ when $n \geq 3$ [5] and $\operatorname{cat}\operatorname{Sp}(n) \leq \binom{n+1}{2}$ [8]. The quaternionic Stiefel manifolds $\mathrm{X}_{n,k}$ are more accessible in certain ranges. For instance, we know that $\operatorname{cat}\mathrm{X}_{n,k} = k$ when $n \geq 2k$. In order to prove this, Nishimoto [9] uses the number of eigenvalues of a complex matrix in a way similar to Singhof's approach. This has also been established by Kadzisa and Mimura [6] using certain Morse-Bott functions on $\mathrm{X}_{n,k}$. In the next proposition, we give a short proof of this result with the Cayley open subsets of Definition 4.1.

Proposition 5.1 ([9, 6]). *If $n \geq 2k$, we have $\operatorname{cat}\mathrm{X}_{n,k} \leq k$.*

Proof. Let $\theta \in]0,\pi/2[$ and take $x_\theta = \begin{bmatrix} T_\theta \\ P_\theta \end{bmatrix} \in \mathrm{X}_{n,k}$, with $P_\theta = (\cos\theta)I_k$ and $T_\theta = \begin{bmatrix} 0 \\ (\sin\theta)I_k \end{bmatrix}$. We know from Theorem 1.2 that Ω^{x_θ} is contractible in $\mathrm{X}_{n,k}$. We choose $(k+1)$ numbers θ_i such that $0 < \theta_0 < \theta_2 < \cdots < \theta_k < \pi/2$ and observe that an element $\pi \in \mathbb{H}^{k\times k}$ such that $\pi + P_{\theta_i}$ is not invertible for all i should have $(k+1)$ distinct real eigenvalues. This is impossible and the family $(\Omega^{x_{\theta_i}})_{0\leq i\leq k}$ is an open cover of $\mathrm{X}_{n,k}$ by subsets contractible in $\mathrm{X}_{n,k}$. □

5.2. Optimisation theory

Let $\mathrm{G}(n) = \mathrm{O}(n)$ be the orthogonal group and $\mathrm{O}_{n,k} = \mathrm{V}_{n,k}$ the real Stiefel manifold. In optimisation theory, the problems with orthogonality constraints are widely known and have concrete applications in many different areas (see [1] for instance). A typical example is looking for k orthogonal n-vectors that are optimal with respect to some parameter f like cost or likehood. These kinds of problems can be seen as optimisation problems on a real Stiefel manifold.

The most popular method of approximate solution is the gradient descent method which can be summarized as follows. Let $x = x_0$ be an initial trial point in the Stiefel manifold $\mathrm{V}_{n,k}$ and let F be the negative gradient of f at x. Then a curve $\alpha(t)$ must be found on the manifold such that $\alpha(0) = x$ and $\alpha'(0) = F$. By fixing a step size τ small enough, the next iterate is obtained by curvilinear search, that is, putting $x_1 = \alpha(\tau)$. Under certain conditions the sequence $x_0, x_1, \ldots$ converges to a local minimun of the function f.

Most existing methods either use matrix factorizations (such as the SVD decomposition) or require the determination of geodesic curves, which is computationally expensive. A different algorithm has been proposed in [14], where the curve is not a geodesic but is constructed from the Cayley transform in the orthogonal group. Specifically, one considers the skew-symmetric matrix $A = Fx^* - xF^*$ and computes the Cayley transform $Q(t) = c_I(tA)$ *on the group* $\mathrm{O}(n)$. Since the group acts on the Stiefel manifold, the desired curve can be given by $\alpha(t) = Q(t)x$. Our construction of a Cayley transform is intrinsic to the Stiefel manifold and should lead to more efficient methods.

Acknowledgements. We thank an anonymous referee for useful comments.

References

[1] Pierre A. Absil, Robert Mahony, and Rodolphe Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, Princeton, NJ, 2008, With a foreword by Paul Van Dooren.

[2] Octav Cornea, Gregory Lupton, John Oprea, and Daniel Tanré, *Lusternik-Schnirelmann category*, Mathematical Surveys and Monographs, vol. 103, American Mathematical Society, Providence, RI, 2003.

[3] Lucía Fernández-Suárez, Antonio Gómez-Tato, Jeffrey Strom, and Daniel Tanré, The Lusternik-Schnirelmann category of Sp(3), *Proc. Amer. Math. Soc.* **132** (2004), no. 2, 587–595.

[4] Antonio Gómez-Tato, Enrique Macías-Virgós, and María José Pereira-Sáez, Trace map, Cayley transform and LS category of Lie groups, *Ann. Global Anal. Geom.* **39** (2011), no. 3, 325–335.

[5] Norio Iwase and Mamoru Mimura, *L-S categories of simply-connected compact simple Lie groups of low rank*, Categorical decomposition techniques in

algebraic topology (Isle of Skye, 2001), Progr. Math., vol. 215, Birkhäuser, Basel, 2004, pp. 199–212.

[6] Hiroyuki Kadzisa and Mamoru Mimura, Morse-Bott functions and the Lusternik-Schnirelmann category, *J. Fixed Point Theory Appl.* **10** (2011), no. 1, 63–85.

[7] Enrique Macías-Virgós and María José Pereira-Sáez, Symplectic matrices with predetermined left eigenvalues, *Linear Algebra Appl.* **432** (2010), no. 1, 347–350.

[8] Enrique Macías-Virgós and María José Pereira-Sáez, An upper bound for the Lusternik-Schnirelmann category of the symplectic group, *Math. Proc. Cambridge Philos. Soc.* **155** (2013), no. 2, 271–276.

[9] Tetsu Nishimoto, On the Lusternik-Schnirelmann category of Stiefel manifolds, *Topology Appl.* **154** (2007), no. 9, 1956–1960.

[10] Mikhail M. Postnikov, *Lie groups and Lie algebras*. Lectures in geometry. Semester V. Mir Publishers, Moscow, 1986.

[11] Paul A. Schweitzer, Secondary cohomology operations induced by the diagonal mapping, *Topology* **3** (1965), 337–355.

[12] Wilhelm Singhof, On the Lusternik-Schnirelmann category of Lie groups, *Math. Z.* **145** (1975), no. 2, 111–116.

[13] Wilhelm Singhof, On the Lusternik-Schnirelmann category of Lie groups II, *Math. Z.* **151** (1976), no. 2, 143–148.

[14] Zaiwen Wen and Wotao Yin, A feasible method for optimization with orthogonality constraints, *Math. Program.* **142** (2013), no. 1-2, Ser. A, 397–434.

[15] Fuzhen Zhang, Quaternions and matrices of quaternions, *Linear Algebra Appl.* **251** (1997), 21–57.