THE CAYLEY TRANSFORM ON LIE GROUPS, SYMMETRIC SPACES AND STIEFEL MANIFOLDS

ENRIQUE MACÍAS-VIRGÓS

Communicated by Vasile Brînzănescu

The Cayley transform for orthogonal groups is a well known construction with applications going from analysis and linear algebra to computer science, physics and biology. This paper surveys joint work with M.J. Pereira Sáez (A Coruña, Spain) and D. Tanré (Lille, France). We explain how to construct Cayley transforms on Lie groups, symmetric spaces and Stiefel manifolds. Several applications to LS-category, Morse-Bott theory and optimization algorithms are discussed.

AMS 2010 Subject Classification: 15A16, 22E15, 58E05, 53C35, 53C30, 55M30.

Key words: Cayley transform, Lie groups, symmetric spaces, Stiefel manifolds, Lusternik-Schnirelmann category, Morse-Bott function, optimization.

1. INTRODUCTION

The classical Cayley transform [13] is a way to express an orthogonal matrix by means of skew-symmetric coordinates. It is given by

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

where the matrix $I + X$ is invertible because all the eigenvalues of the skew-symmetric matrix X have null real part, so they must be different from -1. This transform was discovered by A. Cayley in 1846 [2] and has some advantages over the exponential map; in particular, it equals its own inverse, $c^2 = \mathrm{id}$. It is a well known construction with many applications, going from complex analysis, linear algebra and computer science to nuclear physics or biology.

This talk surveys joint work with M.J. Pereira Sáez (A Coruña, Spain) and D. Tanré (Lille, France). We shall explain how to construct Cayley transforms on orthogonal Lie groups, symmetric spaces and Stiefel manifolds. Most results will be cited without proof. Several applications are discussed, namely:

- Lusternik-Schnirelmann category and topological complexity;
- Morse-Bott theory;
- Optimization.

2. THE CAYLEY TRANSFORM

2.1. Ortogonal groups

Let $\mathbb{K}$ be the algebra of either reals $\mathbb{R}$, complex numbers $\mathbb{C}$ or quaternions $\mathbb{H}$. We denote by $\mathcal{M}(n, \mathbb{K})$ the set of all square $n \times n$ matrices with coefficients in $\mathbb{K}$.

Definition 1. The matrix $A \in \mathcal{M}(n, \mathbb{K})$ is *orthogonal* if $AA^* = I$, where $A^* = \bar{A}^t$ is the conjugate transpose. In the real case $A^* = A^t$ is just the transpose.

Let us denote by $\mathrm{O}(n, \mathbb{K})$ the Lie group of orthogonal matrices. Depending on $\mathbb{K}$ this group corresponds to the (real) orthogonal group $\mathrm{O}(n)$, the (complex) unitary group $\mathrm{U}(n)$ or the (quaternionic) symplectic group $\mathrm{Sp}(n)$.

Remark 2. It would be possible to consider more general J-orthogonal groups $G = \mathrm{O}(n, \mathbb{K}; J)$ defined by the condition $AJA^* = J$, for J an invertible square matrix.

Definition 3 ([5]). Let $A \in \mathrm{O}(n, \mathbb{K})$ be an orthogonal matrix. Denote by $\Omega(A) \subset \mathcal{M}(n, \mathbb{K})$ the open set of matrices X such that $A+X$ is invertible. The *generalized Cayley transform centered at* A is the map

$$c_A \colon \Omega(A) \to \Omega(A^*)$$

given by

$$c_A(X) = (I - A^*X)(A + X)^{-1}.$$

The classical Cayley map corresponds to $A = I$.

The following properties will be useful later.

Proposition 4. *Let $X \in \Omega(A)$. Then*

(1) $c_A(X) = (A + X)^{-1}(I - XA^*)$;
(2) $c_A(X) \in \Omega(A^*)$;
(3) $c_A(X) = c_I(A^*X)A^*$;
(4) $X^* \in \Omega(A^*)$ *and*
$$c_{A^*}(X^*) = c_A(X)^*;$$
(5) $UXU^* \in \Omega(UAU^*)$ *for any matrix* $U \in \mathrm{O}(n, \mathbb{K})$*; moreover,*
$$c_{UAU^*}(UXU^*) = Uc_A(X)U^*;$$
(6) *if the matrix X is invertible then $X^{-1} \in \Omega(A^*)$ and*
$$c_{A^*}(X^{-1}) = -Ac_A(X)A;$$

Corollary 5. *The Cayley transform $c_A \colon \Omega(A) \to \Omega(A^*)$ is a diffeomorphism, with $c_A^{-1} = c_{A^*}$.*

2.2. **Contractibility of the domain**

Our first objective is to prove that the domain of the Cayley transform in an orthogonal group is contractible. This will be important when studying the Lusternik-Schnirelmann category in Section 5.1.

The Lie algebra of $G = \mathrm{O}(n, \mathbb{K})$ is formed by the skew-symmetric (resp. skew-hermitian) matrices,

$$\mathfrak{g} = \{X \in \mathcal{M}(n, \mathbb{K})\colon X + X^* = 0\}.$$

As a vector space $\mathfrak{g}$ equals the tangent space $T_I G$ at the identity, so the tangent space at any other point $A \in G$ is given as

$$T_A G = A \cdot T_I G = \{Y \in \mathcal{M}(n, \mathbb{K})\colon A^* Y + Y^* A = 0\}.$$

We shall denote by $\Omega_G(A)$ the open subset $\Omega(A) \cap G \subset G$ of the matrices $B \in G$ such that $A + B$ is invertible.

THEOREM 6 ([5]). *Let $G = \mathrm{O}(n, \mathbb{K})$ be a compact orthogonal group. The generalized Cayley transform c_A maps diffeomorphically $\Omega_G(A)$ onto $T_{A^*}G$, with $c_A(A) = 0$. As a consequence, the open set $\Omega_G(A)$ is contractible.*

Proof. We sketch the proof.

– First, we must prove that an orthogonal matrix is sent into a skew-symmetric matrix, i.e., $c_A(\Omega_G(A)) \subset T_{A^*}G$, by using the properties of c_A stated in Proposition 4.

– Second, we prove that $T_{A^*}G \subset \Omega(A^*)$: since the eigenvalues of a skew-Hermitian matrix must have null real part, it follows that $A^* + X$ is invertible, for any $X \in T_{A^*}G$.

– Finally, we prove that a skew-symmetric matrix is sent into an orthogonal one, that is, $c_{A^*}(T_{A^*}G) \subset \Omega_G(A)$, by using again the properties of the Cayley transform. □

Example 7. Let z be a unit complex number, $|z| = 1$. Let $G = \mathrm{U}(n)$ be the group of unitary complex matrices. Let $\Omega_G(z)$ be the open set of unitary matrices $A \in \mathrm{U}(n)$ such that $A - zI$ is invertible (i.e., z is not an eigenvalue of A). Then $\Omega_G(z)$ is contractible.

Remark 8. For the J-orthogonal group $G = \mathrm{O}(n, \mathbb{K}; J)$ the Lie algebra is given by

$$\mathfrak{g} = \{X \in \mathcal{M}(n, \mathbb{K})\colon XJ + JX^* = 0\},$$

and the Cyaley transform sends the elements of $G \cap \Omega(A)$ into $\mathfrak{g} \cap \Omega(A^*)$. However, a J-skew-symmetric matrix may have real eigenvalues.

3. SYMMETRIC SPACES

3.1. Cartan model

We extend our results to the most important class of homogeneous spaces, namely, that of symmetric spaces.

Definition 9. Let $\sigma\colon G \to G$ be an involutive automorphism and let

$$K = \{B \in G\colon \sigma(B) = B\}$$

be the closed Lie subgroup of fixed points. The homogeneous space G/K is called a *globally symmetric space.*

Remark 10. In what follows, we shall assume that the automorphism σ is the restriction of an involutive automorphism $\sigma : \mathcal{M}(n, \mathbb{K}) \to \mathcal{M}(n, \mathbb{K})$ of unital algebras. We shall also assume that $\sigma(X^*) = \sigma(X)^*$ for all $X \in \mathcal{M}(n, \mathbb{K})$. These conditions are not too restrictive; for instance, all the compact irreducible Riemannian symmetric spaces in Cartan's classification verify them.

The Lie algebra of K is

$$\mathfrak{k} = \{X \in \mathfrak{g}\colon \sigma(X) = X\}$$

and the tangent space T_oG/K at the base-point $o = [I]$ is isomorphic to

$$\mathfrak{m} = \{X \in \mathfrak{g}\colon \sigma(X) = -X\}.$$

Since $\mathrm{Ad}(k)(\mathfrak{m}) = \mathfrak{m}$ for all $k \in K$, there is an invariant Riemannian metric on G/K which has null torsion and parallel curvature.

Definition 11. The *Cartan embedding* of the symmetric space into the Lie group is the map $\gamma\colon G/K \hookrightarrow G$ given by

$$\gamma([B]) = B\sigma(B)^{-1}.$$

PROPOSITION 12. *Assume that G/K is connected. Then the image $M = \gamma(G/K)$ of the embedding γ is the connected component N_I of the identity of the submanifold*

$$N = \{B \in G\colon \sigma(B) = B^{-1}\}.$$

As a consequence, for each point $A \in M$, the tangent space to the symmetric space can be identified with the vector space

$$T_AM = \{Y \in \mathcal{M}(n, \mathbb{K})\colon YA^* + AY^* = 0,\, \sigma(Y) = -Y^*\}.$$

Example 13. We specialize to *quaternionic Grassmannians.* Fix some $k \leq n$ and consider the matrix

$$J = \mathrm{diag}(-I_k, I_{n-k}).$$

The fixed point set of the automorphism $\sigma\colon \mathrm{Sp}(n) \to \mathrm{Sp}(n)$ defined by

$$\sigma(B) = JBJ$$

is the subgroup $\mathrm{Sp}(k) \times \mathrm{Sp}(n-k)$. Then we obtain the Grassmann manifold of k-planes in $\mathbb{H}^n$,

$$G_{n,k} = \mathrm{Sp}(n)/(\mathrm{Sp}(k) \times \mathrm{Sp}(n-k)).$$

The Cartan embedding $\gamma\colon G_{n,k} \hookrightarrow \mathrm{Sp}(n)$ is given by

$$\gamma([B]) = B\sigma(B)^* = BJB^*,$$

so the Cartan model M is the connected component of the identity of

$$N = \{B \in \mathrm{Sp}(n)\colon JB = B^*J\}.$$

3.2. **Cayley map**

We shall now see that the properties of the Cayley transform in the Lie group G are naturally inherited by the Cartan model of the symmetric space. Let $G = \mathrm{O}(n, \mathbb{K})$ and let $\Omega_G(A)$ be the open set given in Theorem 6.

LEMMA 14. *If $A \in G$ then $c_{\sigma(A)} \circ \sigma = \sigma \circ c_A$ on $\Omega_G(A)$.*

This allows us to prove that the domain of the Cayley transform in a symmetric space is contractible.

THEOREM 15 ([7]). *Let $M \subset G$ be the Cartan model of the symmetric space G/K. Let $A \in M$. Then $\Omega_M(A) := \Omega(A) \cap M$ is a contractible open subspace of M.*

Proof. We know (Theorem 6) that $\Omega_G(A) \cong T_{A^*}G$, so it can be contracted to A by the contraction

$$\nu(X,t) = c_{A^*}(tc_A(X)).$$

We only need to prove that, when $A \in M$ and $X \in M$, the contraction $\nu(X,t)$ remains in M for all $t \in [0,1]$. But since $M = N_I$ (Proposition 12), it suffices to prove that $\nu(X,t) \in N$ for all t, that is,

$$\sigma(\nu(X,t)) = \nu(X,t)^{-1},$$

which follows from the properties of the generalized Cayley transform stated in Proposition 12. □

4. STIEFEL MANIFOLDS

4.1. Cayley transform

We study now another important class of homogeneous spaces: Stiefel manifolds. 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 $\langle u, v\rangle = u^*v$. In the real case this is the standard inner product $v^t w$.

For $0 \leq k \leq n$, the *compact Stiefel manifold* $\mathrm{O}_{n,k}(\mathbb{K})$ 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 $\mathrm{O}_{n,k}(\mathbb{K})$ 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.

We shall write the k-frame as $x = \begin{pmatrix} T \\ P \end{pmatrix} \in \mathrm{O}_{n,k}(\mathbb{K})$, with $T \in \mathbb{K}^{(n-k)\times k}$ and $P \in \mathbb{K}^{k\times k}$. The linear left action of $\mathrm{O}(n,\mathbb{K})$ on $\mathrm{O}_{n,k}(\mathbb{K})$ is transitive and the isotropy group of the base point $x_0 = \begin{pmatrix} 0 \\ I_k \end{pmatrix}$ is isomorphic to $\mathrm{O}(n-k,\mathbb{K})$. Then the manifold $\mathrm{O}_{n,k}(\mathbb{K})$ is diffeomorphic to

$$\mathrm{O}(n,\mathbb{K})/\mathrm{O}(n-k,\mathbb{K})$$

and we have the projection

$$\mathrm{O}(n,\mathbb{K}) \xrightarrow{\rho} \mathrm{O}_{n,k}(\mathbb{K}),$$

given as $\rho(A) = Ax_0$. The Cayley transform in the Stiefel manifold will be obtained by projecting that of $\mathrm{O}(n,\mathbb{K})$.

Let $x \in \mathrm{O}_{n,k}(\mathbb{K})$. Take some $A = \begin{pmatrix} \alpha & T \\ \beta & P \end{pmatrix} \in \mathrm{O}(n,\mathbb{K})$, such that $\rho(A) = x$. It is customary to denote such a matrix A as $(x_\perp\, x)$. The tangent vector space is

$$T_x\mathrm{O}_{n,k} = \{v \in \mathbb{K}^{n\times k} \colon v^*x + x^*v = 0\}.$$

Then any tangent vector $v \in T_x\mathrm{O}_{n,k}(\mathbb{K})$ can be written as $v = A\begin{pmatrix} X \\ Y \end{pmatrix}$, where $X \in \mathbb{K}^{(n-k)\times k}$, $Y \in \mathbb{K}^{k\times k}$ and $Y + Y^* = 0$.

Definition 16 ([9]). The *Cayley transform on the Stiefel manifold*,

$$\gamma^A \colon T_x\mathrm{O}_{n,k}(\mathbb{K}) \to \mathrm{O}_{n,k}(\mathbb{K}),$$

is defined by

$$\gamma^A(v) = \begin{pmatrix} \beta^* \\ -P^* \end{pmatrix} + 2\begin{pmatrix} -X \\ I_k \end{pmatrix}(I_k + X^*X + Y)^{-1}(\beta X + P)^*.$$

The map γ^A depends on the choice of A.

Example 17. If $\mathbb{K} = \mathbb{R}$ and $k = 1$, the Stiefel manifold $V_{n,1}$ is the sphere $S^{n-1} \subset \mathbb{R}^n$. Take for instance the North pole $x = (0, \ldots, 0, 1)^t$ and let be $\beta = (0, \ldots, 0)$ and $\alpha = I_{n-1}$. A vector v tangent to the sphere at x has the form $v = (x_1, \ldots, x_{n-1}, 0)$. The Cayley transform is then given by

$$\gamma(v) = (\frac{-2x_1}{1+\Sigma}, \ldots, \frac{-2x_{n-1}}{1+\Sigma}, \frac{1-\Sigma}{1+\Sigma})^t$$

where $\Sigma = x_1^2 + \cdots + x_{n-1}^2$.

4.2. **Injectivity domain**

THEOREM 18 ([9]). *The Cayley map* $\gamma^A \colon T_x\mathrm{O}_{n,k}(\mathbb{K}) \to \mathrm{O}_{n,k}(\mathbb{K})$ *verifies the following properties:*

(1) $\gamma^A \circ \rho = \rho \circ c_A$.

(2) *The map* γ^A *is injective on the open subset*

$$\Gamma(x) = \left\{ v = A \begin{pmatrix} X \\ Y \end{pmatrix} \in T_x\mathrm{O}_{n,k}(\mathbb{K}) \colon \beta X + P \textit{ is invertible} \right\}.$$

(3) *The map* γ^A *induces a diffeomorphism between* $\Gamma(x) \subset T_x\mathrm{O}_{n,k}(\mathbb{K})$ *and the open subset*

$$\Omega(x) = \left\{ \begin{pmatrix} \tau \\ \pi \end{pmatrix} \in \mathrm{O}_{n,k}(\mathbb{K}) \colon \pi + P^* \textit{ is invertible} \right\}.$$

Remark 19. The subset $\Gamma(x)$ does not depend on the choice of A. Moreover, if γ^A is injective on an open subset $U \subset T_x\mathrm{O}_{n,k}(\mathbb{K})$ then $U \subset \Gamma(x)$.

As we said before, for the group $G = \mathrm{O}(n, \mathbb{K})$, the Cayley transform c_A is a diffeomorphism $T_AG \cong \Omega(A^*) \cap G$. Therefore the open subset $\Omega(A^*) \cap G$ is contractible. This property is no longer true for a Stiefel manifold. However, the image of the injectivity domain of a Cayley transform in $\mathrm{O}_{n,k}(\mathbb{K})$ is contractible in $\mathrm{O}_{n,k}(\mathbb{K})$.

THEOREM 20. *For every* $x \in \mathrm{O}_{n,k}(\mathbb{K})$ *the open subset* $\Omega(x)$ *is contractible in* $\mathrm{O}_{n,k}(\mathbb{K})$.

We omit the proof, that appears in [9].

5. **APPLICATIONS**

5.1. **Lusternik-Schnirelmann category**

Lusternik-Schnirelmann category (in short LS-category) is a homotopical invariant that has been widely studied.

Definition 21. For a topological space X, the *LS-category* $\operatorname{cat} X$ is defined as the minimum number (minus one) of open sets contractible in X which are needed to cover X.

It is important to distinguish subspaces which are contractible into themselves from subspaces which are contractible into the ambient space X. The latter ones are neccessary in the definition above in order to obtain a homotopical invariant. See the book [3] for an introduction to the subject.

Example 22. A contractible space has category zero. A sphere S^n, $n \geq 1$, has category one. The torus T^2 has category two (hint: subspaces contractible in X may not be connected).

Lusternik-Schnirelmann category is important because

- for a connected compact manifold, the LS category (plus one) is a lower bound for the number of critical levels of any smooth function, be Morse or not (*Lusternik-Schnirelmann theorem* [3]).
- for a Lie group, the LS category equals Farber's *topological complexity* [4], a well known invariant with applications in motion planning and robotics.

Unfortunately, the LS-category is very difficult to compute. In what follows we show several spaces where our results are useful.

Theorem 23 ([15]). *For the unitary complex group,* $\operatorname{cat} \mathrm{U}(n)$.

Proof. W. Singhof [15] obtained an explicit covering of $\mathrm{U}(n)$ by the open sets $\Omega_G(z)$, where z is a unitary complex number, as in Example 7. He used the exponential map to prove contractibility. With the Cayley transform we know that these sets are diffeomorphic to the Lie algebra $\mathfrak{u}(n)$. It is then easy to find an explicit covering of the group by $n+1$ open sets because any matrix in $\mathrm{U}(n)$ has at most n different eigenvalues. □

Remark 24. It is important to note that the existence of a given covering by $n+1$ open subsets which are contractible in the ambient space $\mathrm{U}(n)$ only guarantees that $\operatorname{cat} \mathrm{U}(n) \leq n$. For the equality one needs a lower bound for the LS-category, which in all the cases we are considering is given by the so-called *cup product length* [3].

For the symplectic group, the result $\operatorname{cat} \mathrm{Sp}(2) = 3$ (four contractible open sets) was proven by Schweitzer [14] without giving an explicit covering.

Theorem 25 ([5]). *Let us consider the four points: the identity* $I = \operatorname{diag}(1,1)$ *and the matrices* $P = \operatorname{diag}(-1,1)$, $-P$ *and* $-I$. *Then*

$$\{\Omega_G(I), \Omega_G(P), \Omega_G(-I), \Omega_G(-P)\}$$

is an explicit covering of $G = \mathrm{Sp}(2)$ *by contractible open subspaces.*

We shall see later that these four points are the critical points of a Morse function.

Now we consider some symmetric spaces.

THEOREM 26. $\operatorname{cat} \mathrm{U}(2n)/\mathrm{Sp}(n)$ *and* $\operatorname{cat} \mathrm{U}(n)/\mathrm{O}(n)$.

Proof. The original proof is due to Mimura and Sugata [10], and the argument with the Cayley transform is completely analogous to the preceding one. □

For the Grasmannians we have the following result.

THEOREM 27 ([8]). *The Lusternik-Schnirelmann category of the quaternionic Grassmannian*

$$G_{n,k} = \mathrm{Sp}(n)/(\mathrm{Sp}(k) \times \mathrm{Sp}(n-k))$$

is known to be $k(n-k)$. This result can be deduced from Morse theory: the minimum number of critical values of Morse height functions on $G_{n,k}$ is $\operatorname{cat} G_{n,k}+1$.

Finally, for Stiefel manifolds there is only a partial result.

THEOREM 28. *If $n \geq 2k$, the LS-category of the quaternionic Stiefel manifold verifies* $\operatorname{cat} \mathrm{X}_{n,k} \leq k$.

Proof. The original proof is due to Nishimoto [12]. Our proof in [9] is as follows: let $0 < \theta < \pi/2$ and take

$$x_\theta = \begin{pmatrix} 0 \\ (\sin\theta) I_k \\ (\cos\theta) I_k \end{pmatrix}.$$

We know that $\Omega(x_\theta)$ is contractible in $\mathrm{X}_{n,k}$, by Theorem 20. Choose $k+1$ numbers $0 < \theta_0 < \theta_2 < \cdots < \theta_k < \pi/2$. A matrix $\pi \in \mathbb{H}^{k\times k}$ has at most k distinct real eigenvalues, so some of the matrices $\pi + (\cos\theta_i) I_k$ is invertible, that is, $\pi \in \Omega(x_{\theta_i})$. So the family $\{\Omega(x_{\theta_i})\}$ is an open covering of $\mathrm{X}_{n,k}$ by subsets which are contractible in $\mathrm{X}_{n,k}$. □

5.2. Morse-Bott functions

Height functions on a Lie group with respect to some hyperplane as well as distance functions to a given point have been widely studied. In this section, we integrate explicitly the gradient flow of those functions and we give local charts for the critical submanifolds, by means of the Cayley transform.

Let $G = \mathrm{O}(n, \mathbb{K})$ be an orthogonal group embedded in the Euclidean space $E = \mathcal{M}(n, \mathbb{K})$. The Euclidean metric is given by

$$\langle A, B \rangle = \Re \operatorname{Tr}(A^* B),$$

where $\Re \operatorname{Tr}$ means the real part of the trace. Hence, height and distance functions are given, up to a constant, by the formula

$$h_X(A) = \Re \operatorname{Tr}(XA),$$

for some matrix $X \in E$.

A direct computation shows that the gradient of h_X at the point $A \in G$ is

$$(\operatorname{grad} h_X)_A = (1/2)(X^* - AXA).$$

Moreover, if $A \in G$ is a critical point, the Hessian operator is the map

$$(Hh_x)_A \colon T_A G \to T_A G$$

given by

$$(Hh_X)_A(U) = -(1/2)(AXU + UXA), \quad U \in T_A G.$$

Example 29. Let $G = \mathrm{Sp}(n)$ be the symplectic group of quaternionic $n \times n$ matrices A such that $AA^* = I$. For

$$D = \operatorname{diag}(t_1, \ldots, t_n),$$

a positive real diagonal matrix, with $0 < t_1 < \cdots < t_n$, the function h_D is a Morse function, whose critical points are the diagonal matrices

$$\operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_n), \quad \epsilon_k = \pm 1.$$

On the other hand, when $X = I$ the height map h_X is a Morse-Bott function, whose critical manifold is formed by the matrices A such that $A^2 = I$.

We now show how the Cayley transform serves to give a local chart for the set Σ of critical points.

THEOREM 30 ([5]). *Let $h_X(A) = \Re \operatorname{Tr}(XA)$ be an arbitrary height function on the Lie group $G = \mathrm{O}(n, \mathbb{K})$. If $A \in \Sigma$ is a critical point we denote by $S(A)$ the real vector space*

$$S(A) = \{\beta_0 \in T_{A^*} G \colon\ XA\beta_0 + \beta_0 AX = 0\}.$$

Then the Cayley map $c_{A^*} \colon S(A) \to \Sigma \cap \Omega_G(A)$ is a diffeomorphism.

Example 31. Suppose $X = I$ and $\mathbb{K} = \mathbb{C}$. Then the critical points of $h_I \colon \mathrm{U}(n) \to \mathbb{R}$ are the matrices $A \in \mathrm{U}(n)$ such that $A^2 = I$. Such a matrix $A = A^*$ can be diagonalized to $D = \operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_n)$, with $\varepsilon_k = \pm 1$.

On the other hand, $\beta_0 \in T_A G$ if and only if $A\beta_0$ is skew-symmetric, *i.e.*, $A\beta_0 = -\beta_0^* A$, while $\beta_0 \in S(A)$ if and only if $A\beta_0 + \beta_0 A = 0$. It follows that $\beta_0 = \beta_0^*$ (critical direction).

So, for instance, the identity I and its opposite $-I$ are critical points that are isolated because $S(\pm I) = 0$.

On the other hand, let $A = \operatorname{diag}(I_p, -I_q)$. Then $\beta_0 \in T_A G$ must be of the form

$$\beta_0 = \begin{pmatrix} 0 & V^* \\ V & 0 \end{pmatrix},$$

which implies $\dim S(A) = 2pq$. This is in fact the dimension of the (critical) orbit of A under the conjugation action, which is diffeomorphic to the Grasmannian $\mathrm{U}(p+q)/(\mathrm{U}(p) \times \mathrm{U}(q))$.

We shall call *linearization* the process of transforming the gradient flow of h_X in G to a flow in the Lie algebra.

PROPOSITION 32. *Let h_X be an arbitrary height function on $G = \mathrm{O}(n, \mathbb{K})$ and let A be a critical point. The solution of the gradient equation*

$$\alpha' = \frac{1}{2}(X^* - \alpha X \alpha)$$

passing through $\alpha(0) \in \Omega_G(A)$ is the image by the generalized Cayley transform c_{A^*} of the curve in $T_{A^*}G$ defined as

$$\beta(t) = \exp(-XAt/2) \cdot \beta_0 \cdot \exp(-AXt/2), \tag{1}$$

with $\beta_0 = c_A(\alpha(0))$.

This is a generalization of the same result for the classical Cayley transform c_I by Volchenko and Kozachko [16].

Now we consider an analogous result for symmetric spaces. We shall prove that the generalized Cayley transform allows to give explicit local charts for the critical submanifolds of height functions defined on the Cartan model of $M = G/K$ (see Section 3).

Let σ be the automorphism defining K, and let $\widehat{X} := X^* + \sigma(X)$.

PROPOSITION 33. *The gradient of the height function $h_X^M : M \to \mathbb{R}$ at any point $A \in M$ is the projection of $\operatorname{grad} h_X$ onto $T_A M$, that is,*

$$(\operatorname{grad} h_X^M)_A = \frac{1}{4}\left(\widehat{X} - A\sigma(\widehat{X})A\right).$$

The Hessian $H(h_X^M)_A : T_A M \to T_A M$ is given by

$$H(h_X^M)_A(W) = -\frac{1}{4}\left(A\sigma(\widehat{X})W + W\sigma(\widehat{X})A\right).$$

Example 34. Consider the complex Grassmannian $\mathrm{U}(2)/(\mathrm{U}(1)\times\mathrm{U}(1))$ defined by the automorphism $\sigma(A) = I_1 A I_1$, where $I_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. The Cartan model is the sphere $S^2 \subset U(2) \cong S^3 \times S^1$ formed by the matrices $\begin{pmatrix} s & -\overline{z} \\ z & s \end{pmatrix}$ where $(s, z) \in \mathbb{R} \times \mathbb{C}$ verifies $s^2 + |z|^2 = 1$.

We take on M the function h_X^M with $X = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Then $\widehat{X} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$ and the critical points are the two poles $\pm I$. Both tangent spaces $T_{\varepsilon I}M$, are formed by the matrices $W = \begin{pmatrix} 0 & z \\ -\overline{z} & 0 \end{pmatrix}$ with $z \in \mathbb{C}$. The Hessian is

$$(Hh_X^M)_{\varepsilon I}(W) = (-\varepsilon/2)W,$$

so h_X^M is a Morse function on M.

On the other hand, the critical set of h_X^G on the Lie group $\mathrm{U}(2)$ is formed by the two circles $\mathrm{U}(1) \times \{\pm 1\}$ of matrices $\begin{pmatrix} \alpha & 0 \\ 0 & \pm 1 \end{pmatrix}$ with $\alpha \in \mathbb{C}$, $|\alpha| = 1$.

We now integrate explicitly the gradient flow.

Theorem 35 ([7]). *Let h_X^M be an arbitrary height function on the symmetric space M. Let A be a critical point. Then the solution of the gradient equation*

$$4\alpha' = \widehat{X} - \alpha\sigma(\widehat{X})\alpha,$$

with initial condition $\alpha_0 \in \Omega_M(A)$, is the image by the Cayley transform c_{A^*} of the curve

$$\beta(t) = \exp(\frac{-t}{4} A^* \widehat{X})\beta_0 \exp(\frac{-t}{4} \widehat{X} A^*),$$

where $\widehat{X} = X^* + \sigma(X)$ and $\beta_0 = c_A(\alpha_0) \in T_{A^*}M$.

Giving a local chart for the critical set $\Sigma(h_X^M)$ of a Morse-Bott function is another application of the generalized Cayley transform.

Theorem 36. *Let $h_X^M(A) = \Re\,\mathrm{Tr}(XA)$ be a height function on the symmetric space M. Given a critical point $A \in \Sigma(h_X^M)$, let $S^M(A)$ be the vector space*

$$S^M(A) = \{\beta_0 \in T_{A^*}M \colon A^*\widehat{X}\beta_0 + \beta_0\widehat{X}A^* = 0\}.$$

Then, the generalized Cayley transform induces a diffeomorphism

$$c_{A^*} \colon S^M(A) \to \Sigma(h_X^M) \cap \Omega_M(A).$$

Observe that $S^M(A)$ is isomorphic to the kernel of the Hessian $H(h_X^M)_A$.

Example 37. Let $G = \mathrm{Sp}(2)$ and $\sigma(A) = -\mathbf{i}A\mathbf{i}$. Then the fixed point subgroup is $K = \mathrm{U}(2)$. The symmetric space $G/K = \mathrm{Sp}(2)/\mathrm{U}(2)$ can be identified with the manifold of matrices $A \in \mathrm{Sp}(2)$ such that $A^2 = -I$. Let $A_0 = \begin{pmatrix} \mathbf{i} & 0 \\ 0 & \mathbf{i} \end{pmatrix}$. Then the Cartan embedding $\gamma\colon G/K \to \mathrm{Sp}(2)$ is given by $\gamma(A) = -AA_0$. As a consequence, it can be proven that the Cartan model M equals the manifold N of matrices such that $\sigma(A) = A^*$.

Let us take $X = \begin{pmatrix} 1+\mathbf{j} & 0 \\ 0 & \mathbf{i}+\mathbf{j} \end{pmatrix} \in \mathbb{H}^{2\times 2}$.

We can compute the critical points of the restriction h_X^M of the height function to $M \subset G$ and prove that it is a Morse function with four critical points (for a detailed computation see [7]).

5.3. **Optimization**

Unlike most Riemannian manifolds, the gradient flows of height functions can be integrated explicitly on the classical Lie groups and symmetric spaces. As we have seen, this can be achieved by means of the generalized Cayley transform.

On the other hand, the classical Cayley transform has been widely used to linearize differential equations in Lie groups and to solve them numerically. Let us sketch how this works, adapting the explanation given in [6].

Let $f\colon G \to \mathbb{R}$ be a smooth function on the orthogonal Lie group G, and consider the gradient $\mathrm{grad}_A \in T_AG$ at the point $A \in G$. Then the gradient flow $\alpha(t)$ is the solution of the equation $\alpha'(t) = \mathrm{grad}_{\alpha(t)}$, with some initial condition $\alpha(0) \in G$. Since $\mathrm{grad}_{\alpha(t)}$ belongs to $T_{\alpha(t)}G$, which is the image of the Lie algebra $\mathfrak{g} = T_IG$ by the right translation $\alpha(t)$, we can write the equation as

$$\alpha'(t) = X(\alpha(t))\,\alpha(t),$$

where $X(\alpha(t)) \in \mathfrak{g}$ for all t.

In fact, any differential equation on the Lie group G can be written as $\alpha' = X(t,\alpha)\alpha$. To fix ideas, assume that we have a linear equation.

PROPOSITION 38 ([6]). *The differential equation*

$$\alpha'(t) = X(t)\,\alpha(t) \tag{2}$$

on the Lie group G, where $X(t)$ is a $\mathfrak{g}$-valued smooth function, can be reduced to the equation

$$\Omega' = -\frac{1}{2}(I+\Omega)X(I-\Omega)$$

on the Lie algebra $\mathfrak{g}$, where $\Omega(t)$ is a $\mathfrak{g}$-valued function.

Proof. The curve $\alpha(t)\alpha(0)^{-1} \in G$ passing through the identity can be written, for small values of t, as the image, by the classical Cayley transform c_I, of some curve $\Omega(t)$ on the Lie algebra $\mathfrak{g}$, with $\Omega(0) = 0$. That is,

$$\alpha(t) = c_I(\Omega(t))\alpha(0). \tag{3}$$

From the formula

$$c_I(\Omega) = (I - \Omega)(I + \Omega)^{-1} \in G$$

it follows that the derivative

$$\alpha'(t)\alpha(0)^{-1} = (c_I\Omega)'(t)$$

equals

$$\begin{aligned} & -\Omega'(I+\Omega)^{-1} - (I-\Omega)(I+\Omega)^{-1}\Omega'(I+\Omega)^{-1} \\ =& \left(-I - (I-\Omega)(I+\Omega)^{-1}\right)\Omega'(I+\Omega)^{-1} \\ =& -2(I+\Omega)^{-1}\Omega'(I+\Omega)^{-1}, \end{aligned}$$

which combined with formulas (2) and (3) gives

$$-2\Omega(I+\Omega)^{-1}\Omega'(I+\Omega)^{-1} = \alpha'\alpha(0)^{-1} = X(I-\Omega)(I+\Omega)^{-1},$$

so

$$-2\Omega' = (I+\Omega)X(I-\Omega), \tag{4}$$

because Ω and $(I+\Omega)^{-1}$ commute. □

Equation (4) can be written as

$$-2\Omega' = X + [\Omega, X] - \Omega X \Omega.$$

Notice that $X, \Omega \in \mathfrak{g}$ means that they are skew-symmetric matrices, so they are $[\Omega, X] = \Omega X - X\Omega$ and $\Omega X \Omega$.

Remark 39. The equation can be solved by Runge-Kutta methods on Lie groups [11].

Now we shall see how the Cayley transform can be used in the reverse way. As an example, consider the curve given in (1), which is the solution of the linear equation

$$\beta' = (-1/2)\left(XA\beta + \beta AX\right) \tag{5}$$

in the Lie algebra $\mathfrak{g}$ of quaternionic skew-symmetric matrices. Recall that the matrix $X \in \mathcal{M}(n, \mathbb{K})$ defined the height function h_X, while $A \in G$ was a critical point of that function.

By a computation similar to that, done in the proof of Proposition 38 referring to Equation (2), we have the following result.

PROPOSITION 40. *Let $\alpha' = \alpha X\alpha - AXA$ be a differential equation in the orthogonal group G, where $A \in G$ and $X(t) \in \mathcal{M}(n, \mathbb{K})$ is a matrix valued function. Assume that the matrix $AX(t)$ is symmetric for all t. Then the Cayley map c_A transforms the original equation into the linear equation $\beta' = XA\beta + \beta AX$ in the Lie algebra $\mathfrak{g}$.*

Gradient flows related to the minimization of a cost function $f\colon G \to \mathbb{R}$ have been widely studied too. One popular method is the following *gradient descent method* [1]. Let $A = A_0 \in G$ be an initial trial point and let $-\operatorname{grad}_A$ be the negative gradient of f at A. Then a curve $\alpha(t)$ must be found on the manifold G such that $\alpha(0) = A$ and $\alpha'(0) = -\operatorname{grad}_A$. By fixing a step size τ small enough, the next iterate is obtained by curvilinear search, that is, by putting $A_1 = \alpha(\tau)$. Under certain conditions the sequence $A_0, A_1, \dots$ will converge to a local minimun of the function f.

Most existing gradient descent methods require the determination of geodesic curves, which is computationally expensive. Recall that the Riemannian metric induced by the Euclidean one on the compact orthogonal group $G = \mathrm{O}(n, \mathbb{K})$ is bi-invariant, and that the geodesic curve through a point $A \in G$ in the direction $X \in T_A G$ is given by $\exp(tX)A$, where exp is the usual matrix exponential.

Even more, many problems have some *orthogonality constraints.* A typical example is looking for k orthogonal n-vectors that are optimal with respect to some function f like cost or likelihood. This kind of problems are widely known in optimization theory, and they can be seen as optimization problems on a real Stiefel manifold. However, preserving the orthogonality constraints is numerically expensive. Usually, re-orthogonalization requires to use matrix factorizations such as polar decomposition or singular value decomposition to find the nearest orthogonal matrix to the matrix obtained in each step.

For Stiefel manifolds, a different algorithm has been proposed by Z. Wen and W. Yin [17], where the search curve is not a geodesic but it is constructed from the Cayley transform on the orthogonal group. Let us examine their idea more closely.

Remark 41. The method is intended to be used in the real case, but we shall write it for arbitrary coefficients $\mathbb{K}$. Also, as noted in [17], it could be extended to J-orhogonal groups.

Let $G = \mathrm{O}(n, \mathbb{K})$ be an orthogonal group and let $S = \mathrm{O}_{n,k}(\mathbb{K}) \subset \mathbb{R}^{n\times k}$ be a Stiefel manifold. Let $f\colon \mathbb{K}^{n\times k} \to \mathbb{R}$ be a real function and let $f_S\colon S \to \mathbb{R}$ be its restriction to the Stiefel manifold. We denote by

$$D = (\operatorname{grad} f)_x \in \mathbb{K}^{n\times k}$$

the gradient of f at the point $x \in S$.

We recover the notations of Section 4. The Euclidean norm on the space $\mathbb{K}^{n\times k}$ of $n \times k$ matrices is

$$|M|^2 = \Re\operatorname{Tr}(M^*M) = \sum_{i,j} |m_{i,j}|^2,$$

so the gradient of f is defined, for the usual Euclidean inner product, by the condition

$$f_{*x}(M) = \langle D, M\rangle = \Re\operatorname{Tr}(D^*M).$$

Analogously, let

$$D_S = (\operatorname{grad} f_S)_x \in T_x\mathrm{O}_{n,k}$$

be the gradient of the restricted map $f_S\colon S \to \mathbb{R}$. Clearly, D_S is the projection of D onto the tangent space T_xS. Let us denote this as

$$D_S = \operatorname{proj}_x(D).$$

Proposition 42.

$$D_S = (I - \frac{1}{2}xx^*)D - \frac{1}{2}xD^*x.$$

Proof. We know that $x = Ax_0$, with $A \in G$ and $x_0 = \begin{pmatrix} 0 \\ I_k \end{pmatrix}$. Then $T_xS = A \cdot T_{x_0}S$ for the transitive action of $G = \mathrm{O}(n, \mathbb{K})$ onto $S = \mathrm{O}_{n,k}(\mathbb{K})$. This action is isometric, because

$$\langle Av_0, Aw_0\rangle = \Re\operatorname{Tr}((Av_0)^*(Aw_0)) = \Re\operatorname{Tr}(v_0^*w_0) = \langle v_0, w_0\rangle.$$

Then

$$D_S = A \cdot \operatorname{proj}_{x_0}(A^*D),$$

where $A^*D \in T_{x_0}S$.

Next step is to realize that $\mathbb{K}^{n\times k}$ can be decomposed as the orthogonal sum of two subspaces: one is $T_{x_0}S$, where $v_0 \in \mathbb{K}^{n\times k}$ belongs to $T_{x_0}S$ if and only if $v_0^*x_0 + x_0^*v_0 = 0$, that is $v_0 = \begin{pmatrix} X \\ Y \end{pmatrix}$ with $Y + Y^*$. The other subspace is the set of matrices $\begin{pmatrix} 0 \\ Z \end{pmatrix}$ where $Z = Z^*$ is a symmetric $k \times k$ submatrix.

This allows, by using the condition $AA^* = I$, to explicitly compute D_S. □

Remark 43. With the notations of [17], the matrix D_S in Proposition 42 should be written as

$$D_S = (I - xx^*)D + \frac{1}{2}x(x^*D - D^*x).$$

Moreover,

$$D_S^* x = \frac{1}{2}(x^* D - D^* x),$$

which proves that $D_S \in T_x S$ because $D_S^* x + x^* D_S = 0$.

We are ready to find the search curve. The $n \times n$ matrix

$$W = (I - \frac{1}{2} x x^*) D x^* - x D^* (I - \frac{1}{2} x x^*)$$

is skew-symmetric, *i.e.*, W belongs to $\mathfrak{g}$, the Lie algebra of $G = \mathrm{O}(n, \mathbb{K})$. Take the curve $c_I(tW) \in G$, $t \in \mathbb{R}$ on the group, given by the Cayley transform, and make it act on the Stiefel manifold. We obtain in this way a curve

$$\alpha(t) = c_I(tW) x \in S$$

with the required properties.

Proposition 44 ([17]). *The curve $\alpha(t)$ verifies $\alpha(0) = x$ and $\alpha'(0) = -D_S$, the negative gradient of the restriction f_S at the point x.*

Proof. From the formula $c_I(tW) = (I - tW)(I + tW)^{-1}$ it follows that

$$\alpha'(t) = -W(I + tW)^{-1} x - (I - tW)(I + tW)^{-1} W (I + tW)^{-1} x,$$

hence

$$\alpha'(0) = -2Wx = -(I - \frac{1}{2} x x^*) D + \frac{1}{2} x D^* x = -D_S. \quad \square$$

Wen-Yin method has been used in many recent papers, going from biology to computer science. We expect that our formulas for the Cayley transform on the Stiefel manifolds may also be of interest for applications in optimization and control theory, leading to even more efficient methods.

Acknowledgments. This research was supported by Labex CEMPI (ANR-11-LABX-0007-01) and MINECO-FEDER research project MTM2016-78647-P. The author thanks the organizers of the "13th International Workshop on Differential Geometry and its Applications" and Universitatea Petrol-Gaze din Ploieşti for their warm hospitality.

REFERENCES

[1] P.A. Absil, R. Mahony and R. Sepulchre, *Optimization algorithms on matrix manifolds.* Princeton Univ. Press, 2008.

[2] A. Cayley, *Sur quelques propriétés des déterminants gauches.* J. Reine Angew. Math. **32** (1846), 119–123.

[3] O. Cornea, G. Lupton, J. Oprea and D. Tanré, *Lusternik-Schnirelmann category.* Math. Surveys and Monogr. **103**, Amer. Math. Soc., 2003.

[4] M. Farber, *Topological complexity of motion planning.* Discrete Comput. Geom. **29** (2003), 211–221.

[5] A. Gómez-Tato, E. Macías-Virgós and M.J. Pereira-Sáez, *Trace map, Cayley transform and LS category of Lie groups.* Ann. Global Anal. Geom. **39** (2011), *3*, 325–335.

[6] A. Iserles, *On Cayley-transform methods for the discretization of Lie-group equations.* Found. Comput. Math. **1** (2001), 129–160.

[7] E. Macías-Virgós and M.J. Pereira-Sáez, *Height functions on compact symmetric spaces.* Monatsh. Math. **177** (2015), *1*, 119–140.

[8] E. Macías-Virgós, M.J. Pereira-Sáez and D. Tanré, *Morse theory and the Lusternik-Schnirelmann category of quaternionic Grassmannians.* Proc. Edinb. Math. Soc. (2) **60** (2017), *2*, 441–449.

[9] E. Macías-Virgós, M.J. Pereira-Sáez and D. Tanré, *Cayley Transform on Stiefel manifolds.* J. Geom. Phys. (2017). To appear.

[10] M. Mimura and K. Sugata, *On the Lusternik-Schnirelmann category of symmetric spaces of classical type.* In: N. Iwase et al. (Eds.), *Proceedings of the conference on groups, homotopy and configuration spaces*, University of Tokyo, Japan, July 5–11, 2005 in honor of the 60$^{\text{th}}$ birthday of Fred Cohen. Coventry: Geometry & Topology Publications. Geometry and Topology Monographs **13**, 323–334, 2008.

[11] H. Munthe-Kaas, *Runge-kutta methods on Lie groups.* BIT **38** (1998), *1*, 92–111.

[12] T. Nishimoto, *On the Lusternik-Schnirelmann category of Stiefel manifolds.* Topology Appl. **154** (2007), *9*, 1956–1960.

[13] M. Postnikov, *Lie groups and Lie algebras.* Lectures in Geometry **5**, MIR Moscow, 1994.

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

[15] W. Singhof, *On the Lusternik-Schnirelmann category of Lie groups I.* Math. Z. **145** (1975), 11–116; *On the Lusternik-Schnirelmann category of Lie groups II.* Math. Z. **151** (1976), 143–148.

[16] K.Y. Volchenko and A.N. Kozachko, *Integrable gradient flows on classic manifolds and Morse theory.* Moscow Univ. Math. Bull. **52** (1997), *3*, 9–15; translation from Vestnik Moskov. Univ. Ser. I Mat. **3** (1997), 9–15.

[17] Z. Wen and W. Yin, *A feasible method for optimization with orthogonality constraints.* Math. Program. Ser. A **142** (2013), *1-2*, 397–434.

Received 13 October 2017

University of Santiago de Compostela,
Department of Mathematics,
15782 Spain
quique.macias@usc.es