

# self-study: geometry on orthogonal group, Stiefel manifold and Grassmann manifold

Tianpei Xie

Jun. 15th., 2015

## Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Concepts</b>   | <b>2</b>  |
| <b>2</b> | <b>Geometry of orthogonal group <math>\mathcal{O}(n)</math></b>                               | <b>7</b>  |
| 2.1      | The tangent space and normal space of $\mathcal{O}(n)$ and the local orthogonal decomposition | 7         |
| 2.2      | The geodesic on $\mathcal{O}(n)$ and parallel transport                                       | 8         |
| 2.3      | The geodesic for Stiefel manifold   | 9         |
| <b>3</b> | <b>Geometry of Stiefel manifold <math>V_k(\mathbb{R}^n)</math></b>                            | <b>12</b> |
| 3.1      | Geometry of quotient manifold $X/G$   | 12        |
| 3.2      | The tangent space of $V_k(\mathbb{R}^n)$  | 12        |
| 3.3      | canonical metric on $V_k(\mathbb{R}^n)$   | 14        |
| 3.4      | Geodesic on $V_k(\mathbb{R}^n)$   | 15        |
| 3.5      | parallel translation  | 19        |
| 3.6      | Gradient function on Stiefel manifold   | 19        |
| 3.7      | Hessian of function   | 20        |
| <b>4</b> | <b>Geometry of Grassmann manifold</b>   | <b>21</b> |
| 4.1      | The tangent space of $Gr(k, n)$   | 21        |
| 4.2      | canonical metric on $Gr(k, n)$  | 22        |
| 4.3      | Geodesic on $Gr(k, n)$  | 23        |
| 4.4      | parallel translation  | 26        |
| 4.5      | Gradient function on Grassmann manifold   | 26        |
| 4.6      | Hessian of function   | 26        |

# 1 Concepts

- The space of all  $n \times n$  real matrices is denoted as  $M_n(\mathbb{R})$ .
- The *general linear group* ( $GLN$ ) of degree  $n$  is the set of  $n \times n$  invertible matrices, together with the operation of ordinary matrix multiplication. The general linear group over  $\mathbb{R}$  (the set of real numbers) is the group of  $n \times n$  invertible matrices of real numbers, and is denoted by  $GLn(\mathbb{R})$  or  $GL(n, \mathbb{R})$ , i.e.

$$GL(n, \mathbb{R}) \equiv \{ \mathbf{A} \in \mathbb{R}^{n \times n} : \det(\mathbf{A}) \neq 0 \}.$$

- The *special linear group*, written  $SL(n, \mathbb{R})$  or  $SLn(\mathbb{R})$ , is the subgroup of  $GL(n, \mathbb{R})$  consisting of matrices with a determinant of 1.

$$SL(n, \mathbb{R}) \equiv \{ \mathbf{A} \in \mathbb{R}^{n \times n} : \det(\mathbf{A}) = 1 \}.$$

Note that  $\dim SL(n, \mathbb{R}) = \dim GL(n, \mathbb{R}) = n^2$ .

- The *orthogonal group of dimension  $n$* , denoted  $\mathcal{O}(n)$ , is the group of distance-preserving transformations of a Euclidean space of dimension  $n$  that preserve a fixed point, where the group operation is given by composing transformations.

Also,  $(\mathcal{O}(n), \cdot)$  is the group of  $n \times n$  orthogonal matrices, where the group operation  $(\cdot)$  is given by matrix multiplication, and an orthogonal matrix is a real matrix whose inverse equals its transpose.

$$\mathcal{O}(n) \equiv \{ \mathbf{Q} \in GL(n, \mathbb{R}) : \mathbf{Q}^T \mathbf{Q} = \mathbf{Q} \mathbf{Q}^T = \mathbf{I}_n \}.$$

- An important subgroup of  $\mathcal{O}(n)$  is the *special orthogonal group*, denoted  $\mathcal{SO}(n)$ , of the orthogonal matrices of determinant 1. This group is also called the *rotation group*

$$\mathcal{SO}(n) \equiv \{ \mathbf{Q} \in \mathcal{O}(n) : \det(\mathbf{Q}) = 1 \}.$$

Note that  $\dim \mathcal{SO}(n) = \dim \mathcal{O}(n) = n(n-1)/2$ .

- A *topological group*  $(G, \cdot)$  is a *topological space* and *group* satisfying the  $T_1$  axiom (i.e., all finite subset are closed), such that

1. the group operations of product:

$$G \times G \rightarrow G : (x, y) \mapsto x \cdot y$$

2. and taking inverses:

$$G \rightarrow G : x \mapsto x^{-1}$$

are *continuous* functions. Here,  $G \times G$  is viewed as a topological space by using the product topology.

- Define map  $h_a : G \rightarrow G$  as  $h_a(x) = a \cdot x$  and  $g_a : G \rightarrow G$  as  $g_a(x) = x \cdot a$ , for  $a \in G$ . It can be shown that  $h_a, g_a$  are *homomorphism* and  $G$  is a *homogeneous* space, i.e. for every pair  $x, y \in G$ , there exists a homomorphism  $f : G \rightarrow G$  such that  $f(x) = y$ .

- For  $H \subset G$  as the subgroup of  $G$ , define the *left coset* as  $xH = \{x \cdot h \mid h \in H\}$  and the collection of left cosets defines a *quotient group*  $G/H = \{xH \mid x \in G\}$  with the group operation  $xH \cdot yH = (x \cdot y)H$ .  $G/H$  is a homogeneous space.
- $\mathcal{O}(n)$  is a compact topological group.
- Let  $G$  be a topological group and  $X$  is a topological space. An *action*  $\alpha$  of  $G$  on  $X$  is a continuous map  $\alpha : G \times X \rightarrow X$  such that denoting  $\alpha(g \times x) = g \cdot x$ , then
  1. the identity map:

$$e \cdot x = x, \quad \forall x \in X$$

2. and product of maps

$$g_2 \cdot (g_1 \cdot x) = (g_2 \cdot g_1) \cdot x, \quad \forall x \in X, \forall g_1, g_2 \in G$$

Define the *equivalence relationship*  $\sim$  as  $(x \sim g \cdot x)$  for all  $g \in G$ , then the coset (equivalence class)  $Gx$  is called *the orbit of a point  $x$  in  $X$*

$$Gx \equiv \{g \cdot x \mid g \in G\}$$

The *quotient space*  $X/G$  is the set of all orbits of  $X$  under the action of  $G$ , or, *the orbit space of  $X$  under action of  $G$* .

- The orbit space is a partition of space  $X$ ;
- The group action is *transitive* if and only if it has only one orbit, i.e. if there exists  $x \in X$  with  $G \cdot x = X$ . This is the case if and only if  $G \cdot x = X$  for all  $x$  in  $X$ .
- Note that the orbit space of space  $X$  under orthogonal group action  $\mathcal{O}(k)$  is given as  $X/\mathcal{O}(k)$ .
- $(\mathcal{O}(n)/\mathcal{O}(p)) = \{g\mathcal{O}(p) \mid g \in \mathcal{O}(n)\}$  is a quotient group and also a orbit space of  $\mathcal{O}(n)$  under action of  $\mathcal{O}(p)$ . Note that the equivalence class is the (right) orbit at  $g$  as  $g\mathcal{O}(p)$ .
- The (compact) *Stiefel manifold*  $V_k(\mathbb{R}^n)$  is the set of all orthonormal  $k$ -frames in  $\mathbb{R}^n$ . That is, it is the set of ordered  $k$ -tuples of orthonormal vectors in  $\mathbb{R}^n$ .

$$V_k(\mathbb{R}^n) \equiv \left\{ \mathbf{U} \in \mathbb{R}^{n \times k} : \mathbf{U}^T \mathbf{U} = \mathbf{I}_k \right\}, \quad k \leq n.$$

Note that  $V_n(\mathbb{R}^n) = \mathcal{O}(n)$  and  $V_1(\mathbb{R}^n) = \mathbb{S}^{n-1}$ .  $\dim(V_k(\mathbb{R}^n)) = kn - \frac{k(k+1)}{2}$ .

The topology of  $V_k(\mathbb{R}^n)$  is the subspace topology inherited from  $\mathbb{R}^{n \times k}$ . It is a *homogeneous quotient space* as

$$\begin{aligned} V_k(\mathbb{R}^n) &\simeq \mathcal{O}(n)/\mathcal{O}(n-k); \\ V_k(\mathbb{R}^n) &\simeq \mathcal{SO}(n)/\mathcal{SO}(n-k), \quad k < n, \end{aligned}$$

which is a homogeneous space. It means that every orthogonal transformation of a  $k$ -frame in  $\mathbb{R}^n$  results in another  $k$ -frame, and any two  $k$ -frames are related by some orthogonal transformation. In other words, the orthogonal group  $\mathcal{O}(n)$  acts transitively on  $V_k(\mathbb{R}^n)$ . The stabilizer subgroup of a given frame is the subgroup isomorphic to  $\mathcal{O}(n-k)$  which acts nontrivially on the orthogonal complement of the space spanned by that frame.

The point in  $V_k(\mathbb{R}^n) \simeq \mathcal{O}(n)/\mathcal{O}(n-k)$  is the equivalence class

$$[Q] = \left\{ Q \begin{bmatrix} \mathbf{I}_k & 0 \\ 0 & \mathbf{Q}_{n-k} \end{bmatrix} \mid \mathbf{Q}_{n-k} \in \mathcal{O}(n-k) \right\}.$$

- The *Grassmannian* (or, *Grassmann manifold*)  $Gr(k, \mathbb{R}^n)$  or  $Gr(k, n)$  is a space which parameterizes all *linear subspaces* of a vector space  $\mathbb{R}^n$  of given dimension  $k$ . For example, the Grassmann manifold  $Gr(1, \mathbb{R}^n)$  is the space of lines through the origin in  $\mathbb{R}^n$ , so it is the same as the *projective* space of one dimension lower than  $\mathbb{R}^n$ .

$Gr(k, \mathbb{R}^n)$  is a compact smooth manifold. Equipped with topological and differential structure, via  $Gr(k, n)$ , a continuous and smooth choice of subspace or open and closed collections of subspaces.

The Grassmann manifold has a natural motivation: Consider the tangent bundle  $T\mathcal{S}$  of a smooth manifold  $\mathcal{S}^k \subset \mathbb{R}^n$  of dimension  $k$ , which assign for each point  $p \in \mathcal{S}$ , the tangent space  $T_p\mathcal{S} \simeq \mathbb{R}^k$ . It defines a map  $F : \mathcal{S}^k \rightarrow Gr(k, n)$ . The property of tangent bundle is related to the properties of the corresponding map  $F$  viewed as continuous map.

Note that the general linear group  $GL(n, \mathbb{R})$  acts transitively on the  $k$ -dimensional subspaces of  $\mathbb{R}^n$  and it holds also for the orthogonal group  $\mathcal{O}(n)$ . By restricting to orthonormal frames, one gets the identity

$$\begin{aligned} Gr(k, n) &\simeq \mathcal{O}(n) / (\mathcal{O}(k) \times \mathcal{O}(n-k)); \\ &\simeq V_k(\mathbb{R}^n) / \mathcal{O}(k), \end{aligned}$$

and  $\dim(Gr(k, n)) = k(n-k)$

The point in  $Gr(k, n) \simeq \mathcal{O}(n) / (\mathcal{O}(k) \times \mathcal{O}(n-k))$  is the equivalence class

$$[Q] = \left\{ Q \begin{bmatrix} \mathbf{Q}_k & 0 \\ 0 & \mathbf{Q}_{n-k} \end{bmatrix} \mid \mathbf{Q}_k \in \mathcal{O}(k), \mathbf{Q}_{n-k} \in \mathcal{O}(n-k) \right\}.$$

or as  $Gr(k, n) \simeq V_k(\mathbb{R}^n) / \mathcal{O}(k)$ ,

$$[U] = \left\{ U\mathbf{Q}_k \mid \mathbf{Q}_k \in \mathcal{O}(k) \right\}.$$

The point of Grassmann manifold can be equivalently uniquely represented by the projection matrix

$$[U] \simeq UU^T,$$

which use  $n^2$  entries to represent as dimension of  $k(n-k)$ .

- The *matrix exponential* is a matrix function on square matrices analogous to the ordinary exponential function; i.e.

$$\exp(\mathbf{X}) = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{X}^k$$

- It is seen as an exponential map from a *Lie algebra* to a *Lie group*.
- $\exp(\mathbf{0}) = \mathbf{I}$ ;
- $\exp(\mathbf{X}) \exp(\mathbf{Y}) = \exp(\mathbf{X} + \mathbf{Y})$ ,  $\exp(\mathbf{X}) \exp(\mathbf{X}^{-1}) = \mathbf{I}$ .
- $\exp(\mathbf{X}^T) = \exp(\mathbf{X})^T$ . If  $\mathbf{X}$  is skew-symmetric,  $\exp(\mathbf{X})$  is orthogonal; and if  $\mathbf{X}$  is symmetric,  $\exp(\mathbf{X})$  is symmetric.
- If  $\mathbf{Y}$  is invertible, then  $\exp(\mathbf{Y} \mathbf{X} \mathbf{Y}^{-1}) = \mathbf{Y} \exp(\mathbf{X}) \mathbf{Y}^{-1}$
- The solution to the system of differential equations is

$$\dot{\mathbf{x}} = \mathbf{A} \mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0$$

is  $\mathbf{x}(t) = \exp(t \mathbf{A}) \mathbf{x}_0$ .

- If  $\mathbf{A}$  and  $\mathbf{H}$  are *Hermitian* matrices, then

$$\text{tr}(\exp(\mathbf{A} + \mathbf{H})) \leq \text{tr}(\exp(\mathbf{A}) \exp(\mathbf{H})).$$

- For a fixed Hermitian matrix  $\mathbf{H}$ , the function

$$f(\mathbf{A}) = \text{tr}(\exp(\mathbf{H} + \log \mathbf{A})).$$

is *concave* on the cone of positive-definite matrices.

- The exponential map  $\exp : \mathbb{M}_n(\mathbb{C}) \rightarrow GL(n, \mathbb{C})$  is surjective, i.e. every invertible square matrix can be represented as the exponential of some square matrix. However,  $\exp : \mathbb{M}_n(\mathbb{R}) \rightarrow GL(n, \mathbb{R})$  is not surjective.

The exponential map is continuous and *Lipschitz* continuous on compact subsets of  $\mathbb{M}_n(\mathbb{R})$ .

- The map  $t \mapsto \exp(t \mathbf{X})$ ,  $t \in \mathbb{R}$  defines a smooth *curve* in the general linear group which passes through the identity element at  $t = 0$ . The tangent vector is

$$\frac{d}{dt} \exp(t \mathbf{X}) = \exp(t \mathbf{X}) \mathbf{X} = \mathbf{X} \exp(t \mathbf{X})$$

- By Jacobi's formula,

$$\det(\exp(\mathbf{X})) = \exp(\text{tr}(\mathbf{X}))$$

- Any orthogonal matrix  $\mathbf{Q}$  can be represented as

$$\mathbf{Q} = \exp(\mathbf{A}) \tag{1}$$

where  $\mathbf{A}$  is skew-symmetric, i.e.  $\mathbf{A} + \mathbf{A}^T = \mathbf{0}$ .

- For diagonal  $\Sigma_R$ ,

$$\begin{aligned}
\exp\left(t \begin{bmatrix} 0 & -\Sigma_R \\ \Sigma_R & 0 \end{bmatrix}\right) &= \sum_{k=0}^{\infty} \frac{1}{k!} \begin{bmatrix} 0 & -t\Sigma_R \\ t\Sigma_R & 0 \end{bmatrix}^k \\
&= \sum_{k=0}^{\infty} \frac{1}{(2k)!} \begin{bmatrix} 0 & -t\Sigma_R \\ t\Sigma_R & 0 \end{bmatrix}^{2k} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \begin{bmatrix} 0 & -t\Sigma_R \\ t\Sigma_R & 0 \end{bmatrix}^{2k+1} \\
&= \sum_{k=0}^{\infty} \frac{1}{(2k)!} \begin{bmatrix} (-1)^k (t\Sigma_R)^{2k} & 0 \\ 0 & (-1)^k (t\Sigma_R)^{2k} \end{bmatrix} \\
&\quad + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \begin{bmatrix} 0 & (-1)^{k+1} (t\Sigma_R)^{2k+1} \\ (-1)^k (t\Sigma_R)^{2k+1} & 0 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} (t\Sigma_R)^{2k} & 0 \\ 0 & \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} (t\Sigma_R)^{2k} \end{bmatrix} \\
&\quad + \begin{bmatrix} 0 & \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{(2k+1)!} (t\Sigma_R)^{2k+1} \\ \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} (t\Sigma_R)^{2k+1} & 0 \end{bmatrix} \\
&= \begin{bmatrix} \cos t\Sigma_R & -\sin t\Sigma_R \\ \sin t\Sigma_R & \cos t\Sigma_R \end{bmatrix}
\end{aligned}$$

## 2 Geometry of orthogonal group $\mathcal{O}(n)$

### 2.1 The tangent space and normal space of $\mathcal{O}(n)$ and the local orthogonal decomposition

By definition, the matrix  $\mathbf{Q} \in \mathcal{O}(n)$ , if and only if

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_n.$$

Consider the regular curve on  $\mathcal{O}(n)$  as a one-parameter mapping  $\alpha : (-\epsilon, \epsilon) \rightarrow \mathcal{O}(n)$ , as  $t \mapsto \mathbf{Q}(t)$ , and define the tangent vector of the curve as

$$\dot{\mathbf{Q}} = \left. \frac{d}{dt} \mathbf{Q}(t) \right|_{t=0}.$$

By the Leibniz rule,

$$\dot{\mathbf{Q}}^T \mathbf{Q} + \mathbf{Q}^T \dot{\mathbf{Q}} = \mathbf{0}; \quad (2)$$

that is,  $\dot{\mathbf{Q}}^T \mathbf{Q}$  is *skew-symmetric*. Note that  $\dot{\mathbf{Q}} \in T_{\mathbf{Q}}\mathcal{O}(n)$  and it holds for any regular parameterized curve  $\alpha$  on  $\mathcal{O}(n)$ . Thus the tangent space  $T_{\mathbf{Q}}\mathcal{O}(n)$  at  $\mathbf{Q}$  consist of all matrix  $\Delta$  such that  $\mathbf{A} \equiv \Delta^T \mathbf{Q}$  is skew-symmetric.

Define the inner product  $\langle \cdot, \cdot \rangle$  on  $\mathcal{O}(n)$  as induced from  $\mathbb{R}^{n \times n}$ , which is

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^T \mathbf{B}).$$

Then the normal space of the surface  $\mathcal{O}(n)$  at  $\mathbf{Q}$  consists of all matrix  $\mathbf{N}$  such that

$$\text{tr}(\Delta^T \mathbf{N}) = 0.$$

By substituting the fact that  $\Delta^T \mathbf{Q}$  is skew-symmetric, we claim that

$$\mathbf{N} = \mathbf{Q} \mathbf{S}, \quad (3)$$

for some symmetric matrix  $\mathbf{S} \in \mathbb{S}(n)$ . To show this, we check that

$$\begin{aligned} \text{tr}(\Delta^T \mathbf{N}) &= \frac{1}{2} (\text{tr}(\Delta^T \mathbf{Q} \mathbf{S}) + \text{tr}(\mathbf{S} \mathbf{Q}^T \Delta)) \\ &= \frac{1}{2} (\text{tr}(\mathbf{A} \mathbf{S} - \mathbf{S} \mathbf{A})) \\ &= \frac{1}{2} \text{tr}([\mathbf{A}, \mathbf{S}]) = 0, \end{aligned}$$

where  $\mathbf{A} \equiv \Delta^T \mathbf{Q} = -\mathbf{A}^T$  is skew-symmetric and  $[\mathbf{A}, \mathbf{S}] = \mathbf{A} \mathbf{S} - \mathbf{S} \mathbf{A}$  is the *Lie bracket*. The last equality holds as the property of Lie bracket and it follows from the fact that  $\text{tr}(\mathbf{A} \mathbf{S}) = \text{tr}(\mathbf{S} \mathbf{A})$ .

Since the tangent space  $T_{\mathbf{Q}}\mathcal{O}(n)$  has the dimension of all skew-symmetric space, which is  $\frac{n(n-1)}{2}$  and the dimension of the space of all matrices  $\{\mathbf{Q} \mathbf{S}, \mathbf{S} \in \mathbb{S}(n)\}$  is  $\frac{n(n+1)}{2}$ , the latter is the orthogonal complementary space of  $T_{\mathbf{Q}}\mathcal{O}(n)$ . So it is normal space of  $\mathcal{O}(n)$  at  $\mathbf{Q}$ , by definition.

Given the matrix  $\mathbf{R}$ , the projection of  $\mathbf{R}$  onto normal space of  $\mathcal{O}(n)$  at  $\mathbf{Q}$  is given by

$$\min_{\mathbf{S} \in \mathbb{S}(n)} \frac{1}{2} \|\mathbf{R} - \mathbf{U}\mathbf{S}\|_F^2$$

Then

$$\begin{aligned} \frac{\partial}{\partial \mathbf{S}} \left( \frac{1}{2} \text{tr}(\mathbf{S}^T \mathbf{Q}^T \mathbf{U} \mathbf{S}) - \text{tr}(\mathbf{R}^T \mathbf{U} \mathbf{S}) \right) &= \frac{\partial}{\partial \mathbf{S}} \left( \frac{1}{2} \text{tr}(\mathbf{S}^T \mathbf{S}) - \text{tr}(\mathbf{R}^T \mathbf{U} \mathbf{S}) \right) \\ &= (2\mathbf{S} - \text{diag}(\mathbf{S} \odot \mathbf{I})) - (\mathbf{Q}^T \mathbf{R} + \mathbf{R}^T \mathbf{Q} - \text{diag}(\mathbf{R}^T \mathbf{Q} \odot \mathbf{I})) \\ &= 0 \\ \Rightarrow \mathbf{S} &= \frac{1}{2} (\mathbf{Q}^T \mathbf{R} + \mathbf{R}^T \mathbf{Q}). \end{aligned}$$

So the projection of  $\mathbf{R}$  onto normal space of  $\mathcal{O}(n)$  at  $\mathbf{Q}$  is

$$\begin{aligned} \pi_N(\mathbf{R}) &= \mathbf{Q} \left( \frac{1}{2} (\mathbf{Q}^T \mathbf{R} + \mathbf{R}^T \mathbf{Q}) \right) \\ &\equiv \mathbf{Q}_{\text{sym}}(\mathbf{Q}^T \mathbf{R}) \end{aligned} \tag{4}$$

.

Thus, by orthogonal complementary property, this gives the projection of  $\mathbf{R}$  onto tangent space of  $\mathcal{O}(n)$  at  $\mathbf{Q}$  as

$$\pi_T(\mathbf{R}) = \mathbf{R} - \mathbf{Q}_{\text{sym}}(\mathbf{R}^T \mathbf{Q}) \tag{5}$$

$$\begin{aligned} &= \mathbf{R} - \mathbf{Q}_{\text{sym}}(\mathbf{Q}^T \mathbf{R}) \\ &= \mathbf{R} - \mathbf{Q}(\mathbf{Q}^T \mathbf{R} - \text{skew}(\mathbf{Q}^T \mathbf{R})) \\ &= \mathbf{Q}_{\text{skew}}(\mathbf{Q}^T \mathbf{R}) \end{aligned} \tag{6}$$

where  $\text{skew}(\mathbf{A}) = \frac{1}{2}(\mathbf{A} - \mathbf{A}^T)$ .

It suggests from above that  $\Delta$  in the tangent space of  $\mathcal{O}(n)$  at  $\mathbf{Q}$  can be decomposed as

$$\Delta = \mathbf{Q}\mathbf{A} \tag{7}$$

where  $\mathbf{A}$  is skew-symmetric.

## 2.2 The geodesic on $\mathcal{O}(n)$ and parallel transport

Assume that the vector field that assigns  $\Delta \in T_{\mathbf{Q}}\mathcal{O}(n)$  at  $\mathbf{Q}$  is *parallel* along the curve  $\mathbf{Q}(t)$  with the tangent direction  $\dot{\mathbf{Q}}$ . That is the tangential component of the derivative of  $\Delta$ ,  $\dot{\Delta}$  is zero, or,  $\dot{\Delta}$  is in the normal space. It follows from

$$\begin{aligned} \dot{\Delta} &= \pi_N(\dot{\Delta}) \\ \Rightarrow \dot{\Delta} &= \mathbf{Q}_{\text{sym}}(\mathbf{Q}^T \dot{\Delta}) \\ \dot{\Delta} &= \mathbf{Q}(\mathbf{Q}^T \dot{\Delta} + \dot{\Delta}^T \mathbf{Q})/2 \end{aligned}$$



and

$$\begin{aligned}\overline{\Delta^T \dot{Q} + Q^T \dot{\Delta}} &= 0 \\ \dot{\Delta}^T Q + Q^T \dot{\Delta} + \Delta^T \dot{Q} + \dot{Q}^T \Delta &= 0 \\ \dot{\Delta}^T Q + Q^T \dot{\Delta} &= -\Delta^T \dot{Q} + \dot{Q}^T \Delta,\end{aligned}$$

so we have the differential equation for  $(\Delta, Q)$

$$\dot{\Delta} = -Q \left( \Delta^T \dot{Q} + \dot{Q}^T \Delta \right) / 2 \quad (8)$$

To find the geodesic  $Q(t)$  of  $\mathcal{O}(n)$ , we use the fact that  $\Delta = \dot{Q}$ , thus

$$\ddot{Q} = -Q \dot{Q}^T \dot{Q} \quad (9)$$

where the Christoffel function is given as

$$\Gamma_c(\Delta_1, \Delta_2) = Q \Delta_1^T \Delta_2 \quad (10)$$

Let  $A = Q^T \dot{Q}$ ,  $S = \dot{Q}^T \dot{Q}$  and  $B = Q^T Q = I$ . Then

$$\begin{aligned}\dot{A} &= \dot{Q}^T \dot{Q} + Q^T \ddot{Q} \\ &= S - Q^T Q S \\ &= S - B S = 0.\end{aligned}$$

Since  $A = A(0)$  is constant and  $Q Q^T = I$ , we have

$$\begin{aligned}\dot{Q} &= Q Q^T \dot{Q} = Q A \\ Q(t) &= Q(0) \exp(tA) = Q(t-1) \exp(A)\end{aligned} \quad (11)$$

### 2.3 The geodesic for Stiefel manifold

For Stiefel manifold, since

$$U^T U = I_k,$$

we have similar results:

1. The normal space  $N_U$  consists of matrices of the form

$$N = U S, \quad (12)$$

for some symmetric matrix  $S \in \mathbb{S}(k)$ .

2. The tangent space  $T_U$  consists of matrices for the form

$$\Delta = U A + U_{\perp} R, \quad (13)$$

where  $U_{\perp} = (I - U U^T)$ ,  $A = U^T \Delta$  is  $k \times k$  skew-symmetric and  $R$  is arbitrary matrix.

3. The tangential projection

$$\begin{aligned}\pi_T(\mathbf{R}) &= \mathbf{U} \text{skew}(\mathbf{U}^T \mathbf{R}) + (\mathbf{I} - \mathbf{U} \mathbf{U}^T) \mathbf{R} \\ &= \mathbf{U} \text{skew}(\mathbf{U}^T \mathbf{R}) + \mathbf{U}_\perp \mathbf{R},\end{aligned}\tag{14}$$

and the normal projection

$$\pi_N(\mathbf{R}) \equiv \mathbf{U} \text{sym}(\mathbf{U}^T \mathbf{R})\tag{15}$$

4. The differential equation for parallel vector field  $(\Delta, \mathbf{U})$

$$\dot{\Delta} = -\mathbf{U} \left( \Delta^T \dot{\mathbf{U}} + \dot{\mathbf{U}}^T \Delta \right) / 2\tag{16}$$

5. The Christoffel function is given as

$$\Gamma_c(\Delta_1, \Delta_2) = \mathbf{U} \Delta_1^T \Delta_2\tag{17}$$

thus with  $\Delta = \dot{\mathbf{U}}$ , the differential equation for the geodesic is

$$\ddot{\mathbf{U}} = -\mathbf{U} \dot{\mathbf{U}}^T \dot{\mathbf{U}}\tag{18}$$

6. Let  $\mathbf{A} = \mathbf{U}^T \dot{\mathbf{U}}$ ,  $\mathbf{S} = \dot{\mathbf{U}}^T \dot{\mathbf{U}}$  and  $\mathbf{B} = \mathbf{U}^T \ddot{\mathbf{U}} = \mathbf{I}_k$ . Then

$$\begin{aligned}\dot{\mathbf{B}} &= \mathbf{A} + \mathbf{A}^T = \mathbf{0} \\ \dot{\mathbf{A}} &= \dot{\mathbf{U}}^T \ddot{\mathbf{U}} + \mathbf{U}^T \ddot{\mathbf{U}} \\ &= \mathbf{S} - \mathbf{U}^T \mathbf{U} \mathbf{S} \\ &= \mathbf{S} - \mathbf{B} \mathbf{S} \\ \dot{\mathbf{S}} &= \ddot{\mathbf{U}}^T \dot{\mathbf{U}} + \dot{\mathbf{U}}^T \ddot{\mathbf{U}} \\ &= -\mathbf{S} \mathbf{U}^T \dot{\mathbf{U}} - \dot{\mathbf{U}}^T \mathbf{U} \mathbf{S} \\ &= -\mathbf{S} \mathbf{A} - \mathbf{A}^T \mathbf{S} \quad (\text{since } \mathbf{A}^T = -\mathbf{A}) \\ &= \mathbf{A} \mathbf{S} - \mathbf{S} \mathbf{A} \\ &= [\mathbf{A}, \mathbf{S}]\end{aligned}$$

This gives a system of differential equations

$$\begin{aligned}\dot{\mathbf{B}} &= \mathbf{A} + \mathbf{A}^T = \mathbf{0}, \\ \dot{\mathbf{A}} &= \mathbf{S} - \mathbf{B} \mathbf{S} = \mathbf{0}, \\ \dot{\mathbf{S}} &= [\mathbf{A}, \mathbf{S}], \\ \mathbf{B}(0) &= \mathbf{I}, \\ \mathbf{A}(0) &= \mathbf{A}_0, \\ \mathbf{S}(0) &= \mathbf{S}_0.\end{aligned}\tag{19}$$

Note that  $\mathbf{U} \mathbf{U}^T \neq \mathbf{I}$ , but since  $\mathbf{A} = \mathbf{A}_0$  is constant, we can solve above equations to find the following solution [Edelman et al., 1998]

$$\begin{aligned}\mathbf{B}(t) &= \mathbf{I}_k \\ \mathbf{A}(t) &= \mathbf{A}_0 \\ \mathbf{S}(t) &= \exp(t\mathbf{A}_0) \mathbf{S}_0 \exp(-t\mathbf{A}_0)\end{aligned}$$

It follows that

$$\left. \frac{d}{dt} \begin{bmatrix} \mathbf{U} \exp(t\mathbf{A}) & \dot{\mathbf{U}} \exp(t\mathbf{A}) \end{bmatrix} \right|_{t=0} = \begin{bmatrix} \mathbf{U} \exp(t\mathbf{A}) & \dot{\mathbf{U}} \exp(t\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathbf{A} & -\mathbf{S}^{(0)} \\ \mathbf{I}_k & \mathbf{A} \end{bmatrix} \quad (20)$$

with the solution by [Edelman et al., 1998] as

$$\begin{aligned}\begin{bmatrix} \mathbf{U} & \dot{\mathbf{U}} \end{bmatrix} \mathbf{I}_2 \otimes \exp(t\mathbf{A}) &= \begin{bmatrix} \mathbf{U}(0) & \dot{\mathbf{U}}(0) \end{bmatrix} (\mathbf{I}_2 \otimes \exp(t\mathbf{A})) \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{S}^{(0)} \\ \mathbf{I}_k & \mathbf{A} \end{bmatrix}\right) \\ \mathbf{U}(t) &= \begin{bmatrix} \mathbf{U}(0) & \dot{\mathbf{U}}(0) \end{bmatrix} \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{S}^{(0)} \\ \mathbf{I}_k & \mathbf{A} \end{bmatrix}\right) \mathbf{I}_{2k,k} \exp(-t\mathbf{A})\end{aligned} \quad (21)$$

Note that (21) does not contain the tangent  $\dot{\mathbf{U}}(t-1)$  and  $\mathbf{U}(t-1)$  as the update step, as shown in (11) for the orthogonal group  $\mathcal{O}(n)$ . So we have to find another way to find the geodesic.

### 3 Geometry of Stiefel manifold $V_k(\mathbb{R}^n)$

#### 3.1 Geometry of quotient manifold $X/G$

Given the differential structure of the orbit space  $X/G$ , we want to find the tangent space of  $X/G$  at the orbit  $Gx$  through  $x \in X$ .

Note that the map  $\pi : X \rightarrow X/G$  that  $\pi(x) = Gx$  is a surjective submersion. Hence for each  $x \in X$ , you can identify the tangent space  $T_{\pi(x)}(X/G)$  with the quotient of  $T_x X$  by the *tangent space of the orbit*  $Gx$  through  $x$ . The latter is the subspace of  $T_x X$  spanned by the fundamental vector fields for the action associated to elements in the Lie algebra  $\mathfrak{g}$  of  $X$ . Here  $\mathbf{A} \in \mathfrak{g}$ , the fundamental vector field is defined by  $\zeta_{\mathbf{A}}(x) = \frac{d}{dt}\big|_{t=0} (x \cdot \exp(t\mathbf{A}))$ . Note that it is a common strategy for finding the tangent space of orbit space  $X/G$ .

In [Kobayashi and Nomizu, 1963], the subspace of  $T_x X$  that are tangent to the orbit at  $x$  under action of  $G$  is called the *vertical space*. Its orthogonal complement space is called the *horizontal space*. Then the tangent space to the orbit space can be represented by the horizontal space. Intuitively, the motion along the tangent vector in the vertical space does not leave the given orbit and thus the motion that traverse across different orbits should be perpendicular to them. Let  $\zeta$  be the tangent vector of  $(X/G)$  at  $Gx$  and denote the corresponding horizontal vector at  $T_x X$  as  $\zeta \diamond x$  and called it the *horizontal lift* of  $x$ .

Dimensionality of quotient space is equal to the dimensionality of horizontal space. Dimensionality of tangent space is equal to the sum of dimensionalities of Horizontal and Vertical spaces.

We can derive a *Riemannian metric* for the quotient space  $X/G$  using the Riemannian metric of  $X$ . The idea is to define inner product between two tangent vectors  $\zeta$  and  $\eta$  at  $Gx \in X/G$  in terms of inner product between their *horizontal lifts*  $\zeta \diamond x$  and  $\eta \diamond x$  at  $x \in X$ . Note that we need to make sure that the value of the inner product between  $\zeta$  and  $\eta$  at  $Gx \in X/G$  does not depend on the  $x \in X$  used for lifting.

Note that the curve tangent at  $\zeta$  in quotient space is equal to the horizontal curve  $\alpha(t) \in X$ , which is defined to be the curve in  $X$  such that its tangent vector is horizontal lift  $\alpha'(t) = \zeta \diamond x$ .

For the Stiefel and Grassmann manifold case, we can represent the geodesic on the quotient space, i.e. Stiefel and Grassmann manifold as the horizontal geodesic on  $\mathcal{O}(n)$ . In general, if the quotient  $M/G$  is a manifold, with  $M$  being a finite dimensional complete Riemannian manifold, and  $G$  a compact Lie group acting isometrically on  $M$ , *the geodesics on the quotient can be lifted to horizontal geodesics on the original manifold*. In general, it may not be true, since  $M/G$  may not even have the differential structure. Note that the projection of horizontal geodesic onto the quotient space gives the generalized geodesic on the quotient space.

#### 3.2 The tangent space of $V_k(\mathbb{R}^n)$

We use the fact that

$$V_k(\mathbb{R}^n) \simeq \mathcal{O}(n)/\mathcal{O}(n-k).$$

To find the tangent space of  $V_k(\mathbb{R}^n)$  as the quotient manifold  $\mathcal{O}(n)/\mathcal{O}(n-k)$ , we find the horizontal space of  $\mathcal{O}(n)$ .

Note that  $\Delta \in T_Q \mathcal{O}(n)$  follows the decomposition

$$\Delta = QA$$

where  $A$  is skew-symmetric.

Since the orbit at  $Q$  under the action in  $\mathcal{O}(n-k)$  is given as

$$[Q] = \left\{ Q \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q_{n-k} \end{bmatrix} \mid Q_{n-k} \in \mathcal{O}(n-k) \right\}.$$

Denote the vertical space at  $U$  as  $\mathcal{V}_U$  and the horizontal space as  $\mathcal{H}_U$ . Note that the dimension of  $\mathcal{H}_U$  is the same as the dimension of  $\mathcal{O}(n)/\mathcal{O}(n-k)$ , i.e.  $nk - \frac{1}{2}k(k+1)$ . So the dimension of vertical space  $\mathcal{V}_U$  is

$$\begin{aligned} \frac{1}{2}n(n-1) - nk + \frac{1}{2}k(k+1) &= \frac{1}{2}(n-k)^2 - \frac{1}{2}(n-k) \\ &= \frac{1}{2}(n-k-1)(n-k). \end{aligned}$$

Let  $\Delta_H \in \mathcal{H}_U$ , then  $\Delta_H = UA_H$  for  $A_H = -A_H^T$  and  $\Delta_H$  orthogonal to  $[Q]$ , i.e.

$$\begin{aligned} \left\langle \Delta_H, U \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q_{n-k} \end{bmatrix} \right\rangle &= \text{tr} \left( A_H^T U^T U \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q_{n-k} \end{bmatrix} \right) \\ &= \text{tr} \left( \begin{bmatrix} A_{11}^H & A_{12}^H \\ A_{21}^H & A_{22}^H \end{bmatrix}^T \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q_{n-k} \end{bmatrix} \right) \\ &= 0. \quad \forall Q_{n-k} \in \mathcal{O}(n-k) \end{aligned}$$

That is

$$\Rightarrow \text{tr} (A_{22}^H Q_{n-k}) = 0. \quad \forall Q_{n-k} \in \mathcal{O}(n-k)$$

It only happens when  $A_{22}^H = 0$ . So the horizontal space and also the *tangent space* of  $V_k(\mathbb{R}^n) \simeq \mathcal{O}(n)/\mathcal{O}(n-k)$  contains matrix

$$\Delta_H = Q \begin{bmatrix} A_{11}^{(H)} & A_{12}^{(H)} \\ -(A_{12}^{(H)})^T & \mathbf{0} \end{bmatrix} \equiv Q \begin{bmatrix} A & -R^T \\ R & \mathbf{0} \end{bmatrix}$$

for arbitrary skew-symmetric  $k \times k$  matrix  $A$  and arbitrary matrix  $R$  up to a isotropy subgroup  $\begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q_{n-k} \end{bmatrix}$ .

We claim that the vertical space contains matrix

$$\Delta_V = Q \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{22}^V \end{bmatrix} \equiv Q \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C \end{bmatrix}$$

for arbitrary skew-symmetric  $(n-k) \times (n-k)$  matrix  $C$  up to a isotropy subgroup  $\begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q_{n-k} \end{bmatrix}$ .

Note that  $\text{tr}(\Delta_V^T \Delta_H) = 0$  and  $\Delta_V \in T_Q \mathcal{O}(n)$ . So the space of all matrices  $\Delta_V$  is orthogonal complement of  $\mathcal{H}_Q$ , which is by definition the vertical space  $\mathcal{V}_Q$ . See that the space of  $\Delta_V$  has dimension  $\frac{1}{2}(n-k)(n-k-1)$  that is equal to the dimension of  $\mathcal{V}_Q$

Instead of considering the quotient structure, we could use the direction definition of the Stiefel manifold. Let the tangent vector at  $Q$  be  $\bar{\Delta} = QA + Q_\perp R$ , where  $Q \in V_k(\mathbb{R}^n) \subset \mathbb{R}^{n \times k}$  and  $A = Q^T \bar{\Delta} \in \mathbb{R}^{k \times k}$  is skew-symmetric with  $R \in \mathbb{R}^{n \times k}$ ; that is in (7)

$$\bar{\Delta} = QA + Q_\perp R \in \mathbb{R}^{n \times k}$$

and compare with the horizontal vector at  $Q$

$$\begin{aligned} \Delta_H &= Q \begin{bmatrix} A & -R^T \\ R & 0 \end{bmatrix} \\ &= \begin{bmatrix} Q & Q_\perp \end{bmatrix} \begin{bmatrix} A & -R^T \\ R & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & Q_{n-k} \end{bmatrix} \\ &= \begin{bmatrix} \bar{\Delta} & -QR^T Q_{n-k} \end{bmatrix} \in \mathbb{R}^{n \times n} \end{aligned}$$

Thus it follows that

$$\bar{\Delta} = \Delta_H I_{n,k}.$$

and

$$Q^T \Delta_H = I_{n,k}^T \begin{bmatrix} A & -R^T \\ R & 0 \end{bmatrix} = \begin{bmatrix} A & -R^T \end{bmatrix}$$

### 3.3 canonical metric on $V_k(\mathbb{R}^n)$

The canonical metric on  $V_k(\mathbb{R}^n)$  is derived using from Riemannian metric of manifold  $\mathcal{O}(n)$ . That is, via the inner product between the horizontal lifts,

$$\begin{aligned} g_c(\bar{\Delta}_\zeta, \bar{\Delta}_\eta) &= g_c(\Delta_{H,\zeta} I_{n,k}, \Delta_{H,\eta} I_{n,k}) \\ &\equiv g_e(\Delta_{H,\zeta}, \Delta_{H,\eta}) \\ &= \frac{1}{2} \text{tr} \left( \left( Q \begin{bmatrix} A_\zeta & -R_\zeta^T \\ R_\zeta & 0 \end{bmatrix} \right)^T Q \begin{bmatrix} A_\eta & -R_\eta^T \\ R_\eta & 0 \end{bmatrix} \right) \\ &= \frac{1}{2} A_\zeta^T A_\eta + R_\zeta^T R_\eta, \end{aligned} \tag{22}$$

where  $\bar{\Delta}_\zeta \in \mathbb{R}^{n \times k}$  lies in the tangent space of  $V_k(\mathbb{R}^n) \simeq \mathcal{O}(n)/\mathcal{O}(n-k)$  at  $U = [Q]$ , as compared to  $\Delta_H$  and  $\Delta \in \mathbb{R}^{n \times n}$ .

Note that the Euclidean metric in  $T_U V_k(\mathbb{R}^n)$  is given as

$$\begin{aligned} g_e(\bar{\Delta}, \bar{\Delta}) &= g_e(\Delta_{H,\zeta} I_{n,k}, \Delta_{H,\eta} I_{n,k}) \\ &= \frac{1}{2} \text{tr}(\bar{\Delta}^T \bar{\Delta}) \\ &= \frac{1}{2} A^T A + \frac{1}{2} R^T R, \end{aligned}$$

as compared to the canonical metric at the point  $\mathbf{U}$  is given by

$$\begin{aligned} g_c(\bar{\Delta}, \bar{\Delta}) &= \frac{1}{2} \text{tr}(\bar{\Delta}_H^T \bar{\Delta}_H) \\ &= \text{tr} \left( \bar{\Delta}^T \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \bar{\Delta} \right) \end{aligned} \quad (23)$$

for  $\bar{\Delta} = \mathbf{U} \mathbf{A} + \mathbf{U}_\perp \mathbf{R}$ .

### 3.4 Geodesic on $V_k(\mathbb{R}^n)$

There are two ways to find the geodesic on  $V_k(\mathbb{R}^n)$ :

1. Use the definition of  $V_k(\mathbb{R}^n) = \{\mathbf{U} \in \mathbb{R}^{n \times k} \mid \mathbf{U}^T \mathbf{U} = \mathbf{I}_k\}$  and the differential equation for the geodesic in (18), we can obtain the system of differential equations in (19) and the solution is the integral curve of the equations, which is the trajectory of the geodesic, as described in (21)

$$\mathbf{U}(t) = \begin{bmatrix} \mathbf{U}(0) & \dot{\mathbf{U}}(0) \end{bmatrix} \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{S}(0) \\ \mathbf{I} & \mathbf{A} \end{bmatrix} \right) \mathbf{I}_{2k,k} \exp(-t\mathbf{A}),$$

where  $\mathbf{A} = (\mathbf{U}(0))^T \dot{\mathbf{U}}(0)$  is skew-symmetric and  $\mathbf{S}(0) = (\dot{\mathbf{U}}(0))^T \dot{\mathbf{U}}(0)$  is symmetric. [Edelman et al., 1998]

2. Based on the property of the quotient manifold, the geodesic on  $V_k(\mathbb{R}^n) \simeq \mathcal{O}(n)/\mathcal{O}(n-k)$  is the projection of the horizontal geodesic onto  $\mathcal{O}(n)/\mathcal{O}(n-k)$ . The horizontal geodesic is given by projection of geodesic on  $\mathcal{O}(n)$  to the horizontal space at each point; that is,

$$\mathbf{Q}^{(H)}(t) = \mathbf{Q}^{(H)}(0) \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \right),$$

where  $\mathbf{A}(t)$  is skew-symmetric and  $\mathbf{R}(t)$  is arbitrary. To check it, the tangent of the curve is

$$\dot{\mathbf{Q}}^{(H)}(t) = \mathbf{Q}^{(H)}(t) \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}.$$

Note that by definition,  $\mathbf{A} = (\dot{\mathbf{Q}}^{(H)}(t))^T \mathbf{Q}^{(H)}(t)$  is a constant along the geodesic.

Then the geodesic on  $V_k(\mathbb{R}^n)$  is given by the equivalent class  $[\mathbf{Q}^{(H)}(t)] = \pi \circ \mathbf{Q}^{(H)}(t)$ .

3. Use the minimal property of the geodesic as minimizing the functional

$$\begin{aligned} \min L(t, \dot{\mathbf{U}}) &= \int \sqrt{g_c(\dot{\mathbf{U}}, \dot{\mathbf{U}})} dt \\ &= \int \left( \text{tr} \left( \dot{\mathbf{U}}^T \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \dot{\mathbf{U}} \right) \right)^{1/2} dt \end{aligned}$$

$$\leq a \sqrt{\int \left( \text{tr} \left( \dot{\mathbf{U}}^T \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \dot{\mathbf{U}} \right) \right) dt}$$

using the Euler-Lagrange equation on  $E = \text{tr} \left( \dot{\mathbf{U}}^T \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \dot{\mathbf{U}} \right)$ .

Since

$$\begin{aligned} \frac{\partial}{\partial \dot{\mathbf{U}}} \text{tr} \left( \dot{\mathbf{U}}^T \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \dot{\mathbf{U}} \right) &= 2 \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \dot{\mathbf{U}} = 2\dot{\mathbf{U}} - \mathbf{U} \mathbf{U}^T \dot{\mathbf{U}} \\ &= 2\dot{\mathbf{U}} + \mathbf{U} \dot{\mathbf{U}}^T \mathbf{U} \\ \frac{\partial}{\partial \mathbf{U}} \text{tr} \left( \dot{\mathbf{U}}^T \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \dot{\mathbf{U}} \right) &= -\frac{1}{2} \frac{\partial}{\partial \mathbf{U}} \text{tr} \left( \dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} \mathbf{U}^T \right) \\ &= -\dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} \\ \frac{d}{dt} \frac{\partial}{\partial \dot{\mathbf{U}}} \text{tr} \left( \dot{\mathbf{U}}^T \left( \mathbf{I} - \frac{1}{2} \mathbf{U} \mathbf{U}^T \right) \dot{\mathbf{U}} \right) &= 2\ddot{\mathbf{U}} + \dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} + \mathbf{U} \ddot{\mathbf{U}}^T \mathbf{U} + \mathbf{U} \dot{\mathbf{U}}^T \dot{\mathbf{U}}, \end{aligned}$$

so

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{U}} - \frac{d}{dt} \frac{\partial E}{\partial \dot{\mathbf{U}}} &= 0 \\ 2\ddot{\mathbf{U}} + \dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} + \mathbf{U} \ddot{\mathbf{U}}^T \mathbf{U} + \mathbf{U} \dot{\mathbf{U}}^T \dot{\mathbf{U}} + \dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} &= 0 \\ \ddot{\mathbf{U}} + \dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} + \mathbf{U} \left( \frac{1}{2} \ddot{\mathbf{U}}^T \mathbf{U} + \frac{1}{2} \dot{\mathbf{U}}^T \dot{\mathbf{U}} \right) &= 0 \end{aligned} \tag{24}$$

$$\begin{aligned} \text{From } (\mathbf{U}^T \cdot (24)) &\Rightarrow \frac{1}{2} \mathbf{U}^T \ddot{\mathbf{U}} - \left( \dot{\mathbf{U}}^T \mathbf{U} \right)^2 - \frac{1}{2} \dot{\mathbf{U}}^T \dot{\mathbf{U}} = 0 \\ \frac{1}{2} \ddot{\mathbf{U}}^T \mathbf{U} &= \left( \mathbf{U}^T \dot{\mathbf{U}} \right)^2 + \frac{1}{2} \dot{\mathbf{U}}^T \dot{\mathbf{U}} \end{aligned} \tag{25}$$

$$(25) \Rightarrow (24), \quad \ddot{\mathbf{U}} + \dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} + \mathbf{U} \left( \left( \mathbf{U}^T \dot{\mathbf{U}} \right)^2 + \dot{\mathbf{U}}^T \dot{\mathbf{U}} \right) = 0 \tag{26}$$

Let  $\mathbf{U} = \mathbf{Q} \exp(t \mathbf{A}_H) \mathbf{I}_{n,k}$  for  $\mathbf{A}_H$  is skew-symmetric, then

$$\begin{aligned} \dot{\mathbf{U}} &= \mathbf{Q} \exp(t \mathbf{A}_H) \mathbf{A}_H \mathbf{I}_{n,k} \\ \dot{\mathbf{U}}^T \dot{\mathbf{U}} &= \mathbf{I}_{n,k}^T \mathbf{A}_H^T \exp(t \mathbf{A}_H^T) \exp(t \mathbf{A}_H) \mathbf{A}_H \mathbf{I}_{n,k} \\ &= -\mathbf{I}_{n,k}^T \mathbf{A}_H^2 \mathbf{I}_{n,k} \\ \mathbf{A} = \mathbf{U}^T \dot{\mathbf{U}} &= \mathbf{I}_{n,k}^T \exp(t \mathbf{A}_H^T) \exp(t \mathbf{A}_H) \mathbf{A}_H \mathbf{I}_{n,k} \\ &= \mathbf{I}_{n,k}^T \mathbf{A}_H \mathbf{I}_{n,k} \\ \left( \mathbf{U}^T \dot{\mathbf{U}} \right)^2 &= \mathbf{I}_{n,k}^T \mathbf{A}_H^2 \mathbf{I}_{n,k} \\ \text{so } \left( \mathbf{U}^T \dot{\mathbf{U}} \right)^2 + \dot{\mathbf{U}}^T \dot{\mathbf{U}} &= 0 \\ \text{also } \ddot{\mathbf{U}} &= \mathbf{Q} \exp(t \mathbf{A}_H) \mathbf{A}_H^2 \mathbf{I}_{n,k} \\ \text{and } \dot{\mathbf{U}} \dot{\mathbf{U}}^T \mathbf{U} &= -\mathbf{Q} \exp(t \mathbf{A}_H) \mathbf{A}_H \mathbf{I}_{n,k} \mathbf{I}_{n,k}^T \mathbf{A}_H \mathbf{I}_{n,k} \end{aligned}$$



so the equation (26) yields

$$\begin{aligned} \mathbf{A}_H (\mathbf{I}_n - \mathbf{I}_{n,k} \mathbf{I}_{n,k}^T) \mathbf{A}_H &= \mathbf{0} \\ \mathbf{A}_{H,(n-k,n-k)} &= \mathbf{0} \\ \text{that is } \mathbf{A}_H &= \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}, \end{aligned}$$

where  $\mathbf{A} = \mathbf{U}^T \dot{\mathbf{U}}$  is skew-symmetric.

$$\begin{aligned} & \dot{\mathbf{U}}^T (\mathbf{I} - \mathbf{U} \mathbf{U}^T) \dot{\mathbf{U}} \\ &= \mathbf{I}_{n,k}^T \mathbf{A}_H^T (\mathbf{I} - \mathbf{I}_{n,k} \mathbf{I}_{n,k}^T) \mathbf{A}_H \mathbf{I}_{n,k} \\ &= \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \end{bmatrix} \begin{bmatrix} -\mathbf{A} & \mathbf{R}^T \\ -\mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix} \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} -\mathbf{A} & \mathbf{R}^T \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{R} \end{bmatrix} \\ &= \mathbf{R}^T \mathbf{R}. \end{aligned}$$

The trajectory of the geodesic is given as

$$\mathbf{U}(t) = \mathbf{Q} \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \right) \mathbf{I}_{n,k}. \quad (27)$$

Note that

$$\begin{aligned} \mathbf{U}_\perp(0) \dot{\mathbf{U}}(0) &= \mathbf{Q} \mathbf{I}_{n,n-k} \mathbf{I}_{n,n-k}^T \mathbf{Q}^T \mathbf{Q} \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \mathbf{I}_{n,k} \\ &= \mathbf{Q} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix} \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0} \end{bmatrix} \\ &= \mathbf{Q} \mathbf{R} \end{aligned}$$

is the QR-decomposition of the matrix  $\mathbf{U}_\perp(0) \dot{\mathbf{U}}(0)$ . [Edelman et al., 1998]

4. To express  $\mathbf{U}(t)$  using  $\mathbf{U}(0)$  and  $\dot{\mathbf{U}}(0)$ , we see that  $\mathbf{A} = \mathbf{U}(0)^T \dot{\mathbf{U}}(0)$  and  $\mathbf{R}^T \mathbf{R} = \dot{\mathbf{U}}(0)^T (\mathbf{I} - \mathbf{U}(0) \mathbf{U}(0)^T) \dot{\mathbf{U}}(0)$ . From (27), we see that  $\mathbf{U}(0) = \mathbf{Q} \mathbf{I}_{n,k}$  and  $\mathbf{U}_\perp(0) = \mathbf{I} - \mathbf{U}(0) \mathbf{U}(0)^T = \mathbf{Q} (\mathbf{I} - \mathbf{I}_{n,k} \mathbf{I}_{n,k}^T) \mathbf{Q}^T = \mathbf{Q} \mathbf{I}_{n,n-k} \mathbf{I}_{n,n-k}^T \mathbf{Q}^T$ .

Let  $\mathbf{M}(t) = \mathbf{I}_{n,k}^T \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \right) \mathbf{I}_{n,k}$ , then

$$\begin{aligned} \mathbf{U}(t) &= \mathbf{Q} \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \right) \mathbf{I}_{n,k} \\ &= \mathbf{Q} \mathbf{I}_{n,k} \mathbf{I}_{n,k}^T \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \right) \mathbf{I}_{n,k} + \mathbf{Q} \mathbf{I}_{n,n-k} \mathbf{I}_{n,n-k}^T \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \right) \mathbf{I}_{n,k} \\ &= \mathbf{U}(0) \mathbf{M}(t) + \mathbf{Q} \mathbf{I}_{n,n-k} \mathbf{I}_{n,n-k}^T \mathbf{Q}^T \mathbf{Q} \exp \left( t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \right) \mathbf{I}_{n,k} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{U}(0)\mathbf{M}(t) + \mathbf{U}_\perp(0)\mathbf{U}(t) \\
&= \mathbf{U}(0)\mathbf{M}(t) + \mathbf{U}_\perp(0) \int_0^t \dot{\mathbf{U}}(s)ds \\
&= \mathbf{U}(0)\mathbf{M}(t) + \mathbf{U}_\perp(0) \int_0^t \dot{\mathbf{U}}(s)ds \\
&= \mathbf{U}(0)\mathbf{M}(t) + \mathbf{U}_\perp(0)\dot{\mathbf{U}}(0) \int_0^t \mathbf{M}(s)ds
\end{aligned}$$

Denote  $\mathbf{U} = \mathbf{U}(0)$  and  $\mathbf{H} = \dot{\mathbf{U}}(0)$ , then the geodesic trajectory is given as

$$\mathbf{U}(t) = \mathbf{U}\mathbf{M}(t) + (\mathbf{I} - \mathbf{U}\mathbf{U}^T) \mathbf{H} \int_0^t \mathbf{M}(s)ds \quad (28)$$

5. To further simplify (28), define  $\mathbf{N}(t) = \begin{bmatrix} 0 & \mathbf{I}_{n-k} \end{bmatrix} \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}\right) \mathbf{I}_{n,k} \in bR^{(n-k) \times k}$  so that

$$\begin{aligned}
\begin{bmatrix} \mathbf{M}(t) \\ \mathbf{N}(t) \end{bmatrix} &= \begin{bmatrix} \mathbf{I}_{n,k}^T \\ \mathbf{I}_{n,n-k}^T \end{bmatrix} \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}\right) \mathbf{I}_{n,k} \\
&= \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}\right) \mathbf{I}_{n,k}.
\end{aligned}$$

See that

$$\begin{aligned}
\dot{\mathbf{N}}(t) &= \begin{bmatrix} 0 & \mathbf{I}_{n-k} \end{bmatrix} \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}\right) \mathbf{I}_{n,k} \\
&= \begin{bmatrix} \mathbf{R} & \mathbf{0} \end{bmatrix} \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}\right) \mathbf{I}_{n,k} \\
&= \mathbf{R}\mathbf{I}_{n,k}^T \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}\right) \mathbf{I}_{n,k} \\
&= \mathbf{R}\mathbf{M}(t) \\
\Rightarrow \mathbf{N}(t) &= \mathbf{R} \int_0^t \mathbf{M}(s)ds.
\end{aligned}$$

So the geodesic curve in (28) can be represented as

$$\mathbf{U}(t) = \mathbf{U}\mathbf{M}(t) + \mathbf{Q}\mathbf{N}(t) \quad (29)$$

where

$$\begin{bmatrix} \mathbf{M}(t) \\ \mathbf{N}(t) \end{bmatrix} = \exp\left(t \begin{bmatrix} \mathbf{A} & -\mathbf{R}^T \\ \mathbf{R} & \mathbf{0} \end{bmatrix}\right) \mathbf{I}_{n,k}$$

and

$$(\mathbf{I} - \mathbf{U}\mathbf{U}^T) \mathbf{H} = \mathbf{Q}\mathbf{R}$$

is the Q-R decomposition.

### 3.5 parallel translation

From the equation in (26)

$$\ddot{\mathbf{U}} + \dot{\mathbf{U}}\dot{\mathbf{U}}^T\mathbf{U} + \mathbf{U}\left(\left(\mathbf{U}^T\dot{\mathbf{U}}\right)^2 + \dot{\mathbf{U}}^T\dot{\mathbf{U}}\right) = 0,$$

we see that the Christoffel function under the canonical metric is given as

$$\begin{aligned}\Gamma_c(\mathbf{\Delta}, \mathbf{\Delta}) &= \mathbf{\Delta}\mathbf{\Delta}^T\mathbf{U} + \mathbf{U}\left(-\mathbf{\Delta}^T\mathbf{U}\mathbf{U}^T\mathbf{\Delta} + \mathbf{\Delta}^T\mathbf{\Delta}\right) \\ &= \mathbf{\Delta}\mathbf{\Delta}^T\mathbf{U} + \mathbf{U}\mathbf{\Delta}^T\left(\mathbf{I} - \mathbf{U}\mathbf{U}^T\right)\mathbf{\Delta}.\end{aligned}\tag{30}$$

The parallel translation satisfies the differential equation

$$\dot{\mathbf{\Delta}} + \Gamma_c\left(\mathbf{\Delta}, \dot{\mathbf{U}}\right) = 0\tag{31}$$

### 3.6 Gradient function on Stiefel manifold

The gradient of function  $F : V_k(\mathbb{R}^n) \rightarrow \mathbb{R}$  is given by projection of the Euclidean gradient on the tangent plane of the surface, i.e. for all tangent vector  $\mathbf{\Delta}$  at  $\mathbf{U}$ , the gradient  $\nabla F$  is given so that the equation holds

$$\begin{aligned}\text{tr}\left(\frac{\partial F^T}{\partial \mathbf{U}}\mathbf{\Delta}\right) &= g_c(\nabla F, \mathbf{\Delta}) \\ &= \text{tr}\left((\nabla F)^T\left(\mathbf{I} - \frac{1}{2}\mathbf{U}\mathbf{U}^T\right)\mathbf{\Delta}\right),\end{aligned}\tag{32}$$

where  $\mathbf{U}^T(\nabla F)$  is skew-symmetric.

We see that

$$(\nabla F)^T\left(\mathbf{I} - \frac{1}{2}\mathbf{U}\mathbf{U}^T\right)\mathbf{\Delta} = (\nabla F)^T\mathbf{\Delta} + \frac{1}{2}\mathbf{U}^T(\nabla F)\mathbf{U}^T\mathbf{\Delta}$$

So

$$\begin{aligned}\text{tr}\left(\left(\frac{\partial F^T}{\partial \mathbf{U}} - (\nabla F)^T - \frac{1}{2}\mathbf{U}^T(\nabla F)\mathbf{U}^T\right)\mathbf{\Delta}\right) &= 0 \\ \frac{\partial F}{\partial \mathbf{U}} - (\nabla F) - \frac{1}{2}\mathbf{U}(\nabla F)^T\mathbf{U} &= \mathbf{U} \\ \frac{\partial F^T}{\partial \mathbf{U}}\mathbf{U} - (\nabla F)^T\mathbf{U} - \frac{1}{2}\mathbf{U}^T(\nabla F) &= \mathbf{I} \\ \frac{\partial F^T}{\partial \mathbf{U}}\mathbf{U} - \frac{1}{2}(\nabla F)^T\mathbf{U} &= \mathbf{I} \\ \frac{\partial F}{\partial \mathbf{U}} - (\nabla F) + \mathbf{U}\left(\mathbf{I} - \frac{\partial F^T}{\partial \mathbf{U}}\mathbf{U}\right) &= \mathbf{U} \\ \frac{\partial F}{\partial \mathbf{U}} - (\nabla F) - \mathbf{U}\frac{\partial F^T}{\partial \mathbf{U}}\mathbf{U} &= 0\end{aligned}$$

and the gradient  $\nabla F$  is

$$(\nabla F) = \frac{\partial F}{\partial \mathbf{U}} - \mathbf{U}\frac{\partial F^T}{\partial \mathbf{U}}\mathbf{U}\tag{33}$$

### 3.7 Hessian of function

The Hessian is defined as

$$\text{Hess } F(\Delta, \Delta) = \left. \frac{d^2}{dt^2} \right|_{t=0} F(\mathbf{U}(t)) \quad (34)$$

where  $\mathbf{U}(t)$  is the geodesic with tangent  $\dot{\mathbf{U}} = \Delta$ .

Note that

$$\left. \frac{d}{dt} \right|_{t=0} F(\mathbf{U}(t)) = \text{tr} \left( \frac{\partial F^T}{\partial \mathbf{U}} \dot{\mathbf{U}}(t) \right) \Big|_{t=0} \quad (35)$$

$$= \text{tr} \left( \frac{\partial F^T}{\partial \mathbf{U}} \Delta \right) \quad (36)$$

$$\begin{aligned} \text{Hess } F(\Delta, \Delta) &= \left. \frac{d^2}{dt^2} \right|_{t=0} F(\mathbf{U}(t)) \\ &= \dot{\mathbf{U}}(0)^T F_{UU} \dot{\mathbf{U}}(0) + \text{tr} \left( \frac{\partial F^T}{\partial \mathbf{U}} \ddot{\mathbf{U}}(0) \right) \\ &= \Delta^T F_{UU} \Delta - \text{tr} \left( \frac{\partial F^T}{\partial \mathbf{U}} \Gamma_c(\Delta, \Delta) \right) \end{aligned} \quad (37)$$

For  $\Delta_1$  and  $\Delta$  are tangent vectors,

$$\text{Hess } F(\Delta_1, \Delta_2) = \Delta_1^T F_{UU} \Delta_2 - \text{tr} \left( \frac{\partial F^T}{\partial \mathbf{U}} \Gamma_c(\Delta_1, \Delta_2) \right) \quad (38)$$

where

$$\Gamma_c(\Delta_1, \Delta_2) = \Delta_1 \Delta_2^T \mathbf{U} + \mathbf{U} \Delta_1^T (\mathbf{I} - \mathbf{U} \mathbf{U}^T) \Delta_2.$$

The inverse of Hessian times the gradient is given by solving the equation

$$\text{Hess } F(\Delta, \mathbf{Y}) = g_c(-\nabla F, \mathbf{X}), \quad \forall \mathbf{X} \text{ is the tangent vector of } V_k(\mathbb{R}^n) \quad (39)$$

for  $\Delta$  is the tangent vector of  $V_k(\mathbb{R}^n)$  and is denoted as  $-(\text{Hess } F)^{-1} \nabla F$

## 4 Geometry of Grassmann manifold

### 4.1 The tangent space of $Gr(k, n)$

1. The point in  $Gr(k, n) \simeq \mathcal{O}(n)/(\mathcal{O}(k) \times \mathcal{O}(n-k))$  is the equivalence class

$$[Q] = \left\{ Q \begin{bmatrix} Q_k & 0 \\ 0 & Q_{n-k} \end{bmatrix} \mid Q_k \in \mathcal{O}(k), Q_{n-k} \in \mathcal{O}(n-k) \right\}.$$

As in the Stiefel manifold, we can obtain the tangent space of  $Gr(k, n)$  by the horizontal space of  $\mathcal{O}(n)$ . Let  $\Delta_H = QA_H$ , where  $A_H = \begin{bmatrix} A_{11}^H & -R_H^T \\ R_H & A_{22}^H \end{bmatrix}$  is skew-symmetric, be the tangent vector in the horizontal space  $\mathcal{H}$  at  $Q$ . By definition, it should be orthogonal to the orbit. So

$$\begin{aligned} & \text{tr} \left( A_H^T Q^T Q \begin{bmatrix} Q_k & 0 \\ 0 & Q_{n-k} \end{bmatrix} \right) \\ &= \text{tr} \left( \begin{bmatrix} -A_{11}^H & R_H^T \\ -R_H & -A_{22}^H \end{bmatrix} \begin{bmatrix} Q_k & 0 \\ 0 & Q_{n-k} \end{bmatrix} \right) = \text{tr} \left( \begin{bmatrix} -A_{11}^H Q_k & R_H^T Q_k \\ -R_H Q_{n-k} & -A_{22}^H Q_{n-k} \end{bmatrix} \right) \\ &= -\text{tr}(A_{11}^H Q_k) - \text{tr}(A_{22}^H Q_{n-k}) \\ &= \text{tr}((A_{11}^H)^T Q_k) + \text{tr}((A_{22}^H)^T Q_{n-k}) = 0, \quad \forall Q_k \in \mathcal{O}(k), Q_{n-k} \in \mathcal{O}(n-k) \end{aligned}$$

We see that if  $\text{tr}((A_{11}^H)^T Q_k) = 0$  and  $\text{tr}((A_{22}^H)^T Q_{n-k}) = 0$  for all  $Q_k \in \mathcal{O}(k)$ ,  $Q_{n-k} \in \mathcal{O}(n-k)$ , the above expression is satisfied. This means that  $A_{11}^H = 0$  and  $A_{22}^H = 0$ . The horizontal space  $\mathcal{H}$  contains matrix of the form

$$\Delta_H = Q \begin{bmatrix} 0 & -R^T \\ R & 0 \end{bmatrix}$$

up to a postmultiplication of  $\begin{bmatrix} Q_k & 0 \\ 0 & Q_{n-k} \end{bmatrix}$ . Note that the dimension of the horizontal space is  $k(n-k)$  that is equal to the dimension of  $Gr(k, n)$ .

We claim that the vertical space  $\mathcal{V}$  contains all matrices of the form

$$\Delta_V = Q \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$$

up to a postmultiplication of  $\begin{bmatrix} Q_k & 0 \\ 0 & Q_{n-k} \end{bmatrix}$ , where  $A, C$  are skew-symmetric of dimension  $k$ -by- $k$  and  $(n-k)$ -by- $(n-k)$ . See that  $\Delta_V$  is tangent to  $\mathcal{O}(n)$  at  $Q$  and it is orthogonal to all matrices in  $\mathcal{H}$ .

Note that the dimensionalities of all skew-symmetric matrices  $A_{11}^V \in \mathbb{R}^{k \times k}$  and any  $A_{22}^V \in \mathbb{R}^{(n-k) \times (n-k)}$  are  $\frac{k(k-1)}{2}$  and  $\frac{(n-k)(n-k-1)}{2}$ , respectively. Note that the dimension of the vertical space is given by

$$\begin{aligned} \frac{n(n-1)}{2} - k(n-k) &= \frac{1}{2}n^2 - nk + \frac{1}{2}k^2 - \frac{1}{2}n + \frac{1}{2}k + \frac{1}{2}k^2 - \frac{1}{2}k \\ &= \frac{(n-k)(n-k-1)}{2} + \frac{k(k-1)}{2}. \end{aligned}$$

2. The second form is given by  $Gr(k, n) \simeq V_k(\mathbb{R}^n)/\mathcal{O}(k)$  with the point as  $[U] = U\mathcal{O}(k)$ . Since the tangent vector of  $U \in V_k(\mathbb{R}^n)$  is given as

$$\Sigma = UA + U_\perp R$$

for  $A \in \mathbb{R}^{k \times k}$  is skew-symmetric, it is clear by above decomposition that the vertical space at  $U$  is given by

$$\Sigma_V = UA$$

and the horizontal space is given as

$$\Sigma_H = U_\perp R,$$

which is equivalent to all  $n$ -by- $k$  matrices that

$$U^T \Sigma_H = 0.$$

## 4.2 canonical metric on $Gr(k, n)$

Let  $\bar{\Delta}_\zeta, \bar{\Delta}_\eta \in \mathbb{R}^{(n-k) \times k}$  be two tangent vectors of  $[U] \in Gr(k, n) \simeq V_k(\mathbb{R}^n)/\mathcal{O}(k)(\mathbb{R}^n)$ . The canonical metric is defined via the induced canonical metric in the horizontal space of  $V_k(\mathbb{R}^n)$ , i.e.

$$\begin{aligned} g_c(\bar{\Delta}_\zeta, \bar{\Delta}_\eta) &= g_c(\Sigma_{H,\zeta}, \Sigma_{H,\eta}) \\ &= \text{tr} \left( \Sigma_{H,\zeta}^T \left( I - \frac{1}{2} U U^T \right) \Sigma_{H,\eta} \right) \\ &= \text{tr} (\Sigma_{H,\zeta}^T \Sigma_{H,\eta}) \\ &= g_e(\Sigma_{H,\zeta}, \Sigma_{H,\eta}), \end{aligned} \tag{40}$$

which is equal to the Euclidean metric.

It can also be derived using the form

$$\Delta_H = Q \begin{bmatrix} 0 & -R^T \\ R & 0 \end{bmatrix}$$

and

$$\begin{aligned} g_c(\bar{\Delta}_\zeta, \bar{\Delta}_\eta) &= g_e(\Delta_{H,\zeta}, \Delta_{H,\eta}) \\ &= \frac{1}{2} \text{tr} \left( \begin{bmatrix} 0 & R_\zeta^T \\ -R_\zeta & 0 \end{bmatrix} \begin{bmatrix} 0 & -R_\eta^T \\ R_\eta & 0 \end{bmatrix} \right) \\ &= \text{tr} (R_\zeta^T R_\eta) \\ &= g_e(\bar{\Delta}_\zeta, \bar{\Delta}_\eta), \end{aligned} \tag{41}$$

where  $\bar{\Delta}_\zeta = R_\zeta$ .

### 4.3 Geodesic on $Gr(k, n)$

1. [Edelman et al., 1998]

As in Stiefel manifold, the geodesic on  $Gr(k, n)$  is given by restricting the horizontal geodesic on the orbit. Note that given the horizontal vector, the horizontal geodesic follows the differential equation

$$\dot{\mathbf{Q}}(t) = \mathbf{Q}(t) \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix},$$

and then the horizontal geodesic trajectory gives as

$$\mathbf{Q}(t) = \mathbf{Q}(0) \exp \left( t \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix} \right) \quad (42)$$

Therefore, the geodesic of Grassmann manifold is given as  $[\mathbf{Q}(t)]$  and it gives

$$\mathbf{U}(t) = [\mathbf{U}(0) \quad \mathbf{U}_\perp(0)] \exp \left( t \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix} \right) \mathbf{I}_{n,k} \quad (43)$$

2. Note that let SVD of  $\mathbf{R}$  is given by  $[\mathbf{W}(0) \quad \mathbf{W}_\perp(0)] \begin{pmatrix} \Sigma_R \\ 0 \end{pmatrix} \mathbf{V}^T$ , then

$$\begin{aligned} \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix} &= \begin{bmatrix} 0 & \mathbf{V} \begin{pmatrix} -\Sigma_R & 0 \end{pmatrix} \begin{bmatrix} \mathbf{W}(0)^T \\ \mathbf{W}_\perp(0)^T \end{bmatrix} \\ [\mathbf{W}(0) \quad \mathbf{W}_\perp(0)] \begin{pmatrix} \Sigma_R \\ 0 \end{pmatrix} \mathbf{V}^T & 0 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V} & 0 & 0 \\ 0 & \mathbf{W}(0) & \mathbf{W}_\perp(0) \end{bmatrix} \begin{bmatrix} 0 & -\Sigma_R & 0 \\ \Sigma_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}^T & 0 \\ 0 & \mathbf{W}^T(0) \\ 0 & \mathbf{W}_\perp^T(0) \end{bmatrix} \end{aligned}$$

Therefore

$$\begin{aligned} \mathbf{U}(t) &= [\mathbf{U}(0) \quad \mathbf{U}_\perp(0)] \exp \left( t \begin{bmatrix} \mathbf{V} & 0 & 0 \\ 0 & \mathbf{W}(0) & \mathbf{W}_\perp(0) \end{bmatrix} \begin{bmatrix} 0 & -\Sigma_R & 0 \\ \Sigma_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}^T & 0 \\ 0 & \mathbf{W}^T(0) \\ 0 & \mathbf{W}_\perp^T(0) \end{bmatrix} \right) \mathbf{I}_{n,k} \\ &= [\mathbf{U}(0) \quad \mathbf{U}_\perp(0)] \begin{bmatrix} \mathbf{V} & 0 & 0 \\ 0 & \mathbf{W}(0) & \mathbf{W}_\perp(0) \end{bmatrix} \exp \left( t \begin{bmatrix} 0 & -\Sigma_R & 0 \\ \Sigma_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{V}^T & 0 \\ 0 & \mathbf{W}^T(0) \\ 0 & \mathbf{W}_\perp^T(0) \end{bmatrix} \mathbf{I}_{n,k} \\ &= [\mathbf{U}(0)\mathbf{V} \quad \mathbf{U}_\perp(0)\mathbf{W}(0) \quad \mathbf{U}_\perp(0)\mathbf{W}_\perp(0)] \exp \left( t \begin{bmatrix} 0 & -\Sigma_R & 0 \\ \Sigma_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{V}^T \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

and

$$\dot{\mathbf{U}}(t) = [\mathbf{U}(0)\mathbf{V} \quad \mathbf{U}_\perp(0)\mathbf{W}(0) \quad \mathbf{U}_\perp(0)\mathbf{W}_\perp(0)] \begin{bmatrix} 0 & -\Sigma_R & 0 \\ \Sigma_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \exp \left( t \begin{bmatrix} 0 & -\Sigma_R & 0 \\ \Sigma_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{V}^T \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} \mathbf{U}_\perp(0)\mathbf{W}(0)\boldsymbol{\Sigma}_R & -\mathbf{U}(0)\mathbf{V}\boldsymbol{\Sigma}_R & 0 \end{bmatrix} \exp \left( t \begin{bmatrix} 0 & -\boldsymbol{\Sigma}_R & 0 \\ \boldsymbol{\Sigma}_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{V}^T \\ 0 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{U}_\perp(0)\mathbf{W}(0)\boldsymbol{\Sigma}_R & -\mathbf{U}(0)\mathbf{V}\boldsymbol{\Sigma}_R \end{bmatrix} \exp \left( t \begin{bmatrix} 0 & -\boldsymbol{\Sigma}_R \\ \boldsymbol{\Sigma}_R & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{V}^T \\ 0 \end{bmatrix}
\end{aligned}$$

Thus

$$\begin{aligned}
\dot{\mathbf{U}}(0) \equiv \mathbf{H} &= \begin{bmatrix} \mathbf{U}_\perp(0)\mathbf{W}(0)\boldsymbol{\Sigma}_R & -\mathbf{U}(0)\mathbf{V}\boldsymbol{\Sigma}_R \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \\ 0 \end{bmatrix} \\
&= (\mathbf{U}_\perp(0)\mathbf{W}(0)) \boldsymbol{\Sigma}_R \mathbf{V}^T \\
&= \mathbf{U}_H \boldsymbol{\Sigma}_R \mathbf{V}^T
\end{aligned} \tag{44}$$

Note that

$$\begin{aligned}
\exp \left( t \begin{bmatrix} 0 & -\boldsymbol{\Sigma}_R & 0 \\ \boldsymbol{\Sigma}_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) &= \sum_{k=0}^{\infty} \frac{1}{k!} \begin{bmatrix} 0 & -t\boldsymbol{\Sigma}_R & 0 \\ t\boldsymbol{\Sigma}_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^k \\
&= \sum_{k=0}^{\infty} \frac{1}{(2k)!} \begin{bmatrix} 0 & -t\boldsymbol{\Sigma}_R & 0 \\ t\boldsymbol{\Sigma}_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^{2k} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \begin{bmatrix} 0 & -t\boldsymbol{\Sigma}_R & 0 \\ t\boldsymbol{\Sigma}_R & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^{2k+1} \\
&= I + \sum_{k=1}^{\infty} \frac{1}{(2k)!} \begin{bmatrix} (-1)^k (t\boldsymbol{\Sigma}_R)^{2k} & 0 & 0 \\ 0 & (-1)^k (t\boldsymbol{\Sigma}_R)^{2k} & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&\quad + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!} \begin{bmatrix} 0 & (-1)^{k+1} (t\boldsymbol{\Sigma}_R)^{2k+1} & 0 \\ (-1)^k (t\boldsymbol{\Sigma}_R)^{2k+1} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} (t\boldsymbol{\Sigma}_R)^{2k} & 0 & 0 \\ 0 & \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} (t\boldsymbol{\Sigma}_R)^{2k} & 0 \\ 0 & 0 & I \end{bmatrix} \\
&\quad + \begin{bmatrix} 0 & \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{(2k+1)!} (t\boldsymbol{\Sigma}_R)^{2k+1} & 0 \\ \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} (t\boldsymbol{\Sigma}_R)^{2k+1} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \cos t\boldsymbol{\Sigma}_R & -\sin t\boldsymbol{\Sigma}_R & 0 \\ \sin t\boldsymbol{\Sigma}_R & \cos t\boldsymbol{\Sigma}_R & 0 \\ 0 & 0 & I \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathbf{U}(t) &= \begin{bmatrix} \mathbf{U}(0)\mathbf{V} & \mathbf{U}_\perp(0)\mathbf{W}(0) & \mathbf{U}_\perp(0)\mathbf{W}_\perp(0) \end{bmatrix} \begin{bmatrix} \cos t\boldsymbol{\Sigma}_R & -\sin t\boldsymbol{\Sigma}_R & 0 \\ \sin t\boldsymbol{\Sigma}_R & \cos t\boldsymbol{\Sigma}_R & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \\ 0 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{U}(0)\mathbf{V} & \mathbf{U}_\perp(0)\mathbf{W}(0) & \mathbf{U}_\perp(0)\mathbf{W}_\perp(0) \end{bmatrix} \begin{bmatrix} \cos t\boldsymbol{\Sigma}_R \\ \sin t\boldsymbol{\Sigma}_R \\ 0 \end{bmatrix} \mathbf{V}^T
\end{aligned}$$



$$\begin{aligned}
&= \begin{bmatrix} \mathbf{U}(0)\mathbf{V} & \mathbf{U}_\perp(0)\mathbf{W}(0) \end{bmatrix} \begin{bmatrix} \cos t\boldsymbol{\Sigma}_R \\ \sin t\boldsymbol{\Sigma}_R \end{bmatrix} \mathbf{V}^T \\
&= \begin{bmatrix} \mathbf{U}(0)\mathbf{V} & \mathbf{U}_H \end{bmatrix} \begin{bmatrix} \cos t\boldsymbol{\Sigma}_R \\ \sin t\boldsymbol{\Sigma}_R \end{bmatrix} \mathbf{V}^T
\end{aligned}$$

Thus the geodesic of Grassmann manifold is given as

$$\dot{\mathbf{U}}(t) = \begin{bmatrix} \mathbf{U}(0)\mathbf{V} & \mathbf{U}_H \end{bmatrix} \begin{bmatrix} \cos t\boldsymbol{\Sigma}_R \\ \sin t\boldsymbol{\Sigma}_R \end{bmatrix} \mathbf{V}^T \quad (45)$$

where

$$\begin{aligned}
\dot{\mathbf{U}}(0) &\equiv \mathbf{H} = (\mathbf{U}_\perp(0)\mathbf{W}(0)) \boldsymbol{\Sigma}_R \mathbf{V}^T \\
&= \mathbf{U}_H \boldsymbol{\Sigma}_R \mathbf{V}^T
\end{aligned}$$

and  $(\mathbf{W}(0), \boldsymbol{\Sigma}_R, \mathbf{V})$  is economic SVD of  $\mathbf{R}$ , which is given by  $\mathbf{R} = \mathbf{W}(0)\boldsymbol{\Sigma}_R\mathbf{V}^T$ , where  $\mathbf{W}(0) \in \mathbb{R}^{(n-k) \times k}$ ,  $\mathbf{U}_\perp(0) \in \mathbb{R}^{n \times (n-k)}$ .

We make comments on the geodesic on Grassmann manifold in (45):

•

$$\dot{\mathbf{U}}(t) = \begin{bmatrix} \mathbf{U}(0)\mathbf{V} & \mathbf{U}_H \end{bmatrix} \begin{bmatrix} \cos t\boldsymbol{\Sigma}_R \\ \sin t\boldsymbol{\Sigma}_R \end{bmatrix} \mathbf{V}^T,$$

where  $t\boldsymbol{\Sigma}_R = \begin{bmatrix} \phi_1 & & & \\ & \phi_2 & & \\ & & \dots & \\ & & & \phi_k \end{bmatrix}$ . The angle  $\phi_k$  is the  $k$ -th *principal angle* btw the subspaces spanned by  $\mathbf{U}(t)$  and that by origin  $\mathbf{I}_{n,k}$ .

- In general, let  $\mathbf{U} \cos(\boldsymbol{\Theta}) \mathbf{V}^T$  be the SVD of the matrix  $\mathbf{U}_1^T \mathbf{U}_2$ , where  $\boldsymbol{\Theta}$  is the diagonal matrix of principal angles.
- The *arc length* btw  $\mathbf{U}(0)$  and  $\mathbf{U}(t)$  is given by

$$d(\mathbf{U}(0), \mathbf{U}(t)) = t \left\| \dot{\mathbf{U}}(0) \right\|_F = t \sqrt{\sum_{i=1}^k \sigma_i^2} \quad (t \text{ small}). \quad (46)$$

Similarly, one could use the *projection Frobenius-norm* as

$$d_{pF}(\mathbf{U}(s), \mathbf{U}(t)) = \left\| \mathbf{U}(s)\mathbf{U}(s)^T - \mathbf{U}(t)\mathbf{U}(t)^T \right\|_F \quad (47)$$

or the *chordal Frobenius-norm*

$$d_{cF}(\mathbf{U}(s), \mathbf{U}(t)) = \min_{\mathbf{Q}_1, \mathbf{Q}_2 \in \mathcal{O}(k)} \left\| \mathbf{U}(s)\mathbf{Q}_1 - \mathbf{U}(t)\mathbf{Q}_2 \right\|_F \quad (48)$$

The following inequality holds

$$d(\mathbf{U}(s), \mathbf{U}(t)) > d_{cF}(\mathbf{U}(s), \mathbf{U}(t)) > d_{pF}(\mathbf{U}(s), \mathbf{U}(t)), \quad (49)$$

which due to the fact that the chordal F-norm is obtained by embedding the Grassmann manifold into  $\mathbb{R}^{n \times k}$ , whereas the projection Frobenius-norm

#### 4.4 parallel translation

The parallel translation of  $\Delta$  is given by

$$\tau\Delta(t) = \begin{bmatrix} U(0) & U_{\perp}(0) \end{bmatrix} \exp\left(t \begin{bmatrix} 0 & -\mathbf{R}_A^T \\ \mathbf{R}_A & 0 \end{bmatrix}\right) \begin{bmatrix} 0 \\ \mathbf{R}_{\Delta} \end{bmatrix} \quad (50)$$

where  $\Delta = U_{\perp}^T \mathbf{R}_{\Delta}$  and  $H = U_{\perp}^T \mathbf{R}_A$

So we obtain the parallel transport as

$$\tau\Delta(t) = \begin{bmatrix} UV & U_H \end{bmatrix} \begin{bmatrix} \cos \Sigma_R \\ \sin \Sigma_R \end{bmatrix} U_H^T \Delta + (I - U_H U_H^T) \Delta \quad (51)$$

#### 4.5 Gradient function on Grassmann manifold

The gradient of function  $F : Gr(k, n) \rightarrow \mathbb{R}$  is given by projection of the Euclidean gradient on the tangent plane of the surface, i.e. for all tangent vector  $\Delta$  at  $U$ , the gradient  $\nabla F$  is given so that the equation holds

$$\begin{aligned} \text{tr} \left( \frac{\partial F^T}{\partial U} \Delta \right) &= g_c(\nabla F, \Delta) \\ &= \text{tr} \left( (\nabla F)^T \Delta \right), \end{aligned}$$

where  $U^T (\nabla F) = 0$ .

We see that

$$\text{tr} \left( \left( \frac{\partial F^T}{\partial U} - (\nabla F)^T \right) \Delta \right) = 0$$

So

$$\begin{aligned} \frac{\partial F^T}{\partial U} - (\nabla F)^T &= \mathbf{R} U^T \\ \frac{\partial F^T}{\partial U} U &= \mathbf{R} \end{aligned}$$

thus,

$$(\nabla F) = \frac{\partial F}{\partial U} - U U^T \frac{\partial F}{\partial U} \quad (52)$$

#### 4.6 Hessian of function

For the geodesic  $U(t)$ ,

$$U(t) = Q \exp\left(t \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix}\right) I_{n,k}$$

$$\begin{aligned}
\dot{U}(t) &= \mathbf{Q} \exp \left( t \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix} \right) \begin{bmatrix} 0 \\ \mathbf{R} \end{bmatrix} \\
\dot{U}(t)^T \dot{U}(t) &= \begin{bmatrix} 0 & \mathbf{R}^T \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{R} \end{bmatrix} \\
&= \mathbf{R}^T \mathbf{R} \\
\ddot{U}(t) &= \mathbf{Q} \exp \left( t \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix} \right) \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{R} \end{bmatrix} \\
&= \mathbf{Q} \exp \left( t \begin{bmatrix} 0 & -\mathbf{R}^T \\ \mathbf{R} & 0 \end{bmatrix} \right) \begin{bmatrix} -\mathbf{R}^T \mathbf{R} \\ 0 \end{bmatrix} \\
&= -\mathbf{U}(t) \dot{U}(t)^T \dot{U}(t) \\
&= -\Gamma_c \left( \dot{U}(t), \dot{U}(t) \right)
\end{aligned}$$

Apply the formula (38) and we obtain

$$\text{Hess } F(\mathbf{\Delta}_1, \mathbf{\Delta}_2) = \mathbf{\Delta}_1^T F_{UU} \mathbf{\Delta}_2 - \text{tr} \left( \frac{\partial F}{\partial \mathbf{U}}^T \Gamma_c(\mathbf{\Delta}_1, \mathbf{\Delta}_2) \right) \quad (53)$$

where

$$\Gamma_c(\mathbf{\Delta}_1, \mathbf{\Delta}_2) = \mathbf{U} \mathbf{\Delta}_1^T \mathbf{\Delta}_2 \quad (54)$$

Note that since the Riemannian metric of Grassmann manifold  $Gr(k, n)$  is the same as the Riemannian metric for the orthogonal group  $\mathcal{O}(n)$ , the Christoffel function in (54) is equal to (10) and it results in the same form of system of differential equation for the parallel translation.

## References

- Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry*, volume 1. New York, 1963.