SIMPLER FLAG OPTIMIZATION

ZEHUA LAI, LEK-HENG LIM, AND KE YE

ABSTRACT.

CONTENTS

1. INTRODUCTION

Let $d \leq n$ be positive integers and let $(n_1, \ldots, n_d)$ be a sequence integers such that $0 < n_1 < \cdots < n_d < n$. We denote by $\mathrm{Flag}(n_1, \ldots, n_d; \mathbb{R}^n)$ the set of all flags in $\mathbb{R}^n$ of type $(n_1, \ldots, n_d)$:

$$\mathrm{Flag}(n_1, \ldots, n_d; n) := \left\{ \{\mathbb{V}_k\}_{k=1}^d : \mathbb{V}_k \subsetneq \mathbb{V}_{k+1} \subsetneq \mathbb{R}^n, \dim \mathbb{V}_k = n_k, k = 1, \ldots, d-1 \right\}.$$

2. PRELIMINARIES

2.1. differential geometry of Grassmann manifolds. Let $k \leq n$ be positive integer. We denote by $\mathrm{Gr}(k, n)$ the Grassmann manifold of k dimensional subspaces of $\mathbb{R}^n$. According to [4], $\mathrm{Gr}(k, n)$ can be characterized as a submanifold of $\mathrm{O}(n) \cap \mathrm{S}_n$, i.e.,

$$\mathrm{Gr}(k, n) \simeq \{Q \in \mathrm{O}(n) \cap \mathrm{S}_n : \mathrm{tr}(Q) = 2k - n\} \tag{1}$$

$$= \{V \, \mathrm{I}_{k,n-k} \, V^\mathsf{T} : V \in \mathrm{O}(n)\}. \tag{2}$$

Here S_n is the space of $n \times n$ symmetric matrices and $\mathrm{I}_{k,n-k}$ is the diagonal matrix $\begin{bmatrix} \mathrm{I}_k & 0 \\ 0 & -\mathrm{I}_{n-k} \end{bmatrix}$.

Proposition 2.1 (Tangent space I). *Let $Q \in \mathrm{Gr}(k,n)$ with eigendecomposition $Q = V I_{k,n-k} V^{\mathsf{T}}$. The tangent space of $\mathrm{Gr}(k,n)$ at Q is given by*

$$\mathbb{T}_Q \mathrm{Gr}(k,n) = \left\{ X \in \mathbb{R}^{n \times n} : X^{\mathsf{T}} = X,\ XQ + QX = 0,\ \mathrm{tr}(X) = 0 \right\} \tag{3}$$

$$= \left\{ V \begin{bmatrix} 0 & B \\ B^{\mathsf{T}} & 0 \end{bmatrix} V^{\mathsf{T}} \in \mathbb{R}^{n \times n} : B \in \mathbb{R}^{k \times (n-k)} \right\} \tag{4}$$

$$= \left\{ QV \begin{bmatrix} 0 & B \\ -B^{\mathsf{T}} & 0 \end{bmatrix} V^{\mathsf{T}} \in \mathbb{R}^{n \times n} : B \in \mathbb{R}^{k \times (n-k)} \right\}. \tag{5}$$

Proposition 2.2 (Riemannian metric). *Let $Q \in \mathrm{Gr}(k,n)$ with $Q = V I_{k,n-k} V^{\mathsf{T}}$ and*

$$X = V \begin{bmatrix} 0 & B \\ B^{\mathsf{T}} & 0 \end{bmatrix} V^{\mathsf{T}}, \quad Y = V \begin{bmatrix} 0 & C \\ C^{\mathsf{T}} & 0 \end{bmatrix} V^{\mathsf{T}} \in \mathbb{T}_Q \mathrm{Gr}(k,n).$$

Then

$$\langle X, Y \rangle_Q := \mathrm{tr}(XY) = 2\,\mathrm{tr}(B^{\mathsf{T}} C) \tag{6}$$

defines a Riemannian metric. The corresponding Riemannian norm is

$$\|X\|_Q := \sqrt{\langle X, X \rangle_Q} = \|X\|_{\mathsf{F}} = \sqrt{2}\|B\|_{\mathsf{F}}. \tag{7}$$

Theorem 2.3 (Geodesics I). *Let $Q \in \mathrm{Gr}(k,n)$ and $X \in \mathbb{T}_Q \mathrm{Gr}(k,n)$ with*

$$Q = V I_{k,n-k} V^{\mathsf{T}}, \qquad X = V \begin{bmatrix} 0 & B \\ B^{\mathsf{T}} & 0 \end{bmatrix} V^{\mathsf{T}}. \tag{8}$$

The geodesic γ emanating from Q in the direction X is given by

$$\gamma(t) = V \exp\left(\frac{t}{2} \begin{bmatrix} 0 & -B \\ B^{\mathsf{T}} & 0 \end{bmatrix} \right) I_{k,n-k} \exp\left(\frac{t}{2} \begin{bmatrix} 0 & B \\ -B^{\mathsf{T}} & 0 \end{bmatrix} \right) V^{\mathsf{T}}. \tag{9}$$

The differential equation for γ is

$$\gamma(t)^{\mathsf{T}} \ddot{\gamma}(t) - \ddot{\gamma}(t)^{\mathsf{T}} \gamma(t) = 0, \qquad \gamma(0) = Q, \qquad \dot{\gamma}(0) = X. \tag{10}$$

Proposition 2.4 (Parallel transport). *Let $Q \in \mathrm{Gr}(k,n)$ and $X, Y \in \mathbb{T}_Q \mathrm{Gr}(k,n)$ with*

$$Q = V I_{k,n-k} V^{\mathsf{T}}, \qquad X = V \begin{bmatrix} 0 & B \\ B^{\mathsf{T}} & 0 \end{bmatrix} V^{\mathsf{T}}, \qquad Y = V \begin{bmatrix} 0 & C \\ C^{\mathsf{T}} & 0 \end{bmatrix} V^{\mathsf{T}},$$

where $V \in \mathrm{O}(n)$ and $B, C \in \mathbb{R}^{k \times (n-k)}$. Let γ be a geodesic curve emanating from Q in the direction X. Then the parallel transport of Y along γ is

$$Y(t) = V \exp\left(\frac{t}{2} \begin{bmatrix} 0 & -B \\ B^{\mathsf{T}} & 0 \end{bmatrix} \right) \begin{bmatrix} 0 & C \\ C^{\mathsf{T}} & 0 \end{bmatrix} \exp\left(\frac{t}{2} \begin{bmatrix} 0 & B \\ -B^{\mathsf{T}} & 0 \end{bmatrix} \right) V^{\mathsf{T}}. \tag{11}$$

2.2. **some useful functions.** We recall the *Peano–Baker series* associated to a matrix function $\Phi : [a,b] \to \mathbb{R}^{n \times n}$. To define the Peano–Baker series, we first recursively define a sequence $\{M_k(t)\}_{k=0}^{\infty}$ of matrix functions

$$M_0(t) = I_n,$$

$$M_k(t) = I_n + \int_a^t \Phi(s) M_{k-1}(s)\, ds, \quad k \in \mathbb{N}.$$

We have the following:

Theorem 2.5. [2, Section 3, Theorem 1] *The sequence $\{M_k(t)\}_{k=0}^{\infty}$ converges to a matrix function $M(t)$ uniformly on $[a,b]$, which solves the differential equation*

$$\frac{d}{dt} X(t) = \Phi(t) X(t), \quad X(a) = I_n.$$

In particular, given any column vector $u \in \mathbb{R}^n$, $M(t)u$ solves the differential equation

$$\frac{d}{dt}x(t) = \Phi(t)x(t), \quad x(a) = u.$$

The limit matrix function $M(t)$ in Theorem 2.5 is defined to be the Peano–Baker series associated to $\Phi(t)$.

2.3. **vectorization of a matrix.** Let m, n be positive integers and let $A[a_1, \ldots, a_n]$ be a matrix of size $m \times n$ where $a_1, \ldots, a_n \in \mathbb{R}^m$ are column vectors of A. We define the *vectorization* of A to be the column vector

$$\operatorname{vec}(A) := \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^{mn}.$$

We recall that using vectorizations of matrices, we can express the matrix-matrix product in terms of matrix-vector product. Namely, for $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times l}$, we have

$$\operatorname{vec}(AB) = (\mathrm{I}_l \otimes A)\operatorname{vec}(B) = (B^\mathsf{T} \otimes \mathrm{I}_m)\operatorname{vec}(A). \tag{12}$$

Moreover, for any positive integers m, n, there exists a permutation matrix $K^{(m,n)} \in \mathbb{R}^{mn \times mn}$, called *the commutation matrix* such that

$$K^{(m,n)}\operatorname{vec}(A) = \operatorname{vec}(A^\mathsf{T}), \quad A \in \mathbb{R}^{m \times n}. \tag{13}$$

3. Sub-Riemannian geometry of flag manifolds with classical embeddings

According to [5, Proposition 3.2], $\operatorname{Flag}(n_1, \ldots, n_d; n)$ can be embedded into a product of Grassmann manifolds via

$$\iota : \operatorname{Flag}(n_1, \ldots, n_d; n) \hookrightarrow \operatorname{Gr}(n_1, n) \times \operatorname{Gr}(n_2 - n_1, n) \cdots \times \operatorname{Gr}(n_d - n_{d-1}, n)$$

$$(\{\mathbb{V}_k\}_{k=1}^d) \mapsto (\mathbb{W}_1, \mathbb{W}_2, \ldots, \mathbb{W}_d). \tag{14}$$

Here $\mathbb{W}_1 = \mathbb{V}_1$ and $\mathbb{W}_k$ is the orthogonal complement of $\mathbb{V}_{k-1}$ in $\mathbb{V}_k$, $k = 2, \ldots, d$. For simplicity, we denote

$$\boxed{m_1 := n_1, \quad m_{d+1} := n - n_d, \quad m_k := n_k - n_{k-1}, \quad k = 2, \ldots, d} \tag{15}$$

so that ι is an embedding of $\operatorname{Flag}(n_1, \ldots, n_d; n)$ into $\prod_{k=1}^d \operatorname{Gr}(m_k, n)$.

3.1. **an embedding of a flag manifold into a matrix manifold.** By (1), we may also embed each $\operatorname{Gr}(m_k, n)$ into $\mathrm{O}(n)$ and hence we can write $\mathbb{W}_k$ in (14) as $V_k \mathrm{I}_{m_k, n - m_k} V_k^\mathsf{T}$ for some $V_k \in \mathrm{O}(n)$. We denote by τ the induced embedding of $\prod_{k=1}^d \operatorname{Gr}(m_k, n)$ into $\mathrm{O}(n)^d$. In the following, we will explicitly characterize the image $\tau \circ \iota (\operatorname{Flag}(n_1, \ldots, n_d; n))$ in $\mathrm{O}(n)^d$.

Proposition 3.1 (embedding)**.** *The image of the embedding*

$$\epsilon : \operatorname{Flag}(n_1, \ldots, n_d; n) \overset{\iota}{\hookrightarrow} \prod_{k=1}^d \operatorname{Gr}(m_k, n) \overset{\tau}{\hookrightarrow} \prod_{k=1}^d \mathrm{O}(n) \tag{16}$$

is given by

$$\epsilon(\operatorname{Flag}(n_1, \ldots, n_d; n)) = \{(Q_1, \ldots, Q_d) \in \prod_{k=1}^d \mathrm{O}(n) : \operatorname{tr}(Q_k) = 2m_k - n, Q_k^\mathsf{T} = Q_k$$

$$(\mathrm{I}_n + Q_k)(\mathrm{I}_n + Q_{k+1}) = 0, k = 1, \ldots, d\}. \tag{17}$$

In particular, we also have

$$\epsilon(\operatorname{Flag}(n_1, \ldots, n_d; n)) = \{(V \mathrm{J}_1 V^\mathsf{T}, \ldots, V \mathrm{J}_d V^\mathsf{T}) : V \in O(n)\}, \tag{18}$$

where $J_k = \mathrm{diag}(-\mathrm{I}_{m_1}, \cdots, -\mathrm{I}_{m_{k-1}}, \mathrm{I}_{m_k}, -\mathrm{I}_{m_{k+1}}, \cdots, -\mathrm{I}_{m_d}, -\mathrm{I}_{n-n_d})$ *is obtained by permuting diagonal blocks of* $\mathrm{I}_{m_k, n-m_k}$.

Proof. According to (1), we must have $\epsilon(\{\mathbb{R}_k^n\}_{k=1}^d) = (Q_1, \ldots, Q_d) \in (\mathrm{O}(n) \cap \mathrm{S}_n)$ with rank $Q_k = 2m_k - n$. Moreover, since $\mathbb{W}_k$ is perpendicular to $\mathbb{W}_{k+1}$, we must have $P_{\mathbb{W}_k} \circ P_{\mathbb{W}_{k+1}} = 0$ where $P_{\mathbb{U}}$ is the orthogonal projection from $\mathbb{R}^n$ onto a subspace $\mathbb{U}$. Now by [4, Proposition 2.3], we have $P_{\mathbb{W}_k} = \frac{1}{2}(\mathrm{I}_n + Q_k)$ which proves (17). To see (18), we notice that the relation $(\mathrm{I}_n + Q_k)(\mathrm{I}_n + Q_{k+1}) = 0$ implies that $Q_k Q_{k+1} = Q_{k+1} Q_k$ and hence there exists $V_0 \in \mathrm{O}(n)$ diagonalizing Q_k's simultaneously, i.e., $Q_k = V_0 D_k V_0^\mathsf{T}$ where D_k is a diagonal matrix with m_k -1's and $(n - m_k)$ 1's along its diagonal. The restriction $(\mathrm{I}_n + Q_k)(\mathrm{I}_n + Q_{k+1}) = 0$ forces $D_k = \sigma^\mathsf{T} J_k \sigma$ for some permutation matrix σ and hence $V := \sigma V_0$ gives us the desired expression of $\epsilon(\{\mathbb{R}_k^n\}_{k=1}^d)$ in (18). $\qquad\square$

In fact, (18) is a special case of the general fact [3, page 384] that G/P is an adjoint orbit of G if P is a parabolic subgroup of a semi-simple Lie group G. In our case, we have $G = \mathrm{O}(n)$ and $P = \mathrm{O}(n_1) \times \cdots \times \mathrm{O}(n_d)$ so that $G/P \simeq \mathrm{Flag}(n_1, \ldots, n_d; n)$.

Due to Proposition 3.1, in the sequel we abuse the notation by also using $\mathrm{Flag}(n_1, \ldots, n_d; n)$ to denote $\epsilon(\mathrm{Flag}(n_1, \ldots, n_d; n))$. Accordingly, an element in $\mathrm{Flag}(n_1, \ldots, n_d; n)$ is written as a d-tuple

$$(VJ_1V^\mathsf{T}, \ldots, VJ_dV^\mathsf{T}) = V(J_1, \ldots, J_d)V^\mathsf{T}$$

for some $V \in \mathrm{O}(n)$, where $m_1 = n_1$ and $m_k = n_k - n_{k-1}$ for $k = 2, \ldots, d$.

3.2. tangent space, Riemannian metric and normal space.

We first consider the tangent space of $\mathrm{Flag}(n_1, \ldots, n_d; n)$ at a point $V(J_1, \ldots, J_d)V^\mathsf{T}$. To do this, we take a curve $V(t)$ on $\mathrm{O}(n)$ such that $V(0) = V$. It is clear that $\Lambda := V(0)^\mathsf{T} \dot{V}(0) \in \mathfrak{so}(n)$ and hence the tangent vector determined by the curve $V(t)(J_1, \ldots, J_d)V(t)^\mathsf{T}$ is simply

$$\dot{V}(0)(J_1, \ldots, J_d)V(0)^\mathsf{T} + V(0)(J_1, \ldots, J_d)\dot{V}(0)^\mathsf{T}$$

which can be further written as

$$V(0)\left(\Lambda(J_1, \ldots, J_d) - (J_1, \ldots, J_d)\Lambda\right)V(0)^\mathsf{T}.$$

We partition Λ as $\Lambda = (\Lambda(p, q))_{p,q=1}^{d+1}$ where $\Lambda(p, q)$ is a $m_p \times m_q$ matrix such that $\Lambda(q, p) = -\Lambda(p, q)^\mathsf{T}$. This implies that

$$\Lambda J_k - J_k \Lambda = -2 \begin{bmatrix} 0 & \cdots & 0 & \Lambda(k,1)^\mathsf{T} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \Lambda(k,k-1)^\mathsf{T} & 0 & \cdots & 0 \\ \Lambda(k,1) & \cdots & \Lambda(k,k-1) & 0 & \Lambda(k,k+1) & \cdots & \Lambda(k,d+1) \\ 0 & \cdots & 0 & \Lambda(k,k+1)^\mathsf{T} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \Lambda(k,d+1)^\mathsf{T} & 0 & \cdots & 0 \end{bmatrix}. \quad (19)$$

We notice that there is a natural identification $\prod_{1 \le j < k \le d+1} \mathbb{R}^{m_j \times m_k} \simeq \mathfrak{so}(n)$ and hence we have an injective map

$$\psi : \prod_{1 \le j < k \le d+1} \mathbb{R}^{m_j \times m_k} \simeq \mathfrak{so}(n) \hookrightarrow \prod_{j=1}^d \mathrm{S}_n, \quad \psi((A_{jk})_{1 \le j < k \le d+1}) = \frac{1}{2}(AJ - JA),$$

where $J = (J_1, \ldots, J_d)$ and $A \in \mathfrak{so}(n)$ is the skew-symmetric matrix uniquely determined by $(A_{jk})_{1 \le j < k \le d+1}$. The above calculations can be summarized as the following

Proposition 3.2. *Given a point* $\mathfrak{f} := V(J_1, \ldots, J_d)V^\mathsf{T} \in \mathrm{Flag}(n_1, \ldots, n_d; n)$, *the tangent space of* $\mathrm{Flag}(n_1, \ldots, n_d; n)$ *at* $\mathfrak{f}$ *is*

$$\mathbb{T}_\mathfrak{f} \mathrm{Flag}(n_1, \ldots, n_d; n) = V\left\{\psi((A_{jk})_{1 \le j < k \le d+1}) : A_{jk} \in \mathbb{R}^{m_j \times m_k}, 1 \le j < k \le d+1\right\}V^\mathsf{T}.$$

In other words, $\mathbb{T}_{\mathfrak{f}} \operatorname{Flag}(n_1, \ldots, n_d; n)$ *consists of vectors* $V(X_1, \ldots, X_d)V^{\mathsf{T}} \in \prod_{j=1}^d S_n$ *satisfying*

$$X_k(k,l) = -X_l(k,l), X_k(p,q) = 0, X_k(k,k) = 0, \quad 1 \le k, l \le d, 1 \le p, q \le d+1 \text{ and } p, q, l \ne k. \tag{20}$$

Here for each $1 \le s, t \le d+1$, $X_k(s,t) \in \mathbb{R}^{m_s \times m_t}$ *denotes the* (s,t)-*th block of* $X_k \in S_n$ *when we partition* X_k *with respect to* $n = m_1 + \cdots + m_d + m_{d+1}$.

Due to Proposition 3.2, we are able to parametrize a curve on $\operatorname{Flag}(n_1, \ldots, n_d; n)$ easily.

Corollary 3.3 (curves). *If* $c : (-\epsilon, \epsilon) \to \operatorname{Flag}(n_1, \ldots, n_d; n)$ *is a differentiable curve such that* $c(0) = V(J_1, \ldots, J_d)V^{\mathsf{T}}$, *then there exists a differentiable curve* $\Lambda : (-\epsilon, \epsilon) \to \mathfrak{so}(n)$ *such that* $\Lambda(k,k)(t) \equiv 0, k = 1, \ldots, d+1$ *and*

$$c(t) = V \exp(\Lambda(t))(J_1, \ldots, J_d) \exp(-\Lambda(t))V^{\mathsf{T}},$$

where $\Lambda(t) = (\Lambda(j,k))_{j,k=1}^{d+1,d+1}$ *is the partition of* $\Lambda(t)$ *with respect to* $n = m_1 + \cdots + m_{d+1}$.

As a submanifold of $\prod_{k=1}^d \operatorname{Gr}(m_k, n)$ (or equivalently, $\prod_{k=1}^d \operatorname{O}(n)$), $\operatorname{Flag}(n_1, \ldots, n_d; n)$ is equipped with an induced Riemannian metric:

$$\langle V(X_1, \ldots, X_d)V^{\mathsf{T}}, V(Y_1, \ldots, Y_d)V^{\mathsf{T}} \rangle_{\mathfrak{f}} := \sum_{k=1}^d \operatorname{tr}(X_k Y_k), \tag{21}$$

where $\mathfrak{f} = V(J_1, \ldots, J_d)V^{\mathsf{T}}$ is a point in $\operatorname{Flag}(n_1, \ldots, n_d; n)$ and $V(X_1, \ldots, X_d)V^{\mathsf{T}}, V(Y_1, \ldots, Y_d)V^{\mathsf{T}}$ are tangent vectors of $\operatorname{Flag}(n_1, \ldots, n_d; n)$ at $\mathfrak{f}$. More explicitly, we can write

$$\langle V(X_1, \ldots, X_d)V^{\mathsf{T}}, V(Y_1, \ldots, Y_d)V^{\mathsf{T}} \rangle_{\mathfrak{f}} = 2 \sum_{k=1}^d \sum_{l<k<m} \operatorname{tr}(X_k(l,k)Y_k(k,l) + X_k(m,k)Y_k(k,m)). \tag{22}$$

We remark that summands in the formula (22) are not evenly counted. For example, if $d = 2$, then $\langle V(X_1, X_2)V^{\mathsf{T}}, V(Y_1, Y_2)V^{\mathsf{T}} \rangle_{\mathfrak{f}}$ is

$$2(\operatorname{tr}(X_1(2,1)Y_1(1,2)) + \operatorname{tr}(X_1(3,1)Y_1(1,3)) + \operatorname{tr}(X_2(1,2)Y_2(2,1)) + \operatorname{tr}(X_2(3,2)Y_2(2,3))), \tag{23}$$

in which the coefficient of $\operatorname{tr}(X_1(2,1)Y_1(1,2))$ is 4 since $\operatorname{tr}(X_2(1,2)Y_2(2,1)) = \operatorname{tr}(X_1(2,1)Y_1(1,2))$ while coefficients for the other two summands are both 2.

For each $Q \in \operatorname{O}(n)$, we have $\mathbb{T}_Q \operatorname{O}(n) = Q\mathfrak{so}(n)$ and hence for each $(Q_1, \ldots, Q_d) \in \prod_{k=1}^d \operatorname{O}(n)$, we obtain

$$\mathbb{T}_{(Q_1, \ldots, Q_d)}\left(\prod_{k=1}^d \operatorname{O}(n)\right) = \bigoplus_{k=1}^d Q_k \mathfrak{so}(n).$$

To calculate the normal space of $\operatorname{Flag}(n_1, \ldots, n_d; n)$ in $\prod_{k=1}^d \operatorname{O}(n)$ at $\mathfrak{f} = V(J_1, \ldots, J_d)V^{\mathsf{T}}$, we need to determine $Y_1, \ldots, Y_d \in \mathfrak{so}(n)$ such that $Y := (VJ_1V^{\mathsf{T}}Y_1, \ldots, VJ_dV^{\mathsf{T}}Y_d)$ is perpendicular to $\mathbb{T}_{\mathfrak{f}} \operatorname{Flag}(n_1, \ldots, n_d; n)$, i.e., $\langle X, Y \rangle_{\mathfrak{f},\operatorname{O}(n)} = 0$ for all $X \in \mathbb{T}_{\mathfrak{f}} \operatorname{Flag}(n_1, \ldots, n_d; n)$. Here the inner product $\langle \cdot, \cdot \rangle_{\mathfrak{f},\prod_{k=1}^d \operatorname{O}(n)}$ is the canonical Riemannian metric on $\prod_{k=1}^d \operatorname{O}(n)$ at the point $\mathfrak{f}$, which induces (21). We notice that for $X = V(X_1, \ldots, X_d)V^{\mathsf{T}} \in \mathbb{T}_{\mathfrak{f}} \operatorname{Flag}(n_1, \ldots, n_d; n)$

$$VX_kV^{\mathsf{T}} = (VJ_kV^{\mathsf{T}})VJ_kX_kV^{\mathsf{T}}, \quad k = 1, \ldots, d,$$

which implies that

$$\langle X, Y \rangle_{\mathfrak{f}, \prod_{k=1}^d \mathrm{O}(n)} = \sum_{k=1}^d \mathrm{tr}((VJ_kX_kV^\mathsf{T})^\mathsf{T}Y_k)$$

$$= \sum_{k=1}^d \mathrm{tr}((VX_kJ_kV^\mathsf{T})Y_k)$$

$$= \sum_{k=1}^d \mathrm{tr}((X_kJ_k)V^\mathsf{T}Y_kV)$$

Since $\langle X, Y \rangle_{\mathfrak{f}, \prod_{k=1}^d \mathrm{O}(n)} = 0$ holds for any $X \in \mathbb{T}_{\mathfrak{f}}\mathrm{Flag}(n_1, \ldots, n_d; n)$, we can equivalently write this condition as

$$\sum_{k=1}^d \mathrm{tr}((X_kJ_k)Z_k) = 0, \quad (X_1, \ldots, X_d) \in \mathbb{T}_{\mathfrak{f}_0}\mathrm{Flag}(n_1, \ldots, n_d; n),$$

where $\mathfrak{f}_0 = (J_1, \ldots, J_d) \in \mathrm{Flag}(n_1, \ldots, n_d; n)$ and $Z_k = V^\mathsf{T}Y_kV, k = 1, \ldots, d$. If we fix a pair (k, l) such that $1 \le k \le d, 1 \le l \le d+1, k \ne l$ and set $X_m(p, q) = 0$ for

$$(m, p, q) \notin \{(k, k, l), (k, l, k), (l, k, l), (l, l, k)\},$$

then since $X_kJ_{m_k, n-m_k}$ is skew-symmetric, we have

$$0 = \langle X, Y \rangle_{\mathfrak{f}, \prod_{k=1}^d \mathrm{O}(n)} = \mathrm{tr}(X_k(k,l)Z_k(l,k)) - \mathrm{tr}(X_k(k,l)^\mathsf{T}Z_k(k,l)) - \mathrm{tr}(X_l(k,l)^\mathsf{T}Z_l(k,l)) + \mathrm{tr}(X_l(k,l)Z_l(l,k))$$

$$= \mathrm{tr}(X_k(k,l)(Z_k(l,k)) - Z_l(l,k))) - \mathrm{tr}(X_k(k,l)^\mathsf{T}(-Z_k(k,l)) + Z_l(k,l)))$$

$$= \mathrm{tr}(X_k(k,l)(Z_k(l,k)) - Z_l(l,k))) + \mathrm{tr}(X_k(k,l)^\mathsf{T}(Z_k(k,l)) - Z_l(k,l))$$

$$= 2\,\mathrm{tr}(X_k(k,l)(Z_k(l,k) - Z_l(l,k))).$$

Therefore, we may derive the following characterization of $\mathbb{N}_{\mathfrak{f}}\mathrm{Flag}(n_1, \ldots, n_d; n)$:

Proposition 3.4. *At a point* $\mathfrak{f} := V(J_1, \ldots, J_d)V^\mathsf{T} \in \mathrm{Flag}(n_1, \ldots, n_d; \mathbb{R}^n)$, *the normal space* $\mathbb{N}_{\mathfrak{f}}\mathrm{Flag}(n_1, \ldots, n_d; n)$ *consists of vectors*

$$(VJ_1Z_1V^\mathsf{T}, \ldots, VJ_dZ_dV^\mathsf{T})$$

where $Z_1, \ldots, Z_d \in \mathfrak{so}(n)$ *satisfy the relations*

- $Z_k(k, l) - Z_l(k, l) = 0$ *for all* $1 \le k \ne l \le d$.
- $Z_k(k, d+1) = 0, Z_k(d+1, k) = 0$ *for all* $1 \le k \le d$.

In particular, we have a decomposition

$$\mathbb{N}_{\mathfrak{f}}\mathrm{Flag}(n_1, \ldots, n_d; n) = N_{\mathfrak{f}}\left(\prod_{k=1}^d \mathrm{Gr}(m_k, n)\right) \bigoplus N_{\mathfrak{f}}^0 \tag{24}$$

where $\mathbb{N}_{\mathfrak{f}}\left(\prod_{k=1}^d \mathrm{Gr}(m_k, n)\right) := \prod_{k=1}^d \mathbb{N}_{VJ_{m_k, n-m_k}V^\mathsf{T}}\mathrm{Gr}(m_k, n)$ *and*

$$\mathbb{N}_{\mathfrak{f}}^0 := \{(VJ_{m_1, n-m_1}Z_1V^\mathsf{T}, \ldots, VJ_{m_d, n-m_d}Z_dV^\mathsf{T}) : Z_k \in \mathfrak{so}(n), Z_k(k, l) - Z_l(k, l) = 0,$$
$$Z_k(k, k) = 0, Z_k(p, q) = 0, Z_k(k, d+1) = 0, Z_k(d+1, k) = 0, 1 \le k, l \le d, 1 \le p, q \le d+1, p, q \ne k\}. \tag{25}$$

We recall that $\mathrm{Flag}(n_1, \ldots, n_d; n)$ can also be embedded into $\prod_{k=1}^d \mathrm{Gr}(n_k, n)$ as a Riemannian submanifold. Hence we may also characterize the normal space of $\mathrm{Flag}(n_1, \ldots, n_d; n)$ with respect to this embedding.

Corollary 3.5. *The normal space of* $\mathrm{Flag}(n_1, \ldots, n_d; n)$ *in* $\prod_{k=1}^d \mathrm{Gr}(m_k, n)$ *at a point* $\mathfrak{f}$ *is* $\mathbb{N}_{\mathfrak{f}}^0$.

Proposition 3.6 (Projections). *Projections from $\mathbb{T}_{\mathfrak{f}}\left(\prod_{k=1}^{d} O(n)\right)$ onto $\mathbb{T}_{\mathfrak{f}}\operatorname{Flag}(n_1,\ldots,n_d;n)$ and $\mathbb{N}_{\mathfrak{f}}\operatorname{Flag}(n_1,\ldots,n_d;n)$ are respectively given by*

$$\operatorname{proj}_{\mathfrak{f}}^{\mathbb{T}} : \mathbb{T}_{\mathfrak{f}}\left(\prod_{k=1}^{d} O(n)\right) \to \mathbb{T}_{\mathfrak{f}}\operatorname{Flag}(n_1,\ldots,n_d;n)$$

$$V(J_1\Lambda_1,\ldots,J_d\Lambda_d)V^{\mathsf{T}} \mapsto V(X_1,\ldots,X_d)V^{\mathsf{T}}. \tag{26}$$

and

$$\operatorname{proj}_{\mathfrak{f}}^{\mathbb{N}} : \mathbb{T}_{\mathfrak{f}}\left(\prod_{k=1}^{d} O(n)\right) \to \mathbb{N}_{\mathfrak{f}}\operatorname{Flag}(n_1,\ldots,n_d;n)$$

$$V(J_1\Lambda_1,\ldots,J_d\Lambda_d)V^{\mathsf{T}} \mapsto V(Z_1,\ldots,Z_d)V^{\mathsf{T}} \tag{27}$$

where for each $k = 1,\ldots,d$, $X_k \in \mathrm{S}_n$ (resp. $Z_k \in \mathbb{R}^{n\times n}$) is partitioned as $(X_k(p,q))_{p,q=1}^{d+1}$ (resp. $(Z_k(p,q))_{p,q=1}^{d+1}$) with respect to $n = m_1 + \cdots + m_{d+1}$ and

$$X_k(p,q) = \begin{cases} \frac{1}{2}(\Lambda_k(k,q) - \Lambda_q(k,q)), & \text{if } p = k \neq q \leq d \\ \Lambda_k(k,d+1), & \text{if } p = k, q = d+1 \\ -\frac{1}{2}(\Lambda_k(p,k) - \Lambda_p(p,k)), & \text{if } q = k \neq p \leq d \\ -\Lambda_k(d+1,q), & \text{if } q = k, p = d+1 \\ 0, & \text{otherwise.} \end{cases}$$

$$Z_k(p,q) = \begin{cases} \frac{1}{2}(\Lambda_k(k,q) + \Lambda_q(k,q)), & \text{if } p = k \neq q \leq d \\ 0, & \text{if } p = k, q = d+1 \\ -\frac{1}{2}(\Lambda_k(p,k) + \Lambda_p(p,k)), & \text{if } q = k \neq p \leq d \\ 0, & \text{if } q = k, p = d+1 \\ \Lambda_k(p,q), & \text{otherwise.} \end{cases}$$

Before we proceed, we work out the case for $d = 2$ to exhibit our calculations above. In this case, our flag manifold is $\operatorname{Flag}(n_1,n_2;n)$ and hence $m_1 = n_1, m_2 = n_2 - n_1, m_3 = n - n_2$. A point $\mathfrak{f}$ in $\operatorname{Flag}(n_1,n_2;n)$ is written as

$$V(J_1,J_2)V^{\mathsf{T}} = V\left(\begin{bmatrix} \mathrm{I}_{m_1} & 0 & 0 \\ 0 & -\mathrm{I}_{m_2} & 0 \\ 0 & 0 & -\mathrm{I}_{m_3} \end{bmatrix}, \begin{bmatrix} -\mathrm{I}_{m_1} & 0 & 0 \\ 0 & \mathrm{I}_{m_2} & 0 \\ 0 & 0 & -\mathrm{I}_{m_3} \end{bmatrix}\right)V^{\mathsf{T}}, \quad V \in O(n).$$

A tangent vector of $\operatorname{Flag}(n_1,n_2;n)$ at $\mathfrak{f}$ is of the form

$$V\left(\begin{bmatrix} 0 & A & B \\ A^{\mathsf{T}} & 0 & 0 \\ B^{\mathsf{T}} & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -A & 0 \\ -A^{\mathsf{T}} & 0 & C \\ 0 & C^{\mathsf{T}} & 0 \end{bmatrix}\right)V^{\mathsf{T}}, \quad A \in \mathbb{R}^{m_1\times m_2}, B \in \mathbb{R}^{m_1\times m_3}, C \in \mathbb{R}^{m_2\times m_3}.$$

The normal space of $\operatorname{Flag}(n_1,n_2;n)$ as a submanifold of $O(n) \times O(n)$ at $\mathfrak{f}$ consists of vectors

$$V\left(\begin{bmatrix} X & Y & 0 \\ Y^{\mathsf{T}} & Z & W \\ 0 & -W^{\mathsf{T}} & U \end{bmatrix}, \begin{bmatrix} R & Y & S \\ Y^{\mathsf{T}} & T & 0 \\ -S^{\mathsf{T}} & 0 & K \end{bmatrix}\right)V^{\mathsf{T}},$$

where $X, R \in \mathfrak{so}(m_1)$, $Z, T \in \mathfrak{so}(m_2)$, $U, K \in \mathfrak{so}(m_3)$, $Y \in \mathbb{R}^{m_1\times m_2}, W \in \mathbb{R}^{m_2\times m_3}, S \in \mathbb{R}^{m_1\times m_3}$.

A tangent vector ξ of $O(n) \times O(n)$ at $\mathfrak{f}$ can be written as

$$\xi := V\left(\begin{bmatrix} A & B & C \\ B^{\mathsf{T}} & D & E \\ C^{\mathsf{T}} & -E^{\mathsf{T}} & F \end{bmatrix}, \begin{bmatrix} X & Y & Z \\ Y^{\mathsf{T}} & W & S \\ -Z^{\mathsf{T}} & S^{\mathsf{T}} & T \end{bmatrix}\right)V^{\mathsf{T}},$$

where $A, X \in \mathfrak{so}(m_1), D, W \in \mathfrak{so}(m_2), F, T \in \mathfrak{so}(m_3), B, Y \in \mathbb{R}^{m_1 \times m_2}, C, Z \in \mathbb{R}^{m_1 \times m_3}, E, S \in \mathbb{R}^{m_2 \times m_3}$. The projection of ξ onto $T_{\mathfrak{f}} \operatorname{Flag}(n_1, n_2; n)$ is

$$\operatorname{proj}_{\mathfrak{f}}^{\mathbb{T}}(\xi) = V \left(\begin{bmatrix} 0 & \frac{B-Y}{2} & C \\ \frac{B^\mathsf{T} - Y^\mathsf{T}}{2} & 0 & 0 \\ C^\mathsf{T} & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -\frac{B-Y}{2} & 0 \\ -\frac{B^\mathsf{T} - Y^\mathsf{T}}{2} & 0 & S \\ 0 & S^\mathsf{T} & 0 \end{bmatrix} \right) V^\mathsf{T}$$

and its projection onto $N_{\mathfrak{f}} \operatorname{Flag}(n_1, n_2; n)$ is

$$\operatorname{proj}_{\mathfrak{f}}^{\mathbb{N}}(\xi) = V \left(\begin{bmatrix} A & \frac{B+Y}{2} & 0 \\ \frac{B^\mathsf{T} + Y^\mathsf{T}}{2} & D & E \\ 0 & -E^\mathsf{T} & F \end{bmatrix}, \begin{bmatrix} X & \frac{B+Y}{2} & Z \\ \frac{B^\mathsf{T} + Y^\mathsf{T}}{2} & W & 0 \\ -Z^\mathsf{T} & 0 & T \end{bmatrix} \right) V^\mathsf{T}$$

The normal space $\mathbb{N}_f^0$ of $\operatorname{Flag}(n_1, n_2; n)$ as a submanifold of $\operatorname{Gr}(m_1, n) \times \operatorname{Gr}(m_2, n)$ at $\mathfrak{f}$ consists of vectors

$$V \left(\begin{bmatrix} 0 & Y & 0 \\ Y^\mathsf{T} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & Y & 0 \\ Y^\mathsf{T} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) V^\mathsf{T}, \quad Y \in \mathbb{R}^{m_1 \times m_2}.$$

We also recall that the tangent space $\mathbb{T}_{\mathfrak{f}}(\operatorname{Gr}(m_1, n) \times \operatorname{Gr}(m_2, n))$ consists of vectors

$$V \left(\begin{bmatrix} 0 & A & B \\ A^\mathsf{T} & 0 & 0 \\ B^\mathsf{T} & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & D & 0 \\ D^\mathsf{T} & 0 & C \\ 0 & C^\mathsf{T} & 0 \end{bmatrix} \right) V^\mathsf{T}, \quad V \in \operatorname{O}(n), A, D \in \mathbb{R}^{m_1 \times m_2}, B \in \mathbb{R}^{m_1 \times m_3}, C \in \mathbb{R}^{m_2 \times m_3}.$$

The following identities can be directly verified by the above computations.

$$\mathbb{T}_{\mathfrak{f}}(\operatorname{O}(n) \times \operatorname{O}(n)) = \mathbb{T}_{\mathfrak{f}} \operatorname{Flag}(n_1, n_2; n) \bigoplus \mathbb{N}_{\mathfrak{f}} \operatorname{Flag}(n_1, n_2; n),$$

$$\mathbb{T}_{\mathfrak{f}}(\operatorname{Gr}(m_1, n) \times \operatorname{Gr}(m_2, n)) = \mathbb{T}_{\mathfrak{f}} \operatorname{Flag}(n_1, n_2; n) \bigoplus \mathbb{N}_{\mathfrak{f}}^0.$$

3.3. **geodesics.** Recall that we may parametrize a curve $c(t)$ on $\operatorname{Flag}(n_1, \ldots, n_d; n)$ as

$$c(t) = V(t)(J_1, \ldots, J_d)V^\mathsf{T}(t),$$

where $V(t)$ is a curve in $\operatorname{O}(n)$. By differentiating the equation $V(t)^\mathsf{T} V(t) = I_n$, we obtain

$$\dot{V}(t)^\mathsf{T} V(t) + V(t)^\mathsf{T} \dot{V}(t) = 0,$$

from which we may write $\dot{V}(t)$ as

$$\dot{V}(t) = V(t)\Lambda(t),$$

for some $\Lambda(t) \in \mathfrak{so}(n)$. According to Proposition 3.2, we may further partition $\Lambda(t)$ as

$$\Lambda(t) = (\Lambda_{jk})_{j,k=1}^{d+1,d+1}$$

with respect to $n = m_1 + \cdots + m_{d+1}$ and $\Lambda_{kk}(t) \equiv 0, k = 1, \ldots, d+1$. Hence the second derivative of $c(t)$ is

$$\ddot{c}(t) = V(t) (\Delta_1(t), \ldots, \Delta_d(t)) V(t)^\mathsf{T}$$

where

$$\Delta_k(t) = (\dot{\Lambda}(t)J_k - J_k \dot{\Lambda}(t)) + (\Lambda^2(t)J_k + J_k \Lambda^2(t)) + (-2\Lambda(t)J_k\Lambda(t)), \quad k = 1, \ldots, d. \qquad (28)$$

We may rewrite $\ddot{c}(t)$ as

$$\ddot{c}(t) = T_1(t) + T_2(t) - 2T_3(t)$$

where $T_j(t)$ is the j-summand of $V(t)(\Delta_1(t), \ldots, \Delta_d(t))V(t)^\mathsf{T}$ with respect to the decomposition of $\Delta_k(t)$ given in (28). More precisely,

$$T_1(t) = V(t)(\dot\Lambda(t)J_1 - J_1\dot\Lambda(t), \ldots, \dot\Lambda(t)J_d - J_d\dot\Lambda(t))V(t)^\mathsf{T}, \tag{29}$$

$$T_2(t) = V(t)(\Lambda^2(t)J_1 + J_1\Lambda^2(t), \ldots, \Lambda^2(t)J_d + J_d\Lambda^2(t))V(t)^\mathsf{T}, \tag{30}$$

$$T_3(t) = V(t)(\Lambda(t)J_1\Lambda(t), \ldots, \Lambda(t)J_d\Lambda(t))V(t)^\mathsf{T}. \tag{31}$$

We recall that the geodesic equation on $\mathrm{Flag}(n_1, \ldots, n_d; n)$ is given by

$$\mathrm{proj}_{c(t)}^{\mathbb{T}}(\ddot c(t)) = 0.$$

Therefore, to determine the geodesic equation explicitly, we need to compute the projections of $T_1(t), T_2(t), T_3(t)$ to $T_{c(t)}\mathrm{Flag}(n_1, \ldots, n_d; n)$ respectively. From Proposition 3.2, $T_1(t)$ already lies in the tangent space $T_{c(t)}\mathrm{Flag}(n_1, \ldots, n_d; n)$. Hence it is sufficient to determine the projections of $T_2(t)$ and $T_3(t)$.

Lemma 3.7. *Let $c(t), T_2(t)$ be as above. The projection of $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_2(t))$ is zero.*

Proof. We first compute $\Lambda^2(t)J_k + J_k\Lambda^2(t)$ for each $k = 1, \ldots, d$. To do this, we partition $\Lambda^2(t)$ (resp. J_k) as $(\Gamma_{p,q}(t))$ (resp. $(J_k(p, q))$) with respect to the partition $n = m_1 + \cdots + m_{d+1}$ and we recall that

$$J_k(p, q) = \begin{cases} (2\delta_{pk} - 1)\,\mathrm{I}_{m_p}, & \text{if } q = p, \\ 0, & \text{otherwise.} \end{cases}$$

Here δ_{pk} is the Kronecker delta function. Since $\Lambda(t)$ is skew-symmetric, $\Lambda^2(t)$ is symmetric. We have $\Gamma_{q,p} = \Gamma_{p,q}^\mathsf{T}$. Now the (p, q)-th block of $\Lambda^2(t)J_k$ is

$$\sum_{l=1}^{m+1} \Gamma_{p,l}J_k(l, q) = \Gamma_{p,q}J_k(q, q) = (2\delta_{qk} - 1)\Gamma_{p,q}$$

and the (p, q)-th block of $J_k\Lambda^2(t) = (\Lambda^2(t)J_k)^\mathsf{T}$ is $(2\delta_{pk} - 1)\Gamma_{p,q}$. This implies that the (p, q)-th block of $\Lambda^2(t)J_k + J_k\Lambda^2(t)$ is

$$(2\delta_{qk} - 1)\Gamma_{p,q} + (2\delta_{pk} - 1)\Gamma_{p,q} = (-2)(1 - \delta_{pk} - \delta_{qk})\Gamma_{p,q}.$$

In particular, if either $q \neq p = k$ or $p \neq q = k$, we obtain that the (p, q)-th block of $\Lambda^2(t)J_k + J_k\Lambda^2(t)$ is zero and this implies that $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_2(t)) = 0$. $\square$

Lemma 3.8. *Let $c(t), T_3(t)$ be as before. The projection $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_3(t))$ is*

$$V(t)(X_1, \ldots, X_d)V(t)^\mathsf{T}$$

where for each $1 \leq k \leq d$, X_k is a symmetric matrix whose (p, q)-th block vanishes for any (p, q) except $(k, d+1)$ and $(d+1, k)$. Moreover if we partition $\Lambda(t)$ as $\Lambda(t) = \begin{bmatrix} \Lambda_0(t) & \Lambda_1(t) \\ -\Lambda_1(t)^\mathsf{T} & 0 \end{bmatrix}$ where $\Lambda_0(t) \in \mathfrak{so}(n - m_{d+1})$ and $\Lambda_1(t) \in \mathbb{R}^{(n-m_{d+1})\times m_1}$ we have

$$\begin{bmatrix} X_1(1, d+1) \\ \vdots \\ X_d(d, d+1) \end{bmatrix} = -\Lambda_0(t)\Lambda_1(t).$$

Proof. It is sufficient to compute $X_k := \Lambda(t)J_k\Lambda(t)$ for each $k = 1, \ldots, d$. We again partition $\Lambda(t)$ as $(\Lambda(p,q)(t))_{p,q=1}^d$ with respect to $n = m_1 + \cdots + m_{d+1}$. The (p,q)-th block of $\Lambda(t)J_k\Lambda(t)$ is

$$\sum_{l,s=1}^{d+1} \Lambda(p,l)(t)J_k(l,s)\Lambda(s,q)(t) = \sum_{l=1}^{d+1} \Lambda(p,l)(t)J_k(l,l)\Lambda(l,q)(t)$$

$$= \sum_{l=1}^{d+1}(2\delta_{kl} - 1)\Lambda(p,l)(t)\Lambda(l,q)(t). \tag{32}$$

In particular, for $1 \leq q \neq k \leq d$, the (k,q)-th block of $\Lambda(t)J_k\Lambda(t)$ is

$$\sum_{l=1}^{d+1}(2\delta_{kl} - 1)\Lambda(k,l)(t)\Lambda(l,q)(t),$$

while the (k,q)-th block of $\Lambda(t)J_q\Lambda(t)$ is

$$\sum_{l=1}^{d+1}(2\delta_{ql} - 1)\Lambda(k,l)(t)\Lambda(l,q)(t).$$

Using Proposition 3.6, we may conclude that the (k,q)-th block of X_k is zero if $1 \leq k,q \leq d$.

If we take $q = d+1$ and $p = k$ in (32), then the $(k, d+1)$-th block of X_k is

$$X_k(k, d+1) = - \sum_{1 \leq l \neq k \leq d} \Lambda(k,l)(t)\Lambda(l, d+1)(t).$$

We observe that $X_k(k, d+1)$ is the k-th block of the product

$$-\begin{bmatrix} 0 & \Lambda(1,2)(t) & \cdots & \Lambda(1,d-1)(t) & \Lambda(1,d)(t) \\ \Lambda(2,1)(t) & 0 & \cdots & \Lambda(2,d-1)(t) & \Lambda(2,d)(t) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Lambda(d-1,1)(t) & \Lambda(d-1,2)(t) & \cdots & 0 & \Lambda(d-1,d)(t) \\ \Lambda(d,1)(t) & \Lambda(d,2)(t) & \cdots & \Lambda(d,d-1)(t) & 0 \end{bmatrix}\begin{bmatrix} \Lambda(1,d+1)(t) \\ \Lambda(2,d+1)(t) \\ \vdots \\ \Lambda(d-1,d+1)(t) \\ \Lambda(d,d+1)(t) \end{bmatrix},$$

which can be written in a compact form $-\Lambda_0(t)\Lambda_1(t)$. $\square$

By assembling Lemmas 3.7 and 3.8, we can easily derive the geodesic equation on a flag manifold, from which we can even obtain an explicit formula for the geodesic curve. In fact, we have the following:

Proposition 3.9 (geodesics). *Let $c(t)$ be a curve on* $\mathrm{Flag}(n_1, \ldots, n_d; n)$. *We parametrize $c(t)$ as*

$$c(t) = V(t)(J_1, \ldots, J_d)V(t)^\mathsf{T},$$

where $V(t)$ is a curve in $\mathrm{O}(n)$. *We have the following:*

(1) *There exists a unique $\Lambda(t) \in \mathfrak{so}(n)$ such that $\dot{V}(t) = V(t)\Lambda(t)$.*
(2) *If we partition $\Lambda(t)$ as $\Lambda(t) = (\Lambda(p,q)(t))_{p,q=1}^{d+1,d+1} \in \mathfrak{so}(n)$ with respect to $n = m_1 + \cdots + m_{d+1}$, then $\Lambda(p,p)(t) \equiv 0, p = 1, \ldots, d+1$.*
(3) *$c(t)$ is a geodesic curve if and only if*

$$\dot{\Lambda}_0(t) = 0, \quad \dot{\Lambda}_1(t) = \Lambda_0(t)\Lambda_1(t). \tag{33}$$

where $\Lambda_0(t) := (\Lambda(p,q)(t))_{p,q=1}^{d,d}$ and $\Lambda_1(t) := (\Lambda(d+1,q)(t))_{q=1}^d$.

(4) The solution to (33) *is*

$$\Lambda_0(t) = \Lambda_0(0), \quad \Lambda_1(t) = \exp(t\Lambda_0(0))\Lambda_1(0).$$

Hence a geodesic curve $c(t)$ is

$$c(t) = V(t)(J_1, \ldots, J_d)V^\mathsf{T}(t),$$

where $V(t)$ is a curve in $\mathrm{O}(n)$ written as

$$V(t) = V(0) \exp\left(t \begin{bmatrix} 2X_0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix} \right) \begin{bmatrix} \exp(-tX_0) & 0 \\ 0 & \mathrm{I}_{m_{d+1}} \end{bmatrix} \tag{34}$$

for some $X_0 \in \mathfrak{so}(n-m_{d+1})$ satisfying $X_0(k,k) = 0, k = 1, \ldots, d$ and $X_1 \in \mathbb{R}^{(n-m_{d+1})\times m_{d+1}}$.

Proof. (1)–(3) and the first half of (4) are obvious from our earlier discussions, hence it is only left to prove the second part of (4). To that end, we notice that $V(t)$ must satisfy the equation

$$\dot{V}(t) = V(t) \begin{bmatrix} X_0 & \exp(tX_0)X_1 \\ -X_1^\mathsf{T}\exp(-tX_0) & 0 \end{bmatrix} \tag{35}$$

and

$$\begin{bmatrix} X_0 & \exp(tX_0)X_1 \\ -X_1^\mathsf{T}\exp(-tX_0) & 0 \end{bmatrix} = \begin{bmatrix} \exp(tX_0) & 0 \\ 0 & \mathrm{I}_{m_{d+1}} \end{bmatrix} \begin{bmatrix} X_0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \exp(-tX_0) & 0 \\ 0 & \mathrm{I}_{m_{d+1}} \end{bmatrix}.$$

If we set $W(t) = V(t) \begin{bmatrix} \exp(tX_0) & 0 \\ 0 & \mathrm{I}_{m_{d+1}} \end{bmatrix}$, then (35) becomes

$$\dot{W}(t) = W(t) \begin{bmatrix} 2X_0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix}$$

whose solution is simply $W(t) = W(0) \exp\left(t \begin{bmatrix} 2X_0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix} \right) = V(0) \exp\left(t \begin{bmatrix} 2X_0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix} \right)$. Hence we obtain that

$$V(t) = V(0) \exp\left(t \begin{bmatrix} 2X_0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix} \right) \begin{bmatrix} \exp(-tX_0) & 0 \\ 0 & \mathrm{I}_{m_{d+1}} \end{bmatrix}.$$

$\square$

We remark that if $d = 1$, then $X_0 = 0$ in (34) and a geodesic curve on $\mathrm{Gr}(n_1, n)$ passing through VJ_1V^T is

$$c(t) = V \exp\left(t \begin{bmatrix} 0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix} \right) \mathrm{I}_{n_1, n-n_1} \left(-t \begin{bmatrix} 0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix} \right) V^\mathsf{T},$$

which coincides with the formula derived in [4].

We again work out the case $d = 2$ to illustrate the proof of Proposition 3.9. To this end, we write

$$\Lambda(t) = \begin{bmatrix} 0 & A(t) & B(t) \\ -A^\mathsf{T}(t) & 0 & C(t) \\ -B^\mathsf{T}(t) & -C^\mathsf{T}(t) & 0 \end{bmatrix}, \quad A(t) \in \mathbb{R}^{m_1 \times m_2}, B(t) \in \mathbb{R}^{m_1 \times m_3}, C(t) \in \mathbb{R}^{m_2 \times m_3}$$

and suppose that the curve

$$c(t) = V(t)(J_1, J_2)V(t)^\mathsf{T}, \quad \dot{V}(t) = V(t)\Lambda(t), \quad V(t) \in \mathrm{O}(n)$$

is a curve passing through (J_1, J_2) with the direction

$$(\Lambda(0)J_1 - J_1\Lambda(0), \Lambda(0)J_2 - J_2\Lambda(0)) = -2 \left(\begin{bmatrix} 0 & A(0) & B(0) \\ A(0)^\mathsf{T} & 0 & 0 \\ B^\mathsf{T}(0) & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -A(0) & 0 \\ -A^\mathsf{T}(0) & 0 & C(0) \\ 0 & C^\mathsf{T}(0) & 0 \end{bmatrix} \right).$$

We write $\ddot{c}(t) = V(t) (\Delta_1(t), \Delta_2(t)) V(t)^{\mathsf{T}}$ where

$$\Delta_k(t) = (\dot{\Lambda}(t) J_k - J_k \dot{\Lambda}(t)) + (\Lambda^2(t) J_k + J_k \Lambda^2(t)) + (-2\Lambda(t) J_k \Lambda(t)) .$$

It is sufficient to compute the projection of $\Lambda(t) J_k \Lambda(t)$ onto $T_{c(t)} \operatorname{Flag}(n_1, n_2; n)$, which is

$$\Lambda(t) J_1 \Lambda(t) = \left(\begin{bmatrix} * & B(t)C(t)^{\mathsf{T}} & -A(t)C(t) \\ C(t)B(t)^{\mathsf{T}} & * & * \\ -C(t)^{\mathsf{T}}A(t)^{\mathsf{T}} & * & * \end{bmatrix}, \begin{bmatrix} * & B(t)C(t)^{\mathsf{T}} & * \\ C(t)B(t)^{\mathsf{T}} & * & A(t)^{\mathsf{T}}B(t) \\ * & B(t)^{\mathsf{T}}A(t) & * \end{bmatrix} \right),$$

where $*$ denotes those irrelevant blocks. Eventually, we obtain

$$\operatorname{proj}_{c(t)}^{\mathbb{T}}(\dot{c}(t)) = -2 \left(\begin{bmatrix} 0 & \dot{A}(t) & \dot{B}(t) - A(t)C(t) \\ \dot{A}(t)^{\mathsf{T}} & 0 & 0 \\ \dot{B}(t)^{\mathsf{T}} - C(t)^{\mathsf{T}}A(t) & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -\dot{A}(t) & 0 \\ -\dot{A}^{\mathsf{T}}(t) & 0 & \dot{C}(t) + A(t)^{\mathsf{T}}B(t) \\ 0 & \dot{C}(t)^{\mathsf{T}} + B(t)^{\mathsf{T}}A(t) & 0 \end{bmatrix} \right).$$

Hence the geodesic equation for $\operatorname{Flag}(n_1, n_2; n)$ is

$$\dot{A}(t) = 0, \quad \dot{B}(t) - A(t)C(t) = 0, \quad \dot{C}(t) + A(t)^{\mathsf{T}}B(t) = 0,$$

which can be rewritten in a more compact form:

$$\dot{A} = 0, \quad \begin{bmatrix} \dot{B}(t) \\ \dot{C}(t) \end{bmatrix} = \begin{bmatrix} 0 & A(t) \\ -A^{\mathsf{T}}(t) & 0 \end{bmatrix} \begin{bmatrix} B(t) \\ C(t) \end{bmatrix}. \tag{36}$$

The solution to (36) is

$$A(t) = A(0), \quad \begin{bmatrix} B(t) \\ C(t) \end{bmatrix} = \exp \left(t \begin{bmatrix} 0 & A(0) \\ -A^{\mathsf{T}}(0) & 0 \end{bmatrix} \right) \begin{bmatrix} B(0) \\ C(0) \end{bmatrix}.$$

4. Sub-Riemannian geometry of flag manifolds with modified embeddings

In this section, we discuss the embedded geometry of flag manifolds with respect to a modified version of the embedding (14). Namely, we define

$$\tilde{\iota} : \operatorname{Flag}(n_1, \dots, n_d; n) \hookrightarrow \operatorname{Gr}(n_1, n) \times \operatorname{Gr}(n_2 - n_1, n) \cdots \times \operatorname{Gr}(n_d - n_{d-1}, n) \times \operatorname{Gr}(n - n_d, n)$$

$$(\{\mathbb{V}_k\}_{k=1}^d) \mapsto (\mathbb{W}_1, \mathbb{W}_2, \dots, \mathbb{W}_d, \mathbb{W}_{d+1}), \tag{37}$$

Here $\mathbb{W}_k$ is the orthogonal complement of $\mathbb{V}_{k-1}$ in $\mathbb{V}_k$ for $2 \le k \le d$, $\mathbb{W}_1 = \mathbb{R}_1^n$ and $\mathbb{W}_{d+1}$ is the orthogonal complement of $\mathbb{V}_d$ in $\mathbb{R}^n$. We observe that

$$\tilde{\iota}(\{\mathbb{V}_k\}_{k=1}^d) = (\iota(\{\mathbb{V}_k\}_{k=1}^d), \mathbb{W}_{d+1}).$$

In other words, $\tilde{\iota}$ is simply an extension of ι by tautologically adding the orthogonal complement of $\mathbb{V}_d$. Since ι is already an embedding, we may easily conclude that $\tilde{\iota}$ is also an embedding. Adopting the convention (15), $\tilde{\iota}$ embeds $\operatorname{Flag}(n_1, \dots, n_d; n)$ into $\prod_{j=1}^{d+1} \operatorname{Gr}(m_j, n)$. Moreover, by Proposition 3.1 we have the following:

Proposition 4.1 (embedding). *The image of the embedding*

$$\tilde{\epsilon} : \operatorname{Flag}(n_1, \dots, n_d; n) \overset{\tilde{\iota}}{\hookrightarrow} \prod_{j=1}^{d+1} \operatorname{Gr}(m_j, n) \overset{\tilde{\tau}}{\hookrightarrow} \operatorname{O}(n)^{d+1} \tag{38}$$

is given by

$$\tilde{\epsilon}(\operatorname{Flag}(n_1, \dots, n_d; n)) = \{(Q_1, \dots, Q_{d+1}) \in \prod_{j=1}^{d+1} \operatorname{O}(n) : \operatorname{tr}(Q_j) = 2m_j - n, Q_j^{\mathsf{T}} = Q_j$$

$$(\mathrm{I}_n + Q_j)(\mathrm{I}_n + Q_{j+1}) = 0, j = 1, \dots, d+1\}. \tag{39}$$

In particular, we also have

$$\tilde{\epsilon}(\operatorname{Flag}(n_1, \dots, n_d; n)) = \{V(J_1, \dots, J_{d+1})V^{\mathsf{T}} : V \in O(n)\}, \tag{40}$$

where $J_k = \operatorname{diag}(-I_{m_1}, \cdots, -I_{m_{k-1}}, I_{m_k}, -I_{m_{k+1}}, \cdots, -I_{m_{d+1}})$ *is obtained by permuting diagonal blocks of* $I_{m_k, n-m_k}, k = 1, \ldots, d+1$ *and*

$$V(J_1, \ldots, J_{d+1})V^\mathsf{T} := (V\, J_1\, V^\mathsf{T}, \ldots, V\, J_{d+1}\, V^\mathsf{T}).$$

Similarly to Proposition 3.2 and Corollary 3.3, we also have:

Proposition 4.2. *Given a point* $\tilde{\mathsf{f}} := V(J_1, \ldots, J_{d+1})V^\mathsf{T}$, *the tangent space* $\mathbb{T}_{\tilde{\mathsf{f}}}\operatorname{Flag}(n_1, \ldots, n_d; n)$ *consists of vectors* $V(X_1, \ldots, X_{d+1})V^\mathsf{T} \in \prod_{j=1}^{d+1} S_n$ *satisfying*

$$X_k(k, l) = -X_l(k, l), X_k(p, q) = 0, X_k(k, k) = 0, \quad 1 \le k, l, p, q \le d+1 \text{ and } p, q, l \ne k. \tag{41}$$

Here $X_k(s, t) \in \mathbb{R}^{m_s \times m_t}$ *is the* (s, t)-*th block of* $X_k \in S_n$ *when we partition* X_k *with respect to* $n = \sum_{j=1}^{d+1} m_j$. *Moreover, a curve* $c(t)$ *passing through* $c(0) = V(J_1, \ldots, J_{d+1})V^\mathsf{T}$ *on* $\operatorname{Flag}(n_1, \ldots, n_d; n)$ *can be locally parametrized as*

$$c(t) = V \exp(\Lambda(t))(J_1, \ldots, J_{d+1}) \exp(-\Lambda(t))V^\mathsf{T}.$$

For some differentiable curve $\Lambda : (-\epsilon, \epsilon) \to \mathfrak{so}(n)$ *such that* $\Lambda(k, k)(t) \equiv 0$.

If $d = 2$, then a tangent vector of $\operatorname{Flag}(n_1, n_2; n)$ at $\tilde{f}$ can be written as

$$V\left(\begin{bmatrix} 0 & A & B \\ A^\mathsf{T} & 0 & 0 \\ B^\mathsf{T} & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -A & 0 \\ -A^\mathsf{T} & 0 & C \\ 0 & C^\mathsf{T} & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & -B \\ 0 & 0 & -C \\ -B^\mathsf{T} & -C^\mathsf{T} & 0 \end{bmatrix}\right)V^\mathsf{T},$$

where $A \in \mathbb{R}^{m_1 \times m_2}, B \in \mathbb{R}^{m_1 \times m_3}, C \in \mathbb{R}^{m_2 \times m_3}$.

4.1. **induced Riemannian metric, normal space and projections.** As a submanifold of $\prod_{j=1}^{d+1} O(n)$, $\operatorname{Flag}(n_1, \ldots, n_d; n)$ is equipped with a naturally induced Riemannian metric:

$$\langle V(X_1, \ldots, X_{d+1})V^\mathsf{T}, V(Y_1, \ldots, Y_{d+1})V^\mathsf{T} \rangle_{\tilde{\mathsf{f}}} := \sum_{j=1}^{d+1} \operatorname{tr}(X_j Y_j) \tag{42}$$

$$= 2 \sum_{k=1}^{d+1} \sum_{l < k < m} \operatorname{tr}(X_k(l, k)Y_k(k, l) + X_k(m, k)Y_k(k, m)).$$

Unlike (21) in which some summands are weighted differently, all summands in the new metric (42) are evenly weighted. For instance, if we take $d = 2$ then $\langle V(X_1, X_2, X_3)V^\mathsf{T}, V(Y_1, Y_2, Y_3)V^\mathsf{T} \rangle_{\tilde{\mathsf{f}}}$ is simply

$$4\left(\operatorname{tr}(X_1(2, 1)Y_1(1, 2)) + \operatorname{tr}(X_1(3, 1)Y_1(1, 3)) + \operatorname{tr}(X_2(3, 2)Y_2(2, 3))\right). \tag{43}$$

The distinction between (42) and (21) can be easily observed by comparing (43) with (23).

We notice that the tangent space of $\prod_{j=1}^{d+1} O(n)$ at $\tilde{\mathsf{f}} = V(J_1, \ldots, J_{d+1})V^\mathsf{T}$ is

$$\mathbb{T}_{\tilde{\mathsf{f}}}\left(\prod_{j=1}^{d+1} O(n)\right) = \bigoplus_{j=1}^{d+1} (V J_j V^\mathsf{T} \mathfrak{so}(n)) = \bigoplus_{j=1}^{d+1} (V J_j \mathfrak{so}(n) V^\mathsf{T}).$$

Proposition 4.3. *At a point* $\tilde{\mathsf{f}} := V(J_1, \ldots, J_{d+1})V^\mathsf{T} \in \operatorname{Flag}(n_1, \ldots, n_d; \mathbb{R}^n)$, *the normal space* $\mathbb{N}_{\tilde{\mathsf{f}}}\operatorname{Flag}(n_1, \ldots, n_d; n)$ *consists of vectors*

$$V(J_1 Z_1, \ldots, J_{d+1} Z_{d+1})V^\mathsf{T}$$

where $Z_1, \ldots, Z_d \in \mathfrak{so}(n)$ *satisfy the relation*

$$Z_k(k, l) - Z_l(k, l) = 0, \quad \text{for all } 1 \le k \ne l \le d+1.$$

In particular, we have a decomposition

$$\mathbb{N}_{\tilde{\mathfrak{f}}}\operatorname{Flag}(n_1,\ldots,n_d;n) = N_{\tilde{\mathfrak{f}}}\left(\prod_{k=1}^{d+1}\operatorname{Gr}(m_k,n)\right)\bigoplus N_{\tilde{\mathfrak{f}}}^0, \tag{44}$$

where $\mathbb{N}_{\tilde{\mathfrak{f}}}\left(\prod_{k=1}^{d+1}\operatorname{Gr}(m_k,n)\right) = \bigoplus_{k=1}^{d+1}\mathbb{N}_{VJ_{m_k,n-m_k}V^{\mathsf{T}}}\operatorname{Gr}(m_k,n)$ *and*

$$\mathbb{N}_{\tilde{\mathfrak{f}}}^0 = \{V(J_{m_1,n-m_1}Z_1,\ldots,J_{m_{d+1},n-m_{d+1}}Z_{d+1})V^{\mathsf{T}} : Z_k \in \mathfrak{so}(n), Z_k(k,l) - Z_l(k,l) = 0,$$
$$Z_k(k,k) = 0, Z_k(p,q) = 0, 1 \le k,l,p,q \le d+1, p,q \ne k\}. \tag{45}$$

Proposition 4.4 (Projections). *Projections from* $\mathbb{T}_{\tilde{\mathfrak{f}}}\left(\prod_{k=1}^{d+1}\mathrm{O}(n)\right)$ *onto* $\mathbb{T}_{\tilde{\mathfrak{f}}}\operatorname{Flag}(n_1,\ldots,n_d;n)$ *and* $\mathbb{N}_{\tilde{\mathfrak{f}}}\operatorname{Flag}(n_1,\ldots,n_d;n)$ *are respectively given by*

$$\operatorname{proj}_{\tilde{\mathfrak{f}}}^{\mathbb{T}} : \mathbb{T}_{\tilde{\mathfrak{f}}}\left(\prod_{k=1}^{d+1}\mathrm{O}(n)\right) \to \mathbb{T}_{\tilde{\mathfrak{f}}}\operatorname{Flag}(n_1,\ldots,n_d;n)$$
$$V(J_1\Lambda_1,\ldots,J_{d+1}\Lambda_{d+1})V^{\mathsf{T}} \mapsto V(X_1,\ldots,X_{d+1})V^{\mathsf{T}}, \tag{46}$$

and

$$\operatorname{proj}_{\tilde{\mathfrak{f}}}^{\mathbb{N}} : \mathbb{T}_{\tilde{\mathfrak{f}}}\left(\prod_{k=1}^{d+1}\mathrm{O}(n)\right) \to \mathbb{N}_{\tilde{\mathfrak{f}}}\operatorname{Flag}(n_1,\ldots,n_d;n)$$
$$V(J_1\Lambda_1,\ldots,J_{d+1}\Lambda_{d+1})V^{\mathsf{T}} \mapsto V(Z_1,\ldots,Z_{d+1})V^{\mathsf{T}}, \tag{47}$$

where for each $k = 1,\ldots,d$, $X_k \in \mathrm{S}_n$ *(resp.* $Z_k \in \mathbb{R}^{n\times n}$*) is partitioned as* $(X_k(p,q))_{p,q=1}^{d+1}$ *(resp.* $(Z_k(p,q))_{p,q=1}^{d+1}$*) with respect to* $n = m_1 + \cdots + m_{d+1}$ *and*

$$X_k(p,q) = \begin{cases} \frac{1}{2}(\Lambda_k(k,q) - \Lambda_q(k,q)), & \text{if } p = k \ne q \\ -\frac{1}{2}(\Lambda_k(p,k) - \Lambda_p(p,k)), & \text{if } q = k \ne p \\ 0, & \text{otherwise.} \end{cases}$$

$$Z_k(p,q) = \begin{cases} \frac{1}{2}(\Lambda_k(k,q) + \Lambda_q(k,q)), & \text{if } p = k \ne q \\ -\frac{1}{2}(\Lambda_k(p,k) + \Lambda_p(p,k)), & \text{if } q = k \ne p \\ \Lambda_k(p,q), & \text{otherwise.} \end{cases}$$

As an illustrative example, we take a tangent vector ξ of $\mathrm{O}(n) \times \mathrm{O}(n) \times \mathrm{O}(n)$ at some point $\mathfrak{f} = V(J_1, J_2, J_3)V^{\mathsf{T}}$, which can be written as

$$\xi := V\left(\begin{bmatrix} A & B & C \\ B^{\mathsf{T}} & D & E \\ C^{\mathsf{T}} & -E^{\mathsf{T}} & F \end{bmatrix}, \begin{bmatrix} X & Y & Z \\ Y^{\mathsf{T}} & W & S \\ -Z^{\mathsf{T}} & S^{\mathsf{T}} & T \end{bmatrix}, \begin{bmatrix} L & M & N \\ -M^{\mathsf{T}} & P & Q \\ N^{\mathsf{T}} & Q^{\mathsf{T}} & R \end{bmatrix}\right)V^{\mathsf{T}},$$

where $A, X, L \in \mathfrak{so}(m_1)$, $D, W, P \in \mathfrak{so}(m_2)$, $F, T, R \in \mathfrak{so}(m_3)$, $B, Y, M \in \mathbb{R}^{m_1\times m_2}$, $C, Z, N \in \mathbb{R}^{m_1\times m_3}$, $E, S, Q \in \mathbb{R}^{m_2\times m_3}$. The projection of ξ to $T_{\tilde{\mathfrak{f}}}\operatorname{Flag}(n_1, n_2; n)$ is

$$\operatorname{proj}_{\tilde{\mathfrak{f}}}^{\mathbb{T}}(\xi) = V\left(\begin{bmatrix} 0 & \frac{B-Y}{2} & \frac{C-N}{2} \\ \frac{B^{\mathsf{T}}-Y^{\mathsf{T}}}{2} & 0 & 0 \\ \frac{C^{\mathsf{T}}-N^{\mathsf{T}}}{2} & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -\frac{B-Y}{2} & 0 \\ -\frac{B^{\mathsf{T}}-Y^{\mathsf{T}}}{2} & 0 & \frac{S-Q}{2} \\ 0 & \frac{S^{\mathsf{T}}-Q^{\mathsf{T}}}{2} & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & -\frac{C-N}{2} \\ 0 & 0 & -\frac{S-Q}{2} \\ -\frac{C^{\mathsf{T}}-N^{\mathsf{T}}}{2} & -\frac{S^{\mathsf{T}}-Q^{\mathsf{T}}}{2} & 0 \end{bmatrix}\right)V^{\mathsf{T}}$$

and its projection to $N_{\mathfrak{f}}\operatorname{Flag}(n_1, n_2; n)$ is

$$\operatorname{proj}_{\mathfrak{f}}^{\mathbb{N}}(\xi) = V\left(\begin{bmatrix} A & \frac{B+Y}{2} & \frac{C+N}{2} \\ \frac{B^{\mathsf{T}}+Y^{\mathsf{T}}}{2} & D & E \\ \frac{C^{\mathsf{T}}+N^{\mathsf{T}}}{2} & -E^{\mathsf{T}} & F \end{bmatrix}, \begin{bmatrix} X & \frac{B+Y}{2} & Z \\ \frac{B^{\mathsf{T}}+Y^{\mathsf{T}}}{2} & W & \frac{S+Q}{2} \\ -Z^{\mathsf{T}} & \frac{S^{\mathsf{T}}+Q^{\mathsf{T}}}{2} & T \end{bmatrix}, \begin{bmatrix} L & M & \frac{C+N}{2} \\ -M^{\mathsf{T}} & P & \frac{S+Q}{2} \\ \frac{C^{\mathsf{T}}+N^{\mathsf{T}}}{2} & \frac{S^{\mathsf{T}}+Q^{\mathsf{T}}}{2} & T \end{bmatrix}\right)V^{\mathsf{T}}$$

4.2. geodesics. Assume that $c(t)$ is a curve in $\mathrm{Flag}(n_1, \ldots, n_d; n)$, then according to Proposition 4.2 we may parametrize $c(t)$ as

$$c(t) = V(t)(J_1, \ldots, J_{d+1})V(t)^\top \tag{48}$$

for some differentiable curve $V(t)$ in $\mathrm{O}(n)$. Moreover, we have $\dot{V}(t) = V(t)\Lambda(t)$ where $\Lambda(t)$ is a curve in $\mathfrak{so}(n)$ partitioned as $\Lambda(t) = (\Lambda(p,q))_{p,q=1}^{d+1}$ with respect to $m_1 + \cdots + m_{d+1} = n$ and and $\Lambda(k,k)(t) \equiv 0, k = 1, \ldots, d+1$. This implies that we have

$$\ddot{c}(t) = T_1(t) + T_2(t) - 2T_3(t),$$

where $T_j(t)$'s are respectively given by

$$T_1(t) = V(t)(\dot{\Lambda}(t)J_1 - J_1\dot{\Lambda}(t), \ldots, \dot{\Lambda}(t)J_{d+1} - J_{d+1}\dot{\Lambda}(t))V^\top(t), \tag{49}$$

$$T_2(t) = V(t)(\Lambda^2(t)J_1 + J_1\Lambda^2(t), \ldots, \Lambda^2(t)J_{d+1} + J_{d+1}\Lambda^2(t))V^\top(t), \tag{50}$$

$$T_3(t) = V(t)(\Lambda(t)J_1\Lambda(t), \ldots, \Lambda(t)J_{d+1}\Lambda(t))V^\top(t). \tag{51}$$

By similar calculations in proofs of Lemmas 3.7 and 3.8, we may easily obtain the following characterizations of $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_j(t)), j = 1, 2, 3$.

Lemma 4.5. *Let $c(t), \Lambda(t), T_1(t), T_2(t), T_3(t)$ be as above. We have*

(1) $T_1(t) \in \mathbb{T}_{c(t)}\mathrm{Flag}(n_1, \ldots, n_d; n)$.
(2) $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_2(t)) = 0$.
(3) $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_3(t)) = 0$.

Proposition 4.6. *Let $c(t)$ be a curve on $\mathrm{Flag}(n_1, \ldots, n_d; n)$ parametrized as*

$$c(t) = V(t)(J_1, \ldots, J_{d+1})V(t)^\top$$

for some differentiable curve $V(t)$ in $\mathrm{O}(n)$. Let $\Lambda(t)$ be the curve in $\mathfrak{so}(n)$ such that $\dot{V}(t) = V(t)\Lambda(t)$, where $\Lambda(t)$ is a curve in $\mathfrak{so}(n)$ partitioned as $\Lambda(t) = (\Lambda(p,q))_{p,q=1}^{d+1}$ with respect to $m_1 + \cdots + m_{d+1} = n$ and and $\Lambda(k,k)(t) \equiv 0, k = 1, \ldots, d+1$. Then $c(t)$ is a geodesic curve if and only if $V(t) = V(0)\exp(t\Lambda(0))$.

Proof. Since $c(t)$ is a geodesic if and only if $\mathrm{proj}_{c(t)}(\ddot{c}(t)) \equiv 0$, Lemma 4.5 implies that $c(t)$ is a geodesic curve if and only if

$$\dot{\Lambda}(t)J_k - J_k\dot{\Lambda}(t) = 0, \quad k = 1, \ldots, d+1.$$

By (19), we may conclude that $c(t)$ is a geodesic if and only if $\dot{\Lambda}(t) \equiv 0$, i.e., $\Lambda(t) = \Lambda(0)$. This implies that $V(t)$ is determined by the equation $\dot{V}(t) = V(t)\Lambda(0)$, from which we may conclude that $V(t) = V(0)\exp(t\Lambda(0))$. $\qquad\square$

5. THE COMPARISON OF RIEMANNIAN METRICS ON FLAG MANIFOLDS

The goal of this section is to discuss relations among three Riemannian metrics on a flag manifold $\mathrm{Flag}(n_1, \ldots, n_d; n)$. We recall that the two metrics discussed in this paper are respectively induced by the embedding $\epsilon : \mathrm{Flag}(n_1, \ldots, n_d; n) \hookrightarrow \prod_{k=1}^d \mathrm{O}(n)$ given in (16) and $\tilde{\epsilon} : \mathrm{Flag}(n_1, \ldots, n_d; n) \hookrightarrow \prod_{k=1}^{d+1} \mathrm{O}(n)$ given in (38). For notational simplicity, we denote the two induced metrics by g^e and $\tilde{g}^e$, respectively. Yet there is another metric induced from the homogeneous space structure of $\mathrm{Flag}(n_1, \ldots, n_d; n)$, which is discussed thoroughly in [5]. We denote this quotient metric by g^q.

Proposition 5.1. *The Riemaannian metrics $\tilde{g}^e$ and g^q coincide. Moreover, $\tilde{g}^e$ and g^e coincide with g^q when $d = 1$, in which case $\mathrm{Flag}(n_1, \ldots, n_d; n)$ is simply the Grassmann manifold $\mathrm{Gr}(n_1; n)$.*

We will see in Proposition 5.2 that both g^e and $\tilde{g}^e = g^q$ can be constructed by a uniform method. To begin with, we notice that in general, any smooth map

$$\varphi : \left(\mathbb{R}^{n \times n}\right)^d \to \mathbb{R}^{n \times n}$$

induces an embedding $\kappa_\varphi : (\mathbb{R}^{n \times n})^d \to (\mathbb{R}^{n \times n})^{d+1}$ defined by

$$\kappa_\varphi(A_1, \ldots, A_d) = (A_1, \ldots, A_d, \varphi(A_1, \ldots, A_d)), \quad A_j \in \mathbb{R}^{n \times n}, j = 1, \ldots, d.$$

Hence we have another embedding $\kappa_\varphi \circ \epsilon$ of $\mathrm{Flag}(n_1, \ldots, n_d; n)$ into $\mathrm{O}(n)^{d+1} \subseteq (\mathbb{R}^{n \times n})^{d+1}$, which induces a metric g^φ on $\mathrm{Flag}(n_1, \ldots, n_d; n)$ from the Euclidean metric on $(\mathbb{R}^{n \times n})^{d+1}$.

Proposition 5.2. *We have the following:*

- *$g^\varphi = g^e$ if and only if φ is a constant map on $\epsilon(\mathrm{Flag}(n_1, \ldots, n_d; n))$. In particular, $g^\varphi = g^e$ if φ is a constant map.*
- *There exists φ such that $g^\varphi = \tilde{g}^e$.*

Proof. The "if" part of the first statement can be verified by a straightforward calculation. For the "only if" part, we notice that $g^\varphi = g^e$ implies that the differential map $d_{(Q_1, \ldots, Q_d)}\varphi$ must be zero on $T_{(Q_1, \ldots, Q_d)}\epsilon(\mathrm{Flag}(n_1, \ldots, n_d; n))$ at any $(Q_1, \ldots, Q_d) \in \epsilon(\mathrm{Flag}(n_1, \ldots, n_d; n))$. Since $\epsilon(\mathrm{Flag}(n_1, \ldots, n_d; n))$ is connected and φ is continuous, we may conclude that φ is a constant map on $\epsilon(\mathrm{Flag}(n_1, \ldots, n_d; n))$.

For the second statement, we notice that $C := \epsilon(\mathrm{Flag}(n_1, \ldots, n_d; n))$ is a compact subset of $X := (\mathbb{R}^{n \times n})^d$ and we can define

$$\psi : C \to \mathrm{O}(n) \subseteq \mathbb{R}^{n \times n}, \quad \psi(Q_1, \ldots, Q_d) = Q_{d+1},$$

where $(Q_{d+1} + \mathrm{I}_n)/2$ is the projection matrix of $\left(\bigoplus_{j=1}^d \mathrm{im}(Q_j + \mathrm{I}_n)\right)^\perp$. We denote by p_{ij} the projection map from $\mathbb{R}^{n \times n}$ onto its (i, j)-th entry, $1 \le i, j \le n$. It is clear that $p_{ij} \circ \psi : C \to \mathbb{R}$ is a smooth function. The compactness of C in X implies that $p_{ij} \circ \psi$ has a smooth extension $\varphi_{ij} : X \to \mathbb{R}$. Indeed, we can first extend the function $p_{ij} \circ \psi$ smoothly to an open neighbourhood of C and then further extend it smoothly to the whole X by a smooth partition of unity. Now we have a smooth map

$$\varphi := (\varphi_{ij}) : \left(\mathbb{R}^{n \times n}\right)^d \to \mathbb{R}^{n \times n}$$

which extends ψ and hence we have $g^\varphi = \tilde{g}^e$. $\square$

6. An alternating method for optimizations on flag manifolds

Given a strictly increasing sequence $n_1 < \cdots < n_d$, we define

$$m_1 := n_1, \quad m_{d+1} := n - n_d, \quad m_j := n_j - n_{j-1}, \quad j = 2, \ldots, d+1.$$

We recall from (37) that a flag $\{\mathbb{V}_k\}_{k=1}^d \in \mathrm{Flag}(n_1, \ldots, n_d; n)$ can be regarded as $\{\mathbb{W}_j\}_{j=1}^{d+1}$ via the modified embedding $\tilde{\iota} : \mathrm{Flag}(n_1, \ldots, n_d; n) \hookrightarrow \prod_{j=1}^{d+1} \mathrm{Gr}(m_j, n)$, where $\mathbb{W}_j$ is the orthogonal complement of $\mathbb{V}_{j-1}$ in $\mathbb{V}_j, 2 \le j \le d+1$, $\mathbb{W}_1 = \mathbb{V}_1$ and $\mathbb{V}_{d+1} = \mathbb{R}^n$. Therefore, an optimization problem on $\mathrm{Flag}(n_1, \ldots, n_d; n)$ has the following form:

$$
\begin{aligned}
\min \quad & f(\mathbb{W}_1, \ldots, \mathbb{W}_{d+1}) \\
\text{s.t.} \quad & \mathbb{W}_j \in \mathrm{Gr}(m_j, n), 1 \le j \le d+1 \\
& \mathbb{W}_j \perp \mathbb{W}_l, 1 \le j < l \le d+1
\end{aligned}
\tag{52}
$$

Here f is a function on $\mathrm{Flag}(n_1, \ldots, n_d; n)$. We propose Algorithm 1, an alternating type algorithm to solve the problem (52).

Algorithm 1 Alternating method for optimization on flag manifolds

---

Input A differentiable function f on $\mathrm{Flag}(n_1, \ldots, n_d; n)$
Output A critical point of f
Initialization Choose an initial point $(\mathbb{W}_1, \ldots, \mathbb{W}_{d+1}) \in \prod_{j=1}^{d+1} \mathrm{Gr}(m_j, n)$

1: **while** not converge **do**
2:　　set $(s, t) = (1, 2)$
3:　　**for** $1 \leq s < t \leq d+1$ **do**
4:　　　　Solve the following sub-problem for $(\mathbb{X}_s, \mathbb{X}_t) \in \mathrm{Gr}(m_s, n) \times \mathrm{Gr}(m_t, n)$:

$$
\begin{aligned}
\min \quad & f(\mathbb{W}_1, \ldots, \mathbb{W}_{s-1}, \mathbb{X}_s, \mathbb{W}_{s+1}, \ldots, \mathbb{W}_{t-1}, \mathbb{X}_t, \mathbb{W}_{t+1}, \ldots \mathbb{W}_{d+1}) \\
\text{s.t.} \quad & \mathbb{X}_s \perp \mathbb{X}_t \\
& \mathbb{X}_s \perp \mathbb{W}_j, 1 \leq j \neq s \leq d+1 \\
& \mathbb{X}_t \perp \mathbb{W}_j, 1 \leq j \neq t \leq d+1
\end{aligned}
\tag{53}
$$

5:　　　　Update $(\mathbb{W}_s, \mathbb{W}_t)$ by the solution $(\overline{\mathbb{X}}_s, \overline{\mathbb{X}}_t)$ to (53).
6:　　　　Update (s, t) by $(s+1, t)$ if $s+1 < t$ and by $(s, t+1)$ otherwise
7:　　**end for**
8: **end while**

We remark that the sub-problem (53) in Algorithm 1 is an optimization problem on a Grassmann manifold. Indeed, we notice that $\mathbb{W}_j$ in (53) is fixed whenever $j \neq s, t$. This implies

$$
\mathbb{X}_s \oplus \mathbb{X}_t = \left(\bigoplus_{j \neq s,t} \mathbb{W}_j \right)^{\perp}
$$

is a fixed $(m_s + m_t)$-dimensional subspace of $\mathbb{R}^n$. So the submanifold given by fixed $\mathbb{W}_j, j \neq s, t$ and $\mathbb{X}_s \oplus \mathbb{X}_t$ is isomorphic to $\mathrm{Gr}(m_s, m_s + m_t)$. This submanifold is actually a totally-geodesic manifold, which is clear from the geodesic formulas of flag and Grassmann manifolds. Thus the objective function

$$
f(\mathbb{W}_1, \ldots, \mathbb{W}_{s-1}, \mathbb{X}_s, \mathbb{W}_{s+1}, \ldots, \mathbb{W}_{t-1}, \mathbb{X}_t, \mathbb{W}_{t+1}, \ldots \mathbb{W}_{d+1})
$$

can be recognized as a function on this submanifold $\mathrm{Gr}(m_s, m_s + m_t)$. Furthermore, at a given point, there are $d(d+1)/2$ such submanifolds indexed by $1 \leq s < t \leq d+1$. The tangent spaces of those submanifolds are orthogonal to each other and span the whole tangent space. Algorithm 1 is a generalization of coordinate descent algorithm in Euclidean space.

6.1. **separation of subspaces.** Given $d+1$ subspaces $\mathbb{U}_1, \ldots, \mathbb{U}_{d+1}$ of some ambient space $\mathbb{R}^N$. The separation problem can be mathematically formulated as the following optimization problem on a flag manifold:

$$
\begin{aligned}
\min \quad & F(\mathbb{W}) := \sum_{j=1}^{d+1} \| \tau_j(\mathbb{U}_j) - \tau_j(\mathbb{W}_j) \|_F^2 \\
\text{s.t.} \quad & \mathbb{W}_j \in \mathrm{Gr}(m_j, n), 1 \leq j \leq d+1 \\
& \mathbb{W}_j \perp \mathbb{W}_l, 1 \leq j < l \leq d+1
\end{aligned}
\tag{54}
$$

Here $m_j = \dim \mathbb{U}_j, 1 \leq j \leq d+1$, $n = \sum_{j=1}^{d+1} m_j$ and τ_j is the embedding of $\mathrm{Gr}(m_j, n)$ into $\mathrm{O}(n) \cap \mathrm{S}_n$ defined by

$$
\tau_j(\mathbb{W}) = V \begin{bmatrix} -\mathrm{I}_p & 0 & 0 \\ 0 & \mathrm{I}_{m_j} & 0 \\ 0 & 0 & -\mathrm{I}_q \end{bmatrix} V^{\mathsf{T}}
$$

where $p = \sum_{l=1}^{j-1} m_l$, $q = \sum_{l=j+1}^{d+1} m_l$ and $V = [v_1, \ldots, v_n] = [V_1, \ldots, V_{d+1}] \in \mathrm{O}(n)$ such that $[v_{p+1}, \ldots, v_{q-1}] = V_j, \mathrm{span}\{v_{p+1}, \ldots, v_{q-1}\} = \mathbb{W}_\mathsf{J}$.

Lemma 6.1. *Consider the maximization of linear function $f(Q) = \langle A, Q \rangle$ on the Grassmann manifold,*

$$\begin{aligned} \max \quad & \langle A, Q \rangle \\ s.t. \quad & Q \in \mathrm{Gr}(k, n) \end{aligned}$$

The gradient of $f(Q)$ is given by

$$\nabla f(Q) = \frac{1}{4}(A + A^\mathsf{T} - QAQ - QA^\mathsf{T}Q).$$

Let $(A + A^\mathsf{T})/2 = U\Lambda U^\mathsf{T}$ be an eigendecomposition of $(A + A^\mathsf{T})/2$ such that $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $\lambda_1 \geq \cdots \geq \lambda_n$. Then $Q^* = UI_{k,n-k}U^\mathsf{T}$ is a maximizer of $f(Q)$. Furthermore,

$$2\|\Lambda\|(f(Q^*) - f(Q)) \geq \|\nabla f(Q)\|^2.$$

Proof. The formula for gradient is given in [4, Proposition 5.1]. The original problem is equivalent to

$$\begin{aligned} \max \quad & \langle \Lambda, Q \rangle \\ s.t. \quad & Q \in \mathrm{Gr}(k, n) \end{aligned}$$

and we need to prove $Q = I_{k,n-k}$ is a maximizer. Using gradient formula, we can simplify the first order condition $\nabla f(Q^*) = 0$ to

$$Q^*\Lambda = \Lambda Q^*.$$

So Q^*, A can be simultaneously diagonalized, and we can assume Q^* is diagonalized. The original problem is equivalent to

$$\min_{\delta_1 + \cdots + \delta_n = 2k - n, \delta_i = \pm 1} \lambda_1 \delta_1 + \cdots + \lambda_n \delta_n.$$

It is clear that $\delta_1 = \cdots = \delta_k = 1, \delta_{k+1} = \cdots = \delta_n = -1$ is a maximizer. So $Q^* = I_{k,n-k}$ is a maximizer. Now consider the last inequality. The term $\|\nabla f(Q)\|^2, f(Q^*) - f(Q)$ can be simplified

$$\begin{aligned} \|\nabla f(Q)\|^2 &= \frac{1}{4}\langle \Lambda - Q\Lambda Q, \Lambda - Q\Lambda Q \rangle \\ &= \frac{1}{2}\sum_{i=1}^{n} \lambda_i^2 - \frac{1}{2}\mathrm{tr}(\Lambda Q\Lambda Q) \\ &= \frac{1}{2}\mathrm{tr}(\Lambda Q^*\Lambda Q^*) - \frac{1}{2}\mathrm{tr}(\Lambda Q\Lambda Q), \end{aligned}$$

$$f(Q^*) - f(Q) = \langle \Lambda, Q^* \rangle - \langle \Lambda, Q \rangle.$$

For any $c > 2\|\Lambda\|$, we have

$$\begin{aligned} c(f(Q^*) - f(Q)) - \|\nabla f(Q)\|^2 &= c\langle \Lambda, Q^* \rangle - c\langle \Lambda, Q \rangle - \mathrm{tr}(\Lambda Q^*\Lambda Q^*)/2 + \mathrm{tr}(\Lambda Q\Lambda Q)/2 \\ &= g(Q^*) - g(Q), \end{aligned}$$

where $g(Q) = c\langle \Lambda, Q \rangle - \mathrm{tr}(\Lambda Q\Lambda Q)/2$. Assume Q^{**} is a maximizer of $g(Q)$. The first order condition of $g(Q)$ is

$$Q^{**}\Lambda Q^{**}\Lambda Q^{**} - cQ^{**}\Lambda Q^{**} - \Lambda Q^{**}\Lambda + c\Lambda = 0,$$

which is equivalent to

$$(Q^{**}\Lambda - \Lambda Q^{**})(Q^{**}\Lambda + \Lambda Q^{**} - cI) = 0.$$

By definition $c > 2\|\Lambda\| \geq \|Q^{**}\Lambda + \Lambda Q^{**}\|$, so $Q^{**}\Lambda + \Lambda Q^{**} - cI$ is invertible and $Q^{**}\Lambda = \Lambda Q^{**}$. So Q^{**}, Λ can be simultaneously diagonalized. We can assume Q^{**} is diagonalized. So

$$g(Q^{**}) = \sum_{i=1}^{n}(2\|\Lambda\|\lambda_i\delta_i - \frac{1}{2}\lambda_i^2),$$

where δ_i is the diagonal of Q^{**}. Again, Q^* is a maximizer of $g(Q)$. So we have proved that $g(Q^*) \geq g(Q)$, i.e.,

$$c(f(Q^*) - f(Q)) - \|\nabla f(Q)\|^2 \geq 0.$$

Because c is any number larger than $2\|\Lambda\|$, it also holds for $c = 2\|\Lambda\|$ and the proof is finished. $\square$

Proposition 6.2. *If we apply Algorithm 1 to solve the problem* (54), *then for each* $1 \leq s < t \leq d+1$, *the sub-problem has the form*

$$\min \quad \|A_1 - W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}\|_F^2 + \|A_2 + W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}\|_F^2 \tag{55}$$
$$s.t. \quad W \in \mathrm{O}(m_s + m_t)$$

where $A_1, A_2 \in \mathrm{O}(m_s + m_t) \cap S_{m_s+m_t}$ *are some fixed matrices. Moreover, the sub-problem has an explicit solution* W_* *which is given by the SVD of* $A_1 - A_2$, $A_1 - A_2 = W_*\Sigma W_*^\mathsf{T}$.

Denote the change of the value of F at this step by $\Delta_{s,t}$. By previous discussion, the full gradient ∇F can be partition into $d(d-1)/2$ components $\nabla_{s,t}F$ that corresponds to the subproblems. Then

$$\|\tau_s(\mathbb{U}_s) - \tau_t(\mathbb{U}_t)\|\,|\Delta_{s,t}| \geq \|\nabla_{s,t}F\|^2.$$

Proof. Given $1 \leq s < t \leq d+1$, the sub-step in (54) is

$$\min \quad \|\tau_s(\mathbb{U}_s) - \tau_s(\mathbb{W}_s)\|_F^2 + \|\tau_t(\mathbb{U}_t) - \tau_t(\mathbb{W}_t)\|_F^2$$
$$s.t. \quad (\mathbb{W}_s, \mathbb{W}_t) \in \mathrm{Gr}(m_s, m_s + m_t) \times \mathrm{Gr}(m_t, m_s + m_t)$$
$$\mathbb{W}_s \perp \mathbb{W}_t$$

In particular, $\mathbb{W}_s \oplus \mathbb{W}_t = \left(\bigoplus_{j \neq s,t} \mathbb{W}_j\right)^\perp$ is a fixed $(m_s + m_t)$-dimensional vector space represented by $V_{s,t} := [V_s, V_t]$. We use $V^\perp$ to denote the basis of $\left(\bigoplus_{j \neq s,t} \mathbb{W}_j\right)^\perp$. The choice of $\mathbb{W}_s, \mathbb{W}_t$ can be further specified by a orthogonal matrix $W \in \mathrm{O}(m_s + m_t)$ so that $V_{s,t}W = [W_s, W_t]$ where W_s, W_t span $\mathbb{W}_s, \mathbb{W}_t$ respectively. As a result, the images of $\mathbb{W}_s, \mathbb{W}_t$ can be written as

$$\tau_s(\mathbb{W}_s) = V_{s,t}W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}V_{s,t}^\mathsf{T} + V^\perp(V^\perp)^\mathsf{T}, \quad \tau_t(\mathbb{W}_t) = -V_{s,t}W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}V_{s,t}^\mathsf{T} + V^\perp(V^\perp)^\mathsf{T}.$$

The desired (55) follows easily by taking $A_1 = V_{s,t}^\mathsf{T}\tau_s(\mathbb{U}_s)V_{s,t}$, $A_2 = V_{s,t}^\mathsf{T}\tau_t(\mathbb{U}_t)V_{s,t}$.

Next we observe that the objective function in (55) can further be re-written as

$$\|A_1\|_F^2 + \|A_2\|_F^2 + 2(m_s + m_t) - 2\langle A_1, W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}\rangle + 2\langle A_2, W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}\rangle$$
$$=\|A_1\|_F^2 + \|A_2\|_F^2 + 2(m_s + m_t) + 2\langle A_2 - A_1, W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}\rangle.$$

Therefore, the problem (55) is equivalent to

$$\min \quad \langle A_2 - A_1, W\,\mathrm{I}_{m_s,m_t}\,W^\mathsf{T}\rangle \tag{56}$$
$$s.t. \quad W \in \mathrm{O}(m_s + m_t)$$

By previous lemma, we may conclude that a solution to (56) is W_*, which can be obtained by the SVD of $A_1 - A_2$. Furthermore,

$$\|\nabla_{s,t}F\|^2 \leq \|A_2 - A_1\|\,|\Delta_{s,t}| \leq \|\tau_s(\mathbb{U}_s) - \tau_t(\mathbb{U}_t)\|\,|\Delta_{s,t}|.$$

$\square$

Theorem 6.3. *Consider a randomized version of Algorithm 1 for problem (54). At each step, choose (s_i, t_i) uniformly from all possible (s, t). Let $\mathbb{W}_i$ be the point at step i. Then with probability 1, every cluster point of $\mathbb{W}_i$ is a stationary point. (Because flag manifold is compact, cluster point exists.)*

Proof. If $\|\tau_s(\mathbb{U}_s) - \tau_t(\mathbb{U}_t)\| = 0$ for all s, t, then the function is trivial and there is nothing to prove. Otherwise, there is a set $A \subseteq \{(s, t) \mid 1 \le s < t \le d + 1\}$ such that $\|\tau_s(\mathbb{U}_s) - \tau_t(\mathbb{U}_t)\| \ne 0$ if and only if $(s, t) \in A$. At each step i, assume $\operatorname{argmax}_{(s,t) \in A} \|\nabla_{s,t} F(\mathbb{W}_i)\|$ is achieved for (s^*, t^*). If $(s_i, t_i) = (s^*, t^*)$, then

$$
\begin{aligned}
F(\mathbb{W}_i) - F(\mathbb{W}_{i+1}) &\ge \frac{\|\nabla_{s^*,t^*} F(\mathbb{W}_i)\|^2}{\|\tau_{s^*}(\mathbb{U}_{s^*}) - \tau_{t^*}(\mathbb{U}_{t^*})\|} \\
&\ge \frac{\max \|\nabla_{s,t} F(\mathbb{W}_i)\|^2}{\max \|\tau_s(\mathbb{U}_s) - \tau_t(\mathbb{U}_t)\|} \\
&\ge C \|\nabla F(\mathbb{W}_i)\|^2,
\end{aligned}
$$

where C is a constant independent of $\mathbb{W}_i$. If $(s_i, t_i) \ne (s^*, t^*)$, at least we have $F(\mathbb{W}_i) - F(\mathbb{W}_{i+1}) \ge 0$. So

$$
\mathbb{E} F(\mathbb{W}_i) - \mathbb{E} F(\mathbb{W}_{i+1}) \ge \frac{2C}{n(n-1)} \|\nabla F(\mathbb{W}_i)\|^2.
$$

Summing from $i = 0$ to ∞, and take expectation, we have

$$
\mathbb{E}[F(\mathbb{W}_0) - \lim_{i \to \infty} F(\mathbb{W}_i)] \ge C' \mathbb{E} \sum_{i=0}^{\infty} \|\nabla F(\mathbb{W}_i)\|^2.
$$

So with probability 1, $\sum_{i=0}^{\infty} \|\nabla F(\mathbb{W}_i)\|^2$ exists and $\|\nabla F(\mathbb{W}_i)\|$ converges to 0. Any cluster point must be a stationary point. $\qquad\square$

7. Numerical experiments

In this section, we consider the function

$$
f(V) = \sum_{k=1}^{d} \operatorname{tr}(V_k^\mathsf{T} A_k V_k),
$$

where A_i is randomly generated symmetric matrix, V_i is the submatrix of V with index $1 \le i \le n, n_{k-1} < j \le n_k$, i.e., the basis of $\mathbb{W}_k$. This function is clearly a function on the flag manifold $\operatorname{Flag}(n_1, \ldots, n_d; n)$.

We choose $\operatorname{Flag}(5, 5; 200)$ and test five methods: (i) gradient descent method under classical embedding metric; (ii) gradient descent method under modified embedding metric; (iii) coordinate gradient descent method under modified embedding metric; (iv) gradient descent method using the quotient model proposed in Algorithm 1 in [5]; (v) coordinate minimization method under modified metric (Algorithm 1). Figure 1 shows the convergence rate averaged over 10 simulations. We also record the running time to hit $\|\nabla f(V)\| \le 10^{-5}$, averaged over 10 simulations, as shown in Table 1.

Method (ii) is equivalent to (iv), and their convergence rate and running time are similar. All four descent method has comparable performance while the coordinate minimization outperforms them significantly.

The coordiante minimization method works best for this special choice of $f(V)$ because the optimization sub-problem has explicit solution and can be solved sufficiently. For more general problems, it might not be the case. However, This special choice of f covers many common problems appeared in flag manifolds optimization. Most notably, it is equivalent to the projection

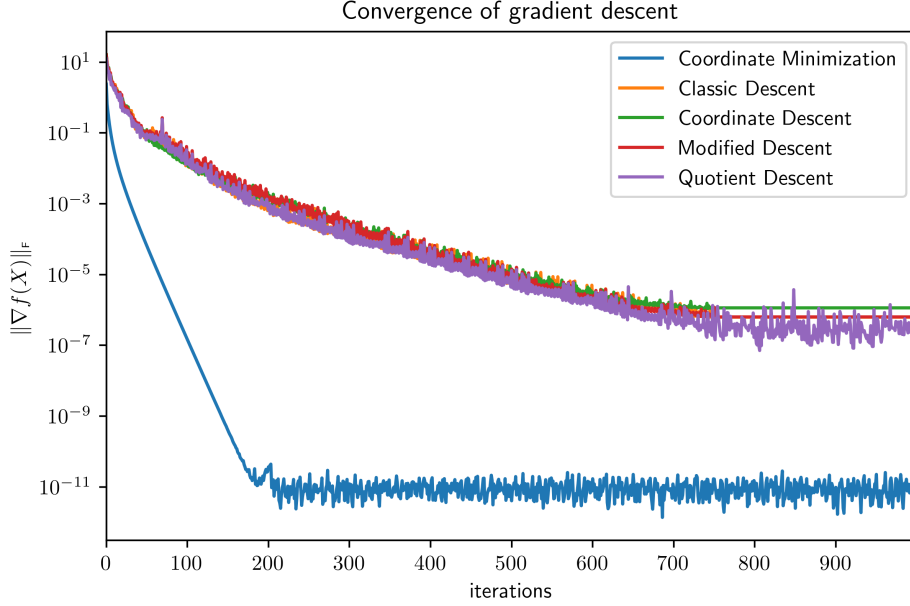|	
(i) Classic Descent	21.428s
(ii) Modified Descent	11.291s
(iii) Coordinate Descent	25.548s
(iv) Quotient Descent	10.238s
(v) Coordinate Minimization	0.552s

TABLE 1. Running time to hit $\|\nabla f(V)\| \leq 10^{-5}$ of different methods.

FIGURE 1. Convergence behavior of different methods.

problem under modified embedding 54. As a result, the extrinsic sample mean problem [1] for flag manifold under modified embedding can be solved efficiently.

REFERENCES

[1] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29, 2003.
[2] R. W. Brockett. *Finite dimensional linear systems*, volume 74 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015. Reprint of the 1970 original.
[3] W. Fulton and J. Harris. *Representation theory*, volume 129 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. A first course, Readings in Mathematics.
[4] Z. Lai, L.-H. Lim, and K. Ye. Simpler grassmannian optimization, 2020.
[5] K. Ye, K. S. Wong, and L. Lim. Optimization on flag manifolds. *arXiv: Optimization and Control*, 2019.

APPENDIX A. PARALLEL TRANSPORT WITH RESPECT TO THE CLASSICAL EMBEDDING

Let $c(t)$ be a curve on $\mathrm{Flag}(n_1, \ldots, n_d; n)$ parametrized as

$$c(t) = V(t)(J_1, \ldots, J_d)V(t)^\mathsf{T}.$$

Here $V(t)$ is a curve in $\mathrm{O}(n)$ and hence $\dot{V}(t) = V(t)\Lambda(t)$ for some $\Lambda(t) \in \mathfrak{so}(n)$. If we partition $\Lambda(t)$ as $\Lambda(t) = (\Lambda(j, k))_{j,k=1}^{d+1, d+1}$ with respect to $n = m_1 + \cdots + m_{d+1}$, then $\Lambda(k, k)(t) \equiv 0, k = 1, \ldots, d+1$.

We notice that by Proposition 3.2 a vector field $Y(t)$ on $\mathrm{Flag}(n_1,\ldots,n_d;n)$ along the curve $c(t)$ can be parametrize as

$$Y(t) = V(t)(X(t)J_1 - J_1 X(t), \ldots, X(t)J_d - J_d X(t))V(t)^\mathsf{T}, \tag{57}$$

where $X(t) = (X(j,k))_{j,k=1}^{d+1,d+1} \in \mathfrak{so}(n)$ is the partition of $X(t)$ with respect to $n = m_1 + \cdots + m_{d+1}$ and $X(k,k)(t) \equiv 0, k = 1,\ldots,d+1$. We recall that $Y(t)$ is the parallel transport of $Y(0)$ along $c(t)$ if and only if

$$\mathrm{proj}_{c(t)}^{\mathbb{T}}(\dot{Y}(t)) = 0.$$

By differentiating (57), we obtain

$$\dot{Y}(t) = V(t)(\Delta_1(t), \ldots, \Delta_d(t))V^\mathsf{T}(t),$$

where

$$\Delta_k(t) = (\dot{X}(t)J_k - J_k \dot{X}(t)) + \Lambda(t)(X(t)J_k - J_k X(t)) - (X(t)J_k - J_k X(t))\Lambda(t)$$
$$= (\dot{X}(t)J_k - J_k \dot{X}(t)) + (\Lambda(t)X(t)J_k + J_k X(t)\Lambda(t)) - (\Lambda(t)J_k X(t) + X(t)J_k \Lambda(t)).$$

Similar to what we have done in Subsection 3.3, we set

$$T_1(t) = V(t)(\dot{X}(t)J_1 - J_1 \dot{X}(t), \ldots, \dot{X}(t)J_d - J_d \dot{X}(t))V(t)^\mathsf{T}, \tag{58}$$
$$T_2(t) = V(t)(\Lambda(t)X(t)J_1 + J_1 X(t)\Lambda(t), \ldots, \Lambda(t)X(t)J_d + J_d X(t)\Lambda(t))V^\mathsf{T}(t), \tag{59}$$
$$T_3(t) = V(t)(\Lambda(t)J_1 X(t) + X(t)J_1 \Lambda(t), \ldots, \Lambda(t)J_d X(t) + X(t)J_d \Lambda(t))V^\mathsf{T}(t), \tag{60}$$

thus $\dot{Y}(t) = T_1(t) + T_2(t) - T_3(t)$. By definition, we conclude that $T_1(t) \in T_{c(t)}\,\mathrm{Flag}(n_1,\ldots,n_d;n)$ and hence to determine $\mathrm{proj}_{c(t)}^{\mathbb{T}}(\dot{Y}(t))$, it is sufficient to compute $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_2(t))$ and $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_3(t))$ respectively.

Lemma A.1. *There exists some symmetric matrices $W_1,\ldots,W_d$ where*

$$W_k(p,q) = \begin{cases} \sum_{s=1}^{d+1} X(k,s)\Lambda(s,q) - \sum_{s=1}^{d+1} \Lambda(s,k)^\mathsf{T} X(q,s)^\mathsf{T}, & \text{if } p = k, q \neq k, \\ \sum_{s=1}^{d+1} \Lambda(s,p)^\mathsf{T} X(k,s)^\mathsf{T} - \sum_{s=1}^{d+1} X(p,s)\Lambda(s,k), & \text{if } q = k, p \neq k, \\ 0, & \text{otherwise} \end{cases}$$

such that $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_2(t)) = V(t)(W_1(t),\ldots,W_d(t))V^\mathsf{T}(t)$.

Proof. For each $k = 1,\ldots,d$, we first compute $\Lambda(t)X(t)J_k + J_k X(t)\Lambda(t)$. Indeed, since $X(t), \Lambda(t) \in \mathfrak{so}(n)$ and J_k is a diagonal matrix, we have

$$J_k X(t)\Lambda(t) = (\Lambda(t)X(t)J_k)^\mathsf{T}.$$

Therefore we only need to compute $\Lambda(t)X(t)J_k$. To do so, we partition $\Lambda(t)$ and $X(t)$ as

$$\Lambda(t) = (\Lambda(p,q)), \quad X(t) = (X(p,q)), \quad 1 \leq p, q \leq d+1. \tag{61}$$

The (p,q)-th entry of $J_k X(t)\Lambda(t)$ is

$$\sum_{l,s=1}^{d+1} J_k(p,l)X(l,s)\Lambda(s,q) = \sum_{s=1}^{d+1} J_k(p,p)X(p,s)\Lambda(s,q)$$
$$= (2\delta_{pk} - 1)\sum_{s=1}^{d+1} X(p,s)\Lambda(s,q).$$

Hence the (p,q)-th block of $\Lambda(t)X(t)J_k + J_k X(t)\Lambda(t)$ is simply

$$(2\delta_{pk} - 1)\sum_{s=1}^{d+1} X(p,s)\Lambda(s,q) + (2\delta_{qk} - 1)\sum_{s=1}^{d+1} \Lambda(s,p)^\mathsf{T} X(q,s)^\mathsf{T}.$$

In particular, for $q \neq k$, the (k, q)-th block of $\Lambda(t)X(t)J_k + J_kX(t)\Lambda(t)$ is

$$\sum_{s=1}^{d+1} X(k,s)\Lambda(s,q) - \sum_{s=1}^{d+1} \Lambda(s,k)^\mathsf{T}X(q,s)^\mathsf{T}$$

and if moreover that $q \leq d$, then the (k, q)-th block of $\Lambda(t)X(t)J_q + J_qX(t)\Lambda(t)$ is

$$-\left(\sum_{s=1}^{d+1} X(k,s)\Lambda(s,q) - \sum_{s=1}^{d+1} \Lambda(s,k)^\mathsf{T}X(q,s)^\mathsf{T}\right).$$

This implies that

$$\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_2(t)) = V(t)(W_1(t),\ldots,W_d(t))V^\mathsf{T}(t),$$

where

$$W_k(p,q) = \begin{cases} \sum_{s=1}^{d+1} X(k,s)\Lambda(s,q) - \sum_{s=1}^{d+1} \Lambda(s,k)^\mathsf{T}X(q,s)^\mathsf{T}, & \text{if } p = k, q \neq k, \\ \sum_{s=1}^{d+1} \Lambda(s,p)^\mathsf{T}X(k,s)^\mathsf{T} - \sum_{s=1}^{d+1} X(p,s)\Lambda(s,k), & \text{if } q = k, p \neq k, \\ 0, & \text{otherwise.} \end{cases}$$

$\square$

Lemma A.2. *There exists symmetric matrices* $Z_1, \ldots, Z_d$ *where*

$$Z_k(p,q) = \begin{cases} -\sum_{1 \leq l \leq d, l \neq k} \left(X(k,l)\Lambda(l,d+1) + \Lambda(l,k)^\mathsf{T}X(d+1,l)^\mathsf{T}\right), & \text{if } p = k, q = d+1, \\ -\sum_{1 \leq l \leq d, l \neq k} \left(X(d+1,l)\Lambda(l,k) + \Lambda(l,d+1)^\mathsf{T}X(k,l)^\mathsf{T}\right), & \text{if } p = d+1, q = k, \\ 0, & \text{otherwise.} \end{cases}$$

such that $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_3(t)) = V(t)(Z_1(t), \ldots, Z_d(t))V^\mathsf{T}(t)$.

Proof. We compute $\Lambda(t)J_kX(t) + X(t)J_k\Lambda(t)$ for each $k = 1, \ldots, d$. We partition $\Lambda(t)$ and $X(t)$ as in (61) respectively. We also notice that

$$X(t)J_k\Lambda(t) = (\Lambda(t)J_kX(t))^\mathsf{T}$$

so that it is sufficient to compute $X(t)J_k\Lambda(t)$. The (p, q)-th block of $X(t)J_k\Lambda(t)$ is

$$\sum_{l,s=1}^{d+1} X(p,l)J_k(l,s)\Lambda(s,q) = \sum_{l=1}^{d+1} X(p,l)J_k(l,l)\Lambda(l,q)$$

$$= \sum_{l=1}^{d+1} (2\delta_{kl} - 1)X(p,l)\Lambda(l,q)$$

Hence the (p, q)-th block of $\Lambda(t)J_kX(t) + X(t)J_k\Lambda(t)$ is

$$\sum_{1 \leq l \leq d+1, l \neq p,q} (2\delta_{kl} - 1)\left(X(p,l)\Lambda(l,q) + \Lambda(l,p)^\mathsf{T}X(q,l)^\mathsf{T}\right).$$

In particular, for $q \neq k$, the (k, q)-th block of $\Lambda(t)J_kX(t) + X(t)J_k\Lambda(t)$ is

$$-\sum_{1 \leq l \leq d+1, l \neq k,q} \left(X(k,l)\Lambda(l,q) + \Lambda(l,k)^\mathsf{T}X(q,l)^\mathsf{T}\right)$$

which is the same as the (k, q)-th block of $\Lambda(t)J_qX(t) + X(t)J_q\Lambda(t)$ if moreover $q \leq d$. Hence we have

$$\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_3(t)) = V(t)(Z_1(t), \ldots, Z_d(t))V(t)^\mathsf{T},$$

where for each $1 \leq k \leq d$,

$$Z_k(p,q) = \begin{cases} -\sum_{1 \leq l \leq d, l \neq k} \left(X(k,l)\Lambda(l,d+1) + \Lambda(l,k)^\mathsf{T} X(d+1,l)^\mathsf{T} \right), & \text{if } p = k, q = d+1, \\ -\sum_{1 \leq l \leq d, l \neq k} \left(X(d+1,l)\Lambda(l,k) + \Lambda(l,d+1)^\mathsf{T} X(k,l)^\mathsf{T} \right), & \text{if } p = d+1, q = k, \\ 0, & \text{otherwise.} \end{cases}$$

$\square$

Proposition A.3 (parallel transport along any curve). *Let $c(t) = V(t)(J_1, \ldots, J_d)V(t)^\mathsf{T}$ be a curve on $\mathrm{Flag}(n_1, \ldots, n_d; n)$ and let $Y(t)$ be a vector field along the curve $c(t)$, parametrized as in (57). Then $Y(t)$ is a parallel transport if and only for each pair (k,q) such that $1 \leq k < q \leq d+1$, we have*

$$-2\dot{X}(k,q) + \sum_{\substack{1 \leq s \leq d+1 \\ s \neq k, q}} (X(k,s)\Lambda(s,q) - \Lambda(k,s)X(s,q)) + \delta_{q,d+1} \sum_{\substack{1 \leq l \leq d \\ l \neq k}} (X(k,l)\Lambda(l,d+1) + \Lambda(k,l)X(l,d+1)) = 0, \quad (62)$$

which can be written in a more compact form:

$$2\dot{X} = \pi\left([X,\Lambda]\right) + \begin{bmatrix} 0 & X_0\Lambda_1 + \Lambda_0 X_1 \\ -(X_0\Lambda_1 + \Lambda_0 X_1)^\mathsf{T} & 0 \end{bmatrix}, \quad (63)$$

where $X_0, \Lambda_0 \in \mathbb{R}^{(n-n_{d+1}) \times (n-n_{d+1})}, X_1, \Lambda_1 \in \mathbb{R}^{(n-n_{d+1}) \times n_{d+1}}$ are submatrices determined by partitions

$$X = \begin{bmatrix} X_0 & X_1 \\ -X_1^\mathsf{T} & 0 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_0 & \Lambda_1 \\ -\Lambda_1^\mathsf{T} & 0 \end{bmatrix},$$

and $\pi(A)$ is defined by setting all diagonal blocks of $A \in \mathfrak{so}(n)$ equal to zero.

Before we proceed, we remark that in particular if $X = \Lambda$, then $[X,\Lambda] = 0$ and (63) reduces to the geodesic equation (33).

Next we re-write each term of (62) using (12). This leads to

$$\mathrm{vec}\left(X(k,s)\Lambda(s,q)\right) = (\Lambda(s,q)^\mathsf{T} \otimes \mathrm{I}_{m_k})\mathrm{vec}(X(k,s)) = -(\Lambda(q,s) \otimes \mathrm{I}_{m_k})\mathrm{vec}(X(k,s))$$

$$\mathrm{vec}\left(\Lambda(s,k)^\mathsf{T} X(q,s)^\mathsf{T}\right) = (\mathrm{I}_{m_q} \otimes \Lambda(s,k)^\mathsf{T})\mathrm{vec}(X(q,s)^\mathsf{T}) = (\mathrm{I}_{m_q} \otimes \Lambda(k,s))\mathrm{vec}(X(s,q))$$

and hence (62) becomes

$$2\mathrm{vec}\left(\dot{X}(k,q)\right) = -\sum_{\substack{1 \leq s \leq d+1 \\ s \neq k, q}} \left((\mathrm{I}_{m_q} \otimes \Lambda(k,s))\mathrm{vec}(X(s,q)) + (\Lambda(q,s) \otimes \mathrm{I}_{m_k})\mathrm{vec}(X(k,s)) \right)$$

$$+ \delta_{q,d+1} \sum_{\substack{1 \leq s \leq d \\ s \neq k}} \left((\mathrm{I}_{m_{d+1}} \otimes \Lambda(k,s))\mathrm{vec}(X(s,d+1)) - (\Lambda(d+1,s) \otimes \mathrm{I}_{m_k})\mathrm{vec}(X(k,s)) \right),$$

which using relations $X(p,q) = -X(q,p)^\mathsf{T}$ and $\mathrm{vec}(A^\mathsf{T}) = K^{(m,n)}\mathrm{vec}(A)$[1], can be written as

$$\begin{bmatrix} \mathrm{vec}(\dot{X}(1,2)) \\ \mathrm{vec}(\dot{X}(1,3)) \\ \cdots \\ \mathrm{vec}(\dot{X}(d-1,d)) \end{bmatrix} = \Phi(t) \begin{bmatrix} \mathrm{vec}(X(1,2)) \\ \mathrm{vec}(X(1,3)) \\ \cdots \\ \mathrm{vec}(X(d,d+1)) \end{bmatrix} \quad (64)$$

for some $\binom{n}{2} \times \binom{n}{2}$ matrix function $\Phi(t)$. Now according to Theorem 2.5, (64) can be solved by Peano–Baker series associated to the coefficient matrix $\Phi(t)$.

We again take $d = 2$ for example. In this case, we write

$$\Lambda(t) = \begin{bmatrix} 0 & A(t) & B(t) \\ -A^\mathsf{T}(t) & 0 & C(t) \\ -B^\mathsf{T}(t) & -C^\mathsf{T}(t) & 0 \end{bmatrix}, \quad A(t) \in \mathbb{R}^{m_1 \times m_2}, B(t) \in \mathbb{R}^{m_1 \times m_3}, C(t) \in \mathbb{R}^{m_2 \times m_3},$$

---

[1] $K^{(m,n)}$ is the commutation matrix defined in (13).

$$X(t) = \begin{bmatrix} 0 & U(t) & V(t) \\ -U^{\mathsf{T}}(t) & 0 & W(t) \\ -V^{\mathsf{T}}(t) & -W^{\mathsf{T}}(t) & 0 \end{bmatrix}, \quad U(t) \in \mathbb{R}^{m_1 \times m_2}, V(t) \in \mathbb{R}^{m_1 \times m_3}, W(t) \in \mathbb{R}^{m_2 \times m_3}.$$

Then we have

$$\text{proj}_{c(t)}^{\mathbb{T}}(T_1(t)) = 2V(t) \left(\begin{bmatrix} 0 & -\dot{U}(t) & -\dot{V}(t) \\ -\dot{U}^{\mathsf{T}}(t) & 0 & 0 \\ -\dot{V}^{\mathsf{T}}(t) & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \dot{U}(t) & 0 \\ \dot{U}^{\mathsf{T}}(t) & 0 & -\dot{W}(t) \\ 0 & -\dot{W}^{\mathsf{T}}(t) & 0 \end{bmatrix} \right) V(t)^{\mathsf{T}},$$

$$\text{proj}_{c(t)}^{\mathbb{T}}(T_3(t)) = V(t) \left(\begin{bmatrix} 0 & 0 & -(A(t)W(t) + U(t)C(t)) \\ 0 & 0 & 0 \\ -(A(t)W(t) + U(t)C(t))^{\mathsf{T}} & 0 & 0 \end{bmatrix}, \right.$$
$$\left. \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & A(t)^{\mathsf{T}}V(t) + U(t)^{\mathsf{T}}B(t) \\ 0 & V(t)^{\mathsf{T}}A(t) + B(t)^{\mathsf{T}}U(t) & 0 \end{bmatrix} \right) V(t)^{\mathsf{T}}.$$

$$\text{proj}_{c(t)}^{\mathbb{T}}(T_2(t)) = V(t) \left(\begin{bmatrix} 0 & B(t)W(t)^{\mathsf{T}} - V(t)C(t)^{\mathsf{T}} & -A(t)W(t) + U(t)C(t) \\ W(t)B(t)^{\mathsf{T}} - C(t)V(t)^{\mathsf{T}} & 0 & 0 \\ -W(t)^{\mathsf{T}}A(t)^{\mathsf{T}} + C(t)^{\mathsf{T}}U(t)^{\mathsf{T}} & 0 & 0 \end{bmatrix}, \right.$$
$$\left. \begin{bmatrix} 0 & -B(t)W(t)^{\mathsf{T}} + V(t)C(t)^{\mathsf{T}} & 0 \\ -W(t)B(t)^{\mathsf{T}} + C(t)V(t)^{\mathsf{T}} & 0 & A(t)^{\mathsf{T}}V(t) - U(t)^{\mathsf{T}}B(t) \\ 0 & V(t)^{\mathsf{T}}A(t) - B(t)^{\mathsf{T}}U(t) & 0 \end{bmatrix} \right) V(t)^{\mathsf{T}}.$$

Hence the system for $X(t)$ to be a parallel transport is given by

$$2\dot{U}(t) = -V(t)C(t)^{\mathsf{T}} + B(t)W(t)^{\mathsf{T}} = (-C(t) \otimes \mathrm{I}_{m_1})\,\text{vec}(V(t)) + (\mathrm{I}_{m_2} \otimes B(t))K^{(m_2,m_3)}\,\text{vec}(W(t)),$$
$$\dot{V}(t) = U(t)C(t) = (C(t)^{\mathsf{T}} \otimes \mathrm{I}_{m_1})\,\text{vec}(U(t)),$$
$$\dot{W}(t) = -U(t)^{\mathsf{T}}B(t) = -(B(t)^{\mathsf{T}} \otimes \mathrm{I}_{m_2})\,\text{vec}(U(t)).$$

Hence we have

$$\begin{bmatrix} \text{vec}(\dot{U}(t)) \\ \text{vec}(\dot{V}(t)) \\ \text{vec}(\dot{W}(t)) \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2}C(t) \otimes \mathrm{I}_{m_1} & \frac{1}{2}(\mathrm{I}_{m_2} \otimes B(t))K^{(m_2,m_3)} \\ C(t)^{\mathsf{T}} \otimes \mathrm{I}_{m_1} & 0 & 0 \\ -B(t) \otimes \mathrm{I}_{m_2} & 0 & 0 \end{bmatrix} \begin{bmatrix} \text{vec}(U(t)) \\ \text{vec}(V(t)) \\ \text{vec}(W(t)) \end{bmatrix}.$$

Appendix B. Parallel transport with respect to the modified embedding

Now we proceed to discuss the parallel transport of a tangent vector along a curve on a flag manifold. Again we parametrize a curve $c(t)$ on $\text{Flag}(n_1, \ldots, n_d; n)$ as (48). Let $Y(t)$ be a vector field on $\text{Flag}(n_1, \ldots, n_d; n)$ along the curve $c(t)$. Then we may correspondingly parametrize $Y(t)$ as

$$Y(t) = V(t)(X(t)J_1 - J_1X(t), \ldots, X(t)J_{d+1} - J_{d+1}X(t))V(t)^{\mathsf{T}}, \tag{65}$$

where $X(t) = (X(j,k))_{j,k=1}^{d+1,d+1} \in \mathfrak{so}(n)$ is the partition of $X(t)$ with respect to $n = m_1 + \cdots + m_{d+1}$ and $X(k,k)(t) \equiv 0, k = 1, \ldots, d+1$. We notice that $\dot{Y}(t) = T_1(t) + T_2(t) - T_3(t)$ where

$$T_1(t) = V(t)(\dot{X}(t)J_1 - J_1\dot{X}(t), \ldots, \dot{X}(t)J_{d+1} - J_{d+1}\dot{X}(t))V^{\mathsf{T}}(t), \tag{66}$$
$$T_2(t) = V(t)(\Lambda(t)X(t)J_1 + J_1X(t)\Lambda(t), \ldots, \Lambda(t)X(t)J_{d+1} + J_{d+1}X(t)\Lambda(t))V^{\mathsf{T}}(t), \tag{67}$$
$$T_3(t) = V(t)(\Lambda(t)J_1X(t) + X(t)J_1\Lambda(t), \ldots, \Lambda(t)J_{d+1}X(t) + X(t)J_{d+1}\Lambda(t))V^{\mathsf{T}}(t). \tag{68}$$

Similar computations in proofs of Lemmas A.1 and A.2 lead to the following

Lemma B.1. *Let* $c(t), \Lambda(t), Y(t), X(t), T_1(t), T_2(t), T_3(t)$ *be as above. We have*

(1) $T_1(t) \in \mathbb{T}_{c(t)} \text{Flag}(n_1, \ldots, n_d; n)$.

(2) $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_2(t)) = V(t)(W_1(t), \ldots, W_{d+1}(t))V^{\mathsf{T}}(t)$, where

$$W_k(p,q) = \begin{cases} \sum_{s=1}^{d+1} X(k,s)\Lambda(s,q) - \sum_{s=1}^{d+1} \Lambda(s,k)^{\mathsf{T}}X(q,s)^{\mathsf{T}}, & \text{if } p = k, q \neq k, \\ \sum_{s=1}^{d+1} \Lambda(s,p)^{\mathsf{T}}X(k,s)^{\mathsf{T}} - \sum_{s=1}^{d+1} X(p,s)\Lambda(s,k), & \text{if } q = k, p \neq k, \\ 0, & \text{otherwise.} \end{cases}$$

(3) $\mathrm{proj}_{c(t)}^{\mathbb{T}}(T_3(t)) = 0$.

Now we recall that $Y(t)$ is a parallel transport along the curve $c(t)$ if and only if $\mathrm{proj}_{c(t)}(\dot{Y}(t)) = 0$. Combining this with Lemma B.1, we may derive the equation for parallel transport.

Proposition B.2 (Parallel transport along any curve). *Let $c(t), \Lambda(t), Y(t), X(t)$ be as above. The vector field $Y(t)$ along $c(t)$ is a parallel transport if and only if*

$$\dot{X}(t) = \frac{1}{2}\pi\left([X(t), \Lambda(t)]\right), \tag{69}$$

where $\pi(A)$ is the matrix obtained by setting all diagonal blocks equal to zero for $A \in \mathfrak{so}(n)$.

In particular, if $c(t) = V(t)(J_1, \ldots, J_{d+1})V^{\mathsf{T}}(t)$ is a geodesic, then Proposition 4.6 implies $\dot{V}(t) = V(t)A$ for some constant matrix $A \in \mathfrak{so}(n)$, from which we obtain the following characterization of a parallel transport along a geodesic in $\mathrm{Flag}(n_1, \ldots, n_d; n)$.

Corollary B.3 (Parallel transport along a geodesic). *If $c(t) = V(t)(J_1, \ldots, J_{d+1})V^{\mathsf{T}}(t)$ is the geodesic curve passing through $V(0)(J_1, \ldots, J_{d+1})V^{\mathsf{T}}(0)$ with the tangent direction*

$$V(0)(AJ_1 - J_1A, \ldots, AJ_{d+1} - J_{d+1}A)V^{\mathsf{T}}(0), \quad A \in \mathfrak{so}(n), A(k,k) = 0, k = 1, \ldots, d+1,$$

then the parallel transport $Y(t)$ of

$$Y(0) = V(0)(BJ_1 - J_1B, \ldots, BJ_{d+1} - J_{d+1}B)V^{\mathsf{T}}(0), \quad B \in \mathfrak{so}(n), B(k,k) = 0, k = 1, \ldots, d+1,$$

is

$$Y(t) = V(t)(X(t)J_1 - J_1X(t), \ldots, X(t)J_{d+1} - J_{d+1}X(t))V^{\mathsf{T}}(t),$$

where

$$\dot{X}(t) = \frac{1}{2}\pi\left([X(t), A]\right), \quad X(0) = B,$$

where $\pi(A)$ is the matrix obtained by setting all diagonal blocks equal to zero for $A \in \mathfrak{so}(n)$.

Computational and Applied Mathematics Initiative, University of Chicago, Chicago, IL 60637-1514.
Email address: `laizehua@uchicago.edu, lekheng@uchicago.edu`

KLMM, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
Email address: `keyk@amss.ac.cn`