HIGH-ORDER RETRACTIONS ON MATRIX MANIFOLDS USING PROJECTED POLYNOMIALS

EVAN S. GAWLIK* AND MELVIN LEOK*

Abstract. We derive a family of high-order, structure-preserving approximations of the Riemannian exponential map on several matrix manifolds, including the group of unitary matrices, the Grassmannian manifold, and the Stiefel manifold. Our derivation is inspired by the observation that if Ω is a skew-Hermitian matrix and t is a sufficiently small scalar, then there exists a polynomial of degree n in $t\Omega$ (namely, a Bessel polynomial) whose polar decomposition delivers an approximation of $e^{t\Omega}$ with error $O(t^{2n+1})$. We prove this fact and then leverage it to derive high-order approximations of the Riemannian exponential map on the Grassmannian and Stiefel manifolds. Along the way, we derive related results concerning the supercloseness of the geometric and arithmetic means of unitary matrices.

1. Introduction. Approximating the Riemannian or Lie-theoretic exponential map on a matrix manifold is a task of importance in a variety of applications, including numerical integration on Lie groups [17, 23, 18, 5], optimization on manifolds [2, 8, 4, 30], interpolation of manifold-valued data [33, 34, 16, 13], rigid body simulation [5, 26], fluid simulation [15], and computer vision [36, 29, 11]. Often, special attention is paid to preserving the structure of the exponential map [6], which, for instance, should return a unitary matrix when the input Ω is skew-Hermitian. In this paper, we construct structure-preserving approximations to the Riemannian exponential map on matrix manifolds using *projected polynomials* – polynomial functions of matrices which, when projected onto a suitable set, deliver approximations to the Riemannian exponential with a desired order of accuracy. These projected polynomials can be thought of as high-order generalizations of the "projection-like retractions" considered in [3]. The matrix manifolds we consider are:

1. The group of unitary $m \times m$ matrices.
2. The Grassmannian manifold $Gr(p, m)$, which consists of all p-dimensional linear subspaces of $\mathbb{C}^m$, where $m \geq p$.
3. The Stiefel manifold $St(p, m) = \{Y \in \mathbb{C}^{m\times p} \mid Y^*Y = I\}$, where $m \geq p$.

The projector we use is to accomplish this task is the map which sends a full-rank matrix $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ to the nearest matrix with orthonormal columns. The latter matrix is precisely the factor U in the polar decomposition $A = UH$, where $U \in \mathbb{C}^{m\times p}$ has orthonormal columns and $H \in \mathbb{C}^{p\times p}$ is Hermitian positive-definite [9, Theorem 1]. In the case of the Grassmannian manifold, the QR decomposition can be used in place of the polar decomposition, leading to methods with very low computational cost.

Interestingly, in the case of the unitary group and the Grassmannian manifold, superconvergent approximations of the exponential are constructible with this approach. By this we mean that it is possible to construct polynomials of degree n that, upon projection, deliver approximations to the exponential with error of order $n + 2$ or higher. The appropriate choices of polynomials turn out to be intimately related to the Bessel polynomials, a well-known orthogonal sequence of polynomials [27], and the resulting approximations have error of order $2n + 1$; see Theorems 1 and 3 and Corollary 4.

One of the major advantages of this approach is that it delivers approximations

*Department of Mathematics, University of California, San Diego (egawlik@ucsd.edu, mleok@math.ucsd.edu).

to the exponential on the unitary group, the Grassmannian manifold, and the Steifel manifold that, to machine precision, have orthonormal columns. This is of obvious importance for the unitary group and the Stiefel manifold, and it is even desirable on the Grassmannian manifold, where it is common in computations to represent elements of the Grassmannian – p-dimensional subspaces of $\mathbb{C}^m$ – as $m \times p$ matrices whose columns form orthonormal bases for those subspaces [8].

Furthermore, when the polar decomposition is adopted as the projector, projected polynomials have the advantage that they can be computed using only rudimentary operations on matrices: matrix addition, multiplication, and inversion. This follows from the fact that the polar decomposition can be computed iteratively [20, Chapter 8]. Rudimentary algorithms for calculating the exponential on the Grassmannian and Stiefel manifolds are particularly desirable, since the most competitive existing algorithms for accomplishing this task involve singular value and/or eigenvalue decompositions [8, Theorem 2.1, Corollary 2.2, and Theorem 2.3], a feature that renders existing algorithms less ideally suited for parallel computation than projected polynomials.

In spite of these advantages, it is worth noting that not all of the constructions in this paper lead to algorithms that outshine their competitors. Diagonal Padé approximations of the exponential deliver, to machine precision, unitary approximations of e^{Ω} when Ω is skew-Hermitian [23, p. 97]. It is clear that the projected polynomials we present below (in Theorem 1) for approximating e^{Ω} are more expensive to compute, at least when the comparison is restricted to approximations of e^{Ω} with equal order of accuracy. On the Stiefel manifold, Padé approximation is not an option, rendering projected polynomials more attractive. However, they are not superconvergent on the Stiefel manifold; see Theorem 5. The setting in which projected polynomials appear to shine the brightest is the Grassmannian manifold $Gr(p, m)$, where they provide superconvergent, orthonormal approximations to the Riemannian exponential with algorithmic complexity $O(mp^2)$; see Theorem 3 and Corollary 4. To our knowledge, these are the first such approximations (other than the lowest-order versions) to appear in the literature on the Grassmannian manifold.

Structure-preserving approximations of the exponential map on matrix manifolds have a long history, particularly for matrix manifolds that form Lie groups. On Lie groups, techniques involving rational approximation [23, p. 97], splitting [6, 37], canonical coordinates of the second kind [7], and the generalized polar decomposition [24] have been studied, and many of these strategies lead to high-order approximations. For more general matrix manifolds like the Grassmannian and Stiefel manifolds, attention has been primarily restricted to methods for calculating the exponential exactly [8, 1, 2] or approximating it to low order [2, 3, 25, 10]. In this context, structure-preserving approximations of the exponential are commonly referred to as *retractions* [2, Definition 4.1.1]. High-order retractions on the Grassmannian and Stiefel manifolds have received very little attention, but there are good reasons to pursue them. For instance, in optimization, exactly evaluating the Riemannian Hessian of a function defined on a matrix manifold requires the use of a retraction with second-order accuracy or higher, at least if one is interested in its value away from critical points [2, p. 107]. In addition, existing algorithms for calculating the exponential on the Grassmannian and Stiefel manifolds exactly [8, Theorem 2.1, Corollary 2.2, and Theorem 2.3] are relatively expensive, which raises the question of whether more efficient options, perhaps with nonzero but controllable error, are available. On the Grassmannian manifold, the answer seems to be yes, at least for small-normed input; see Corollary 4.

Organization. This paper is organized as follows. In Section 2, we give statements of our results, deferring their proofs to Section 3. Our main results are Theorems 1, 3, and 5, which detail families of approximants to the exponential on the unitary group, the Grassmannian manifold, and the Steifel manifold, respectively. A fourth noteworthy result is Corollary 4, which provides a computationally inexpensive variant of the approximants in Theorem 3. We also detail two related results, Proposition 6 and Theorem 7, that concern the supercloseness of the geometric and arithmetic means of unitary matrices. In Section 3, we prove each of the results just mentioned. In Section 4, we describe algorithms for calculating our proposed approximations, with an emphasis on iterative methods for computing the polar decomposition. We conclude that section with numerical examples.

2. Statement of Results. In this section, we give statements of our results. Proofs are detailed in Section 3.

2.1. Exponentiation on the Unitary Group. Our first result deals with the approximation of the exponential of a skew-Hermitian matrix $\Omega \in \mathbb{C}^{m\times m}$ with projected polynomials. To motivate the forthcoming theorem, consider the Taylor polynomial $q_n(t\Omega)$ of degree n for $e^{t\Omega}$:

$$q_n(t\Omega) = \sum_{k=0}^{n} \frac{(t\Omega)^k}{k!}.$$

This quantity, in general, is not an unitary matrix, even though the matrix it aims to approximate, $e^{t\Omega}$, is unitary. If t is sufficiently small (small enough so that $q_n(t\Omega)$ is nonsingular), then $q_n(t\Omega)$ can be made unitary by computing the polar decomposition $q_n(t\Omega) = UH$, where $U \in \mathbb{C}^{m\times m}$ is unitary and $H \in \mathbb{C}^{m\times m}$ is Hermitian positive-definite. The matrix U is easily seen to be an unitary approximation to $e^{t\Omega}$ with error at worst $O(t^{n+1})$, owing to the fact that

$$\|U - q_n(t\Omega)\| \leq \|V - q_n(t\Omega)\|$$

for every unitary matrix $V \in \mathbb{C}^{m\times m}$, where $\|\cdot\|$ denotes the Frobenius norm [9, Theorem 1]. Indeed,

$$\begin{aligned}\|U - e^{t\Omega}\| &\leq \|U - q_n(t\Omega)\| + \|q_n(t\Omega) - e^{t\Omega}\| \\ &\leq \|e^{t\Omega} - q_n(t\Omega)\| + \|q_n(t\Omega) - e^{t\Omega}\| \\ &= O(t^{n+1}).\end{aligned}$$

Below we address the question of whether a better approximation to $e^{t\Omega}$ can be constructed by computing the unitary factor in the polar decomposition of

$$q_n(t\Omega) = \sum_{k=0}^{n} a_k t^k \Omega^k$$

for suitably chosen coefficients a_k. We show that if the coefficients a_k are chosen carefully, then an approximation with error of order t^{2n+1} can be constructed. The choice of coefficients a_k corresponds to the selection of a Bessel polynomial of degree n in $t\Omega$. In what follows, we use $\mathcal{P}$ to denote the map which sends a full-rank matrix $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ to the factor $\mathcal{P}(A) = U$ in its polar decomposition $A = UH$, where $U \in \mathbb{C}^{m\times p}$ has orthonormal columns and $H \in \mathbb{C}^{p\times p}$ is Hermitian positive-definite [20, Theorem 8.1].

THEOREM 1. *Let $\Omega \in \mathbb{C}^{m\times m}$ be skew-Hermitian, and let $n \geq 0$ be an integer. Define*

$$\Theta_n(z) = \sum_{k=0}^{n} \binom{n}{k} \frac{(2n-k)!}{(2n)!} (2z)^k. \tag{1}$$

Then

$$\mathcal{P}(\Theta_n(t\Omega)) = e^{t\Omega} + O(t^{2n+1}).$$

In fact, we will show that the polar decomposition of $\Theta_n(t\Omega)$ delivers the highest order approximation of $e^{t\Omega}$ among all polynomials in $t\Omega$ of degree n, up to rescaling. That is, if $r_n(t\Omega)$ is any other polynomial in $t\Omega$ of degree n satisfying $r_n(0) = I$, then

$$\mathcal{P}(r_n(t\Omega)) = e^{t\Omega} + O(t^k)$$

for some $0 \leq k \leq 2n$.

The polynomials (1) are scaled versions of Bessel polynomials [27]. More precisely, we have

$$\Theta_n(z) = \frac{2^n n!}{(2n)!} \theta_n(z),$$

where

$$\theta_n(z) = \sum_{k=0}^{n} \frac{(n+k)!}{(n-k)!k!} \frac{z^{n-k}}{2^k}$$

denotes the *reverse Bessel polynomial* of degree n. The first few polynomials $\Theta_n(z)$ are given by

$$\begin{aligned}
\Theta_0(z) &= 1,\\
\Theta_1(z) &= 1 + z,\\
\Theta_2(z) &= 1 + z + \frac{1}{3}z^2,\\
\Theta_3(z) &= 1 + z + \frac{2}{5}z^2 + \frac{1}{15}z^3,\\
\Theta_4(z) &= 1 + z + \frac{3}{7}z^2 + \frac{2}{21}z^3 + \frac{1}{105}z^4.
\end{aligned}$$

Note that, rather surprisingly, $\Theta_n(z)$ agrees with e^z only to first order for every $n \geq 1$.

2.2. Exponentiation on the Grassmannian. We now consider the task of approximating the Riemannian exponential map on the Grassmannian manifold $Gr(p, m)$, which consists of all p-dimensional subspaces of $\mathbb{C}^m$, where $m \geq p$. We begin by reviewing the geometry of $Gr(p, m)$, with an emphasis on computational aspects.

In computations, it is convenient to represent each subspace $\mathcal{V} \in Gr(p, m)$ with a matrix $Y \in \mathbb{C}^{m\times p}$ having orthonormal columns that span $\mathcal{V}$. The choice of Y is not unique, so we are of course thinking of Y as a representative of an equivalence class of $m \times p$ matrices sharing the same column space. With this identification, the tangent space to $Gr(p, m)$ at Y is given by

$$T_Y Gr(p, m) = \{Y_\perp K \mid K \in \mathbb{C}^{(m-p)\times p}\},$$

where $Y_\perp \in \mathbb{C}^{m\times(m-p)}$ is any matrix such that $\begin{pmatrix} Y & Y_\perp \end{pmatrix}$ is unitary, and $Y_\perp K$ is regarded as a representative of an equivalence class of matrices sharing the same

column space [8, p. 15]. With respect to the canonical metric on $Gr(p,m)$, the Riemannian exponential $\operatorname{Exp}_Y^{Gr} : T_Y Gr(p,m) \to Gr(p,m)$ at $Y \in \mathbb{C}^{m\times p}$ in the direction $H = Y_\perp K \in \mathbb{C}^{m\times p}$ is given by [8, p. 10]

$$\operatorname{Exp}_Y^{Gr} H = \begin{pmatrix} Y & Y_\perp \end{pmatrix} \exp \begin{pmatrix} 0 & -K^* \\ K & 0 \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix}. \tag{2}$$

The goal of this subsection is to construct computationally inexpensive approximations of $\operatorname{Exp}_Y^{Gr} H$. In order to be competitive with existing methods, such approximations must have computational complexity $O(mp^2)$ or better, owing to the following well-known result [8, Theorem 2.3].

THEOREM 2. *[8, Theorem 2.3] Let $H = U\Sigma V^*$ be the thin singular value decomposition of H, i.e. $U \in \mathbb{C}^{m\times p}$ has orthonormal columns, $\Sigma \in \mathbb{C}^{p\times p}$ is diagonal with nonnegative entries, and $V \in \mathbb{C}^{p\times p}$ is unitary. Then*

$$\operatorname{Exp}_Y^{Gr} H = YV \cos(\Sigma) V^* + U \sin(\Sigma) V^*.$$

The preceding theorem reveals that $\operatorname{Exp}_Y^{Gr} H$ can be computed exactly with $O(mp^2)$ operations, since this is the cost of computing the thin singular value decomposition of $H \in \mathbb{C}^{m\times p}$. With this in mind, we aim to derive approximations of $\operatorname{Exp}_Y^{Gr} H$ with smaller or comparable computational complexity.

Since the matrix $Z := \begin{pmatrix} 0 & -K^* \\ K & 0 \end{pmatrix}$ appearing in (2) is skew-Hermitian, an obvious option is to approximate $\exp Z$ in (2) with a projected polynomial $\mathcal{P}(\Theta_n(Z))$ in accordance with Section 2.1. This leads to approximants of the form

$$\operatorname{Exp}_Y^{Gr}(tH) = \begin{pmatrix} Y & Y_\perp \end{pmatrix} \mathcal{P}(\Theta_n(tZ)) \begin{pmatrix} I \\ 0 \end{pmatrix} + O(t^{2n+1}), \tag{3}$$

which, unfortunately, have computational complexity $O(m^3)$. Remarkably, we show in Lemma 15 below that

$$\mathcal{P}(\Theta_n(tZ)) \begin{pmatrix} I \\ 0 \end{pmatrix} = \mathcal{P}\left(\Theta_n(tZ) \begin{pmatrix} I \\ 0 \end{pmatrix} \right) \tag{4}$$

if t is sufficiently small (small enough so that $\Theta_n(tZ)$ is nonsingular). This is significant, since the right-hand side of this equality involves the polar decomposition of an $m \times p$ matrix $\Theta_n(tZ) \begin{pmatrix} I \\ 0 \end{pmatrix}$, which can be computed in $O(mp^2)$ operations. A few more algebraic manipulations (detailed in Section 3.4) lead to the following scheme for approximating the exponential on the Grassmannian in $O(mp^2)$ operations.

THEOREM 3. *Let $Y \in \mathbb{C}^{m\times p}$ have orthonormal columns, and let $H \in T_Y Gr(p,m)$. Then, for any $n \geq 0$,*

$$\operatorname{Exp}_Y^{Gr}(tH) = \mathcal{P}\left(Y\alpha_n(t^2 H^*H) + tH\beta_n(t^2 H^*H) \right) + O\left(t^{2n+1}\right),$$

where

$$\alpha_n(z) = \sum_{j=0}^{\lfloor n/2 \rfloor} a_{2j}(-z)^j,$$

$$\beta_n(z) = \sum_{j=0}^{\lfloor (n-1)/2 \rfloor} a_{2j+1}(-z)^j,$$

and

$$a_k = \binom{n}{k} \frac{(2n-k)!}{(2n)!} 2^k, \quad k = 0, 1, \ldots, n.$$

The first few nontrivial approximants provided by Theorem 3 read

$$\mathrm{Exp}_Y^{Gr}(tH) = \mathcal{P}\left(Y + tH\right) + O(t^3), \tag{5}$$

$$\mathrm{Exp}_Y^{Gr}(tH) = \mathcal{P}\left(Y\left(I - \frac{1}{3}t^2 H^*H\right) + tH\right) + O(t^5), \tag{6}$$

$$\mathrm{Exp}_Y^{Gr}(tH) = \mathcal{P}\left(Y\left(I - \frac{2}{5}t^2 H^*H\right) + tH\left(I - \frac{1}{15}t^2 H^*H\right)\right) + O(t^7), \tag{7}$$

$$\begin{aligned}\mathrm{Exp}_Y^{Gr}(tH) = \mathcal{P}\Bigg(&Y\left(I - \frac{3}{7}t^2 H^*H + \frac{1}{105}t^4 (H^*H)^2\right)\\ &+tH\left(I - \frac{2}{21}t^2 H^*H\right)\Bigg) + O(t^9).\end{aligned} \tag{8}$$

Note that, rather interestingly, the commonly used retraction $\mathcal{P}(Y + tH)$ (see, for instance, [1]) is in fact an approximation of $\mathrm{Exp}_Y^{Gr}(tH)$ with error $O(t^3)$, despite its appearance.

Note also that the approximants provided by Theorem 3 are rotationally equivariant. That is, if $V \in \mathbb{C}^{m\times m}$ is a unitary matrix, $\widetilde{Y} = VY$, and $\widetilde{H} = VH$, then $\widetilde{H}^*\widetilde{H} = H^*V^*VH = H^*H$ and hence

$$\begin{aligned}\mathcal{P}\left(\widetilde{Y}\alpha_n(t^2\widetilde{H}^*\widetilde{H}) + t\widetilde{H}\beta_n(t^2\widetilde{H}^*\widetilde{H})\right) &= \mathcal{P}\left(V\left(Y\alpha_n(t^2H^*H) + tH\beta_n(t^2H^*H)\right)\right)\\ &= V\mathcal{P}\left(Y\alpha_n(t^2H^*H) + tH\beta_n(t^2H^*H)\right),\end{aligned} \tag{9}$$

where the last line follows from the fact that

$$\mathcal{P}(VA) = V\mathcal{P}(A) \tag{10}$$

for every full-rank $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ and every unitary $V \in \mathbb{C}^{m\times m}$.

Replacing the Polar Decomposition with the QR Decomposition. An extraordinary feature of Theorem 3 is that it applies, with slight modification, even if the map $\mathcal{P}$ is replaced by the map $\mathcal{Q}$ which sends a full-rank matrix $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ to the factor Q in the QR decomposition $A = QR$, where $Q \in \mathbb{C}^{m\times p}$ has orthonormal columns and $R \in \mathbb{C}^{p\times p}$ is upper triangular. (The map $\mathcal{Q}$ is denoted qf in [2].) This follows from the fact that the columns of A, $\mathcal{P}(A)$, and $\mathcal{Q}(A)$ span the same space, so $\mathcal{P}(A)$ and $\mathcal{Q}(A)$ represent the same element of the Grassmannian manifold.

The only modification of Theorem 3 needed to make this idea precise is to use a genuine distance on the Grassmannian, such as

$$\mathrm{dist}^{Gr}(X, Y) = \min_{\substack{V,W\in\mathbb{C}^{p\times p}\\ V^*V=W^*W=I}} \|XV - YW\|,$$

to measure the distance between subspaces [8, p. 30]. We then have the following corollary.

COROLLARY 4. *Let $Y \in \mathbb{C}^{m\times p}$ have orthonormal columns, and let $H \in T_Y Gr(p, m)$. Then, for any $n \geq 0$,*

$$\mathrm{dist}^{Gr}(\mathcal{Q}\left(Y\alpha_n(t^2H^*H) + tH\beta_n(t^2H^*H)\right), \mathrm{Exp}_Y(tH)) = O(t^{2n+1}),$$

where α_n and β_n are given in the statement of Theorem 3.

This corollary is quite powerful, since the factor Q in the QR decomposition of an $m \times p$ matrix can be computed in merely $2mp^2 - \frac{2}{3}p^3$ operations [20, Appendix C], rendering the approximations $\mathcal{Q}\left(Y\alpha_n(t^2H^*H) + tH\beta_n(t^2H^*H)\right)$ very cheap to compute. Note also that these approximations are rotationally equivariant, by an argument similar to the one leading up to (9).

2.3. Exponentiation on the Stiefel Manifold. The next manifold we consider is the Stiefel manifold

$$St(p,m) = \{Y \in \mathbb{C}^{m\times p} \mid Y^*Y = I\}.$$

Unlike the Grassmannian, here we do not regard matrices $Y \in \mathbb{C}^{m\times p}$ as representatives of equivalence classes; each $Y \in \mathbb{C}^{m\times p}$ corresponds to a distinct element of $St(p,m)$. The tangent space to $St(p,m)$ at $Y \in St(p,m)$ is given by

$$T_Y St(p,m) = \{Y\Omega + Y_\perp K \mid \Omega = -\Omega^* \in \mathbb{C}^{p\times p}, K \in \mathbb{C}^{(m-p)\times p}\},$$

where $Y_\perp \in \mathbb{C}^{m\times(m-p)}$ is any matrix such that $\begin{pmatrix} Y & Y_\perp \end{pmatrix}$ is unitary [8, Equation 2.5]. With respect to the canonical metric on $St(p,m)$ [8, Section 2.4], the Riemannian exponential $\operatorname{Exp}_Y^{St} : T_Y St(p,m) \to St(p,m)$ at $Y \in \mathbb{C}^{m\times p}$ in the direction $H = Y\Omega + Y_\perp K \in \mathbb{C}^{m\times p}$ is given by [8, Equation 2.42]

$$\operatorname{Exp}_Y^{St} H = \begin{pmatrix} Y & Y_\perp \end{pmatrix} \exp \begin{pmatrix} \Omega & -K^* \\ K & 0 \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix}. \tag{11}$$

As an aside, we remark that a different exponential map is obtained if one endows $St(p,m)$ with the metric inherited from the embedding of $St(p,m)$ in Euclidean space [8, Section 2.2]. We do not consider the latter exponential map in this paper.

There exist algorithms for calculating (11) in $O(mp^2)$ operations, the simplest of which involves calculating the QR decomposition of a certain $m \times p$ matrix and then exponentiating a $2p \times 2p$ skew symmetric matrix [8, Corollary 2.2]. Our aim below is to derive a competitive algorithm for approximating (11) to high order using projected polynomials.

Before doing so, it is important to note that the right-hand side of (11) reduces to more familiar expressions in two special cases. First, when $Y^*H = \Omega = 0$, the right-hand side of (11) coincides with the right-hand side of (2), the Riemmannian exponential on the Grassmannian. On the other hand, if $m = p$ and $Y = I$, then K and $Y_\perp$ are empty matrices and the right-hand side of (11) reduces to e^Ω, the exponential of a skew-Hermitian matrix. Thus, in an effort to generalize Theorems 1 and 3, we seek to approximate (11) with projected polynomials that reduce to the ones appearing in Theorems 1 and 3 in those special cases.

In view of Theorem 3 and the identities $Y^*H = \Omega$ and $H^*H = -\Omega^2 + K^*K$, it is natural to consider approximations of (11) of the form

$$\operatorname{Exp}_Y^{St}(tH) \approx \mathcal{P}(Yq(t^2H^*H, tY^*H) + tHr(t^2H^*H, tY^*H)), \tag{12}$$

where $q(x,y)$ and $r(x,y)$ are polynomials in the (non-commuting) variables x and y. In order to ensure that these approximations recover those appearing in Theorems 1 and 3, we insist that:

(2.3.i) $q(x,0) = \alpha_n(x)$.
(2.3.ii) $r(x,0) = \beta_n(x)$.

(2.3.iii) $q(-x^2,x)$ and $r(-x^2,x)$ are polynomials of degree at most n and $n-1$, respectively, satisfying

$$q(-x^2,x)+xr(-x^2,x)=\Theta_n(x).$$

It turns out that such approximations can be constructed, but they lack the superconvergence enjoyed by the approximations in Theorems 1 and 3 (unless $\Omega=0$ or $m=p$). The difficulty becomes apparent if one compares $\mathcal{P}(Y+tH)$ with $\mathrm{Exp}_Y^{St}(tH)$ for generic $Y\in St(p,m)$ and $H=Y\Omega+Y_\perp K\in T_YSt(p,m)$. As $t\to 0$, one observes numerically that $\mathcal{P}(Y+tH)=\mathrm{Exp}_Y^{St}(tH)+O(t^2)$ (unless $\Omega=0$ or $m=p$).[1] This contrasts starkly with the situation in Theorems 1 and 3, where the polar decomposition of the first-order Taylor approximant of the exponential had superconvergent error $O(t^3)$. The following theorem confirms this observation and provides a couple of higher-order approximations of (11). In it, we use $\|\cdot\|$ to denote the Frobenius norm.

THEOREM 5. *Let $Y\in St(p,m)$ and $H\in T_YSt(p,m)$. Define*

$$\gamma_1(x,y)=1, \qquad \delta_1(x,y)=1,$$
$$\gamma_2(x,y)=1-\frac{1}{3}x-\frac{1}{2}y^2, \qquad \delta_2(x,y)=1+\frac{1}{2}y,$$
$$\gamma_3(x,y)=1-\frac{2}{5}x-\frac{1}{2}y^2-\frac{1}{6}y^3-\frac{1}{6}xy, \qquad \delta_3(x,y)=1-\frac{1}{15}x+\frac{1}{2}y.$$

For $n=1,2,3$, we have

$$\mathrm{Exp}_Y^{St}(tH)=\mathcal{P}\left(Y\gamma_n(t^2H^*H,tY^*H)+tH\delta_n(t^2H^*H,tY^*H)\right)+E, \tag{13}$$

where

$$E=\begin{cases}O(t^{2n+1}) & \text{if } Y^*H=0 \text{ or } m=p,\\ O(t^{n+1}) & \text{otherwise.}\end{cases} \tag{14}$$

In addition, for every polynomial $q(x,y)$ and $r(x,y)$ satisfying (2.3.i-2.3.iii) ($1\le n\le 3$), there exists $Y\in St(p,m)$, $H\in T_YSt(p,m)$, $C>0$, and $t_0>0$ such that

$$\left\|\mathrm{Exp}_Y^{St}(tH)-\mathcal{P}\left(Yq(t^2H^*H,tY^*H)+tHr(t^2H^*H,tY^*H)\right)\right\|\ge Ct^{n+1} \tag{15}$$

for every $t\le t_0$.

Written more explicitly, the approximants provided by Theorem 5 read

$$\mathrm{Exp}_Y^{St}(tH)\approx\mathcal{P}(Y+tH), \tag{16}$$

$$\mathrm{Exp}_Y^{St}(tH)\approx\mathcal{P}\left(Y\left(I-\frac{1}{3}t^2H^*H-\frac{1}{2}t^2(Y^*H)^2\right)+tH\left(I+\frac{1}{2}tY^*H\right)\right), \tag{17}$$

$$\mathrm{Exp}_Y^{St}(tH)\approx\mathcal{P}\left(Y\left(I+t^2\left(-\frac{2}{5}H^*H-\frac{1}{2}(Y^*H)^2\right)+t^3\left(-\frac{1}{6}H^*HY^*H-\frac{1}{6}(Y^*H)^3\right)\right)\right.$$
$$\left.+tH\left(I+\frac{1}{2}tY^*H-\frac{1}{15}t^2H^*H\right)\right). \tag{18}$$

[1] The astute reader may notice that this appears to contradict Theorem 4.9 of [3], which states, among other things, that projective retractions (see [3, Example 4.5]) are automatically second-order. However, it is not the exponential map (11) that $\mathcal{P}(Y+tH)$ approximates to second order. Rather, it is exponential map associated with the metric inherited from the embedding of $St(p,m)$ in $\mathbb{C}^{m\times p}$.

All of these are rotationally equivariant by an argument similar to the one leading up to (9).

Observe that when $Y^*H = \Omega = 0$, the right-hand sides of (16-18) reduce to those in (6-7), respectively. Likewise, when $m = p$ (so that $H = Y\Omega$ and $YY^* = I$), they reduce to $Y\mathcal{P}(\Theta_1(tH))$, $Y\mathcal{P}(\Theta_2(tH))$, and $Y\mathcal{P}(\Theta_3(tH))$, respectively, which are precisely the approximations of Ye^{Ω} provided by Theorem 1. See Section 3.5 for details.

In view of the complexity of (17) and (18), we do not believe that a general formula (valid for all n) can be conveniently written down for polynomials $q(x, y)$ and $r(x, y)$ satisfying (2.3.i-2.3.iii) that deliver approximations of (11) of the form (12) with error of optimal order. (As a matter of fact, the polynomials $\gamma_3(x, y)$ and $\delta_3(x, y)$ are not even uniquely determined by these conditions.) However, the proofs presented in Section 3.5 demonstrate how one can construct such polynomials.

Note that if the conditions (2.3.i-2.3.iii) are relaxed, then it is straightforward to construct approximations of (11) with error $O(t^{n+1})$: Simply truncate the Taylor series for $\exp\begin{pmatrix}\Omega & -K^*\\ K & 0\end{pmatrix}$, insert the result into (11), and express the result in terms of Y and H using the identities $H = Y\Omega + Y_\perp K$, $Y^*H = \Omega$, and $H^*H = -\Omega^2 + K^*K$. If desired, the result can be orthonormalized with the map $\mathcal{P}$, retaining the order of accuracy of the approximation. For this reason, Theorem 5 is less powerful than Theorems 1 and 3, and it underscores the complexity of the Stiefel manifold relative to the Grassmannian and the unitary group. For more evidence of the computational difficulties inherent to the Stiefel manifold, we refer the reader to [8].

2.4. Geometric and Arithmetic Means of Unitary Matrices. We conclude this section by stating two results that concern the supercloseness of certain means of unitary matrices. At the surface, these results might not appear to be closely related to Theorems 1, 3, 5, but in fact they follow from the same general theory.

Our first result reveals that the polar decomposition of the (componentwise) linear interpolant of two unitary matrices U_1 and U_2 is superclose to the geodesic joining U_1 and U_2.

PROPOSITION 6. *Let $U_1 \in \mathbb{C}^{m\times m}$ be unitary, let $\Omega \in \mathbb{C}^{m\times m}$ be skew-Hermitian, and let $U_2 = U_1e^{t\Omega}$. Assume that $(1-s)U_1 + sU_2$ is nonsingular for each $s \in [0, 1]$. Then*

$$\mathcal{P}((1-s)U_1 + sU_2) = U_1e^{st\Omega} + O(t^3)$$

for every $s \in [0, 1/2)\cup(1/2, 1]$. When $s = 1/2$, the equality $\mathcal{P}((1-s)U_1+sU_2) = U_1e^{st\Omega}$ holds exactly.

The case $s = 1/2$ in the preceding lemma recovers the well-known observation (see [21, Theorem 4.7] and [31, Equation (3.14)]) that the unitary factor $U = \mathcal{P}(\frac{1}{2}(U_1+U_2))$ in the polar decomposition of $\frac{1}{2}(U_1 + U_2)$ is given by

$$U = U_1e^{\frac{1}{2}\log(U_1^*U_2)} = U_1(U_1^*U_2)^{1/2}$$

whenever U_1 and U_2 are (non-antipodal) members of the unitary group.

Our second result of this subsection generalizes the preceding proposition in the following way. Let

$$\mathbb{A}(U_1, \ldots, U_n; w) = \operatorname*{arg\,min}_{\substack{V\in\mathbb{C}^{m\times m},\\ V^*V=I}} \sum_{i=1}^{n} w_i\|V - U_i\|^2$$

denote the weighted *arithmetic mean* [31, Definition 5.1] of unitary matrices $U_1, \dots, U_n \in \mathbb{C}^{m\times m}$, where $w \in \mathbb{R}^n$ is a vector of weights summing to 1. Let

$$\mathbb{G}(U_1, \dots, U_n; w) = \operatorname*{arg\,min}_{\substack{V\in\mathbb{C}^{m\times m},\\ V^*V=I}} \sum_{i=1}^{n} w_i \operatorname{dist}(V, U_i)^2$$

denote their weighted *geometric mean* [31, Definition 5.2], where

$$\operatorname{dist}(U, V) = \frac{1}{\sqrt{2}} \|\log(U^*V)\|$$

denotes the geodesic distance on the unitary group [31, Equation (2.6)]. It can be shown [12, Proposition 4] that $\mathbb{A}(U_1, \dots, U_n; w)$ exists whenever $\sum_{i=1}^n w_i U_i$ is nonsingular, and is given explicitly by

$$\mathbb{A}(U_1, \dots, U_n; w) = \mathcal{P}\left(\sum_{i=1}^{n} w_i U_i\right). \tag{19}$$

On the other hand, $\mathbb{G}(U_1, \dots, U_n; w)$ is characterized implicitly by the condition [31, p. 14]

$$\sum_{i=1}^{n} w_i \log(\mathbb{G}(U_1, \dots, U_n; w)^* U_i) = 0. \tag{20}$$

The following theorem reveals that if the data $U_1, \dots, U_n$ are nearby, then their weighted arithmetic and geometric means are superclose.

THEOREM 7. *Let $U_i : [0, T] \to \mathbb{C}^{m\times m}$, $i = 1, 2, \dots, n$, be continuous functions on $[0, T]$ such that $U_i(t)$ is unitary for each $t \in [0, T]$. Suppose that there exists $C > 0$ such that*

$$\operatorname{dist}(U_i(t), U_j(t)) \le Ct \tag{21}$$

for every $i, j = 1, 2, \dots, n$ and every $t \in [0, T]$. Then, for any $w \in \mathbb{R}^n$ with entries summing to 1,

$$\mathbb{A}(U_1(t), \dots, U_n(t); w) = \mathbb{G}(U_1(t), \dots, U_n(t); w) + O(t^3).$$

3. Proofs. In this section, we prove Theorems 1, 3, 5, and 7 and Proposition 6. Our proofs are structured as follows. In Section 3.1, we consider the general problem of estimating

$$\|\mathcal{P}(A) - \widetilde{U}\|, \tag{22}$$

where $A \in \mathbb{C}^{m\times p}$ $(m \ge p)$ is a full-rank matrix and $\widetilde{U} \in \mathbb{C}^{m\times p}$ has orthonormal columns. We show in Lemma 9 that this quantity can be estimated by measuring (1) the extent to which $\widetilde{U}^*A$ fails to be Hermitian and (2) the discrepancy between the range of A and the range of $\widetilde{U}$. We then leverage this lemma to prove Theorem 1 in Section 3.2, Theorem 3 in Section 3.4, Theorem 5 in Section 3.5, and Theorem 6 and Proposition 7 in Section 3.6.

It turns out that one of the theorems proved below, Theorem 1, admits an alternative proof that does not rely on Lemma 9. This alternative proof, which relies instead

on a relationship between projected polynomials and Padé approximation, is shorter than the one we present in Section 3.2, so we detail it in Section 3.3 for completeness. We have chosen to retain both proofs in this paper for several reasons. The proof in Section 3.2, despite being longer, highlights the versatility of Lemma 9, a lemma whose wide-ranging applicability is, in our opinion, one of the key contributions of this paper. The proof in Section 3.2 is also more elementary, in a certain sense, than that in Section 3.3, since the former relies merely on well-known perturbation estimates for the polar decomposition, whereas the latter relies on Padé approximation theory and certain results concerning the commutativity of functions of matrices.

3.1. Perturbations of the Polar Decomposition. We begin our examination of (22) by studying the sensitivity of the polar decomposition to perturbations. In what follows, we continue to use $\|\cdot\|$ to denote the Frobenius norm. We denote the i^{th} largest singular value of a matrix $A \in \mathbb{C}^{m\times p}$ by $\sigma_i(A)$. If $m = p$, we use $\rho(A)$ to denote the spectral radius of A. If furthermore A has real eigenvalues, we denote them by $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \lambda_p(A)$. Note that with this convention, it need not be true that $|\lambda_1(A)| \geq |\lambda_2(A)| \geq \ldots |\lambda_p(A)|$.

We denote by

$$\mathrm{sym}(A) = \frac{1}{2}(A + A^*)$$

and

$$\mathrm{skew}(A) = \frac{1}{2}(A - A^*)$$

the Hermitian and skew-Hermitian parts of a square matrix A, respectively. Note that since

$$\|A\| = \|A^*\|, \tag{23}$$

we have

$$\|\mathrm{sym}(A)\| \leq \|A\| \tag{24}$$

and

$$\|\mathrm{skew}(A)\| \leq \|A\| \tag{25}$$

for any square matrix A.

We will make use of the following additional properties of the Frobenius norm. For any $A \in \mathbb{C}^{m\times p}$, any $B \in \mathbb{C}^{p\times q}$, any $C \in \mathbb{C}^{p\times p}$, and any $U \in \mathbb{C}^{m\times p}$ with orthonormal columns:

(3.1.i) $\|A^*B\| \leq \|A^*\|\sigma_1(B) = \|A\|\sigma_1(B)$ [20, Equation (B.7)].
(3.1.ii) $\|U^*A\| \leq \|A\|$ (This follows from (3.1.i) and (23)).
(3.1.iii) $\|UC\| = \|C\|$ [20, Problem B.7].
(3.1.iv) $\rho(C) \leq \|C\|$ [20, Equation (B.8)].

Note that (3.1.i) is sharper than the estimate $\|A^*B\| \leq \|A^*\|\|B\|$, so we will frequently use (3.1.i) instead of the latter estimate.

We first recall a result concerning the stability of the polar decomposition under perturbations. A proof is given in [28].

LEMMA 8. *[28, Theorem 2.4] Let $A, \widetilde{A} \in \mathbb{C}^{m\times p}$ $(m \geq p)$ be full-rank matrices with polar decompositions $A = UH$ and $\widetilde{A} = \widetilde{U}\widetilde{H}$, where $U, \widetilde{U} \in \mathbb{C}^{m\times p}$ have orthonormal columns and $H, \widetilde{H} \in \mathbb{C}^{p\times p}$ are Hermitian positive-definite. Then*

$$\|U - \widetilde{U}\| \leq \frac{2}{\sigma_p(A) + \sigma_p(\widetilde{A})}\|A - \widetilde{A}\|.$$

Next, we consider a full-rank matrix $A \in \mathbb{C}^{m\times p}$ with polar decomposition $A = UH$, and we use the preceding lemma to show that the distance from U to any other matrix $\widetilde{U} \in \mathbb{C}^{m\times p}$ (sufficiently close to A) with orthonormal columns is controlled by two properties: (1) the extent to which $\widetilde{U}^*A$ fails to be Hermitian, and (2) the discrepancy between the range of A and the range of $\widetilde{U}$.

LEMMA 9. *Let $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ be a full-rank matrix with polar decomposition $A = UH$, where $U = \mathcal{P}(A) \in \mathbb{C}^{m\times p}$ has orthonormal columns and $H \in \mathbb{C}^{p\times p}$ is Hermitian positive-definite. Then, for any matrix $\widetilde{U} \in \mathbb{C}^{m\times p}$ with orthonormal columns satisfying $\|A - \widetilde{U}\| < 1$, we have*

(26)
$$\frac{\max\left\{2\|\mathrm{skew}(\widetilde{U}^*A)\|, \|(I - \widetilde{U}\widetilde{U}^*)A\|\right\}}{2\sigma_1(A)} \leq \|U - \widetilde{U}\| \leq \frac{2\left(\|\mathrm{skew}(\widetilde{U}^*A)\| + \|(I - \widetilde{U}\widetilde{U}^*)A\|\right)}{\sigma_p(A) + \sigma_p(\mathrm{sym}(\widetilde{U}^*A))}.$$

Proof. Define $\widetilde{H} = \mathrm{sym}(\widetilde{U}^*A)$. This matrix is positive-definite, since it is a small-normed perturbation of the identity matrix. Indeed, the relation

$$\widetilde{H} - I = \mathrm{sym}\left(\widetilde{U}^*(A - \widetilde{U})\right)$$

implies $\|\widetilde{H} - I\| \leq \|\widetilde{U}^*(A - \widetilde{U})\|$. Thus, using (3.1.ii) and (3.1.iv), the smallest eigenvalue of $\widetilde{H}$ satisfies

$$\begin{aligned}\lambda_p(\widetilde{H}) &= 1 + \lambda_p(\widetilde{H} - I)\\ &\geq 1 - \|\widetilde{H} - I\|\\ &\geq 1 - \|\widetilde{U}^*(A - \widetilde{U})\|\\ &\geq 1 - \|A - \widetilde{U}\|\\ &> 0.\end{aligned}$$

Now define $\widetilde{A} = \widetilde{U}\widetilde{H}$. Observe that

$$\begin{aligned}A - \widetilde{A} &= (\widetilde{U}\widetilde{U}^* + I - \widetilde{U}\widetilde{U}^*)A - \widetilde{U}\mathrm{sym}(\widetilde{U}^*A)\\ &= \widetilde{U}\mathrm{skew}(\widetilde{U}^*A) + (I - \widetilde{U}\widetilde{U}^*)A,\end{aligned}$$

so

$$\begin{aligned}\|A - \widetilde{A}\| &\leq \|\widetilde{U}\mathrm{skew}(\widetilde{U}^*A)\| + \|(I - \widetilde{U}\widetilde{U}^*)A\|\\ &= \|\mathrm{skew}(\widetilde{U}^*A)\| + \|(I - \widetilde{U}\widetilde{U}^*)A\|\end{aligned}$$

by (3.1.iii). The right-hand inequality in (26) then follows from Lemma 8 upon noting that $\widetilde{A}$ and $\mathrm{sym}(\widetilde{U}^*A)$ have the same singular values.

To prove the left-hand inequality in (26), observe that since $H = U^*A$ is Hermitian,

$$\mathrm{skew}(\widetilde{U}^*A) = \mathrm{skew}\left((\widetilde{U}^* - U^*)A\right).$$

Thus, using (3.1.i) and (25),

$$\begin{aligned}\|\mathrm{skew}(\widetilde{U}^*A)\| &\leq \|(\widetilde{U}^* - U^*)A\|\\ &\leq \|\widetilde{U}^* - U^*\|\sigma_1(A)\\ &= \|\widetilde{U} - U\|\sigma_1(A).\end{aligned} \tag{27}$$

On the other hand, since $UU^*A = UH = A$, we have

$$\begin{aligned}(I - \widetilde{U}\widetilde{U}^*)A &= (UU^* - \widetilde{U}\widetilde{U}^*)A \\ &= (U - \widetilde{U})U^*A + \widetilde{U}(U - \widetilde{U})^*A.\end{aligned}$$

Thus, using (3.1.i), (3.1.iii), and the fact that $\sigma_1(U^*A) = \sigma_1(A)$, it follows that

$$\begin{aligned}\|(I - \widetilde{U}\widetilde{U}^*)A\| &\leq \|U - \widetilde{U}\|\sigma_1(U^*A) + \|(U - \widetilde{U})^*A\| \\ &\leq \|U - \widetilde{U}\|\sigma_1(A) + \|U - \widetilde{U}\|\sigma_1(A) \\ &= 2\|U - \widetilde{U}\|\sigma_1(A). \qquad (28)\end{aligned}$$

Combining (27) and (28) proves the left-hand inequality in (26). ∎

The following less sharp version of Lemma 9, applicable in the square case ($m = p$), will be useful in the upcoming sections.

LEMMA 10. *Let A, U, and $\widetilde{U}$ be as in Lemma 9. If A is square (i.e. $m = p$), then*

$$\|A\|^{-1}\|\mathrm{skew}(\widetilde{U}^*A)\| \leq \|U - \widetilde{U}\| \leq 2\|A^{-1}\|\|\mathrm{skew}(\widetilde{U}^*A)\|.$$

Proof. Use (26) together with the fact that $\sigma_1(A) \leq \|A\|$, $\sigma_p(A)^{-1} \leq \|A^{-1}\|$, and $\widetilde{U}\widetilde{U}^* = I$ when $\widetilde{U}$ is square. ∎

3.2. Exponentiation on the Unitary Group. We now prove Theorem 1. Fix an integer $n \geq 0$ and consider a polynomial of the form

$$q_n(z) = \sum_{k=0}^{n} a_k z^k,$$

with $a_0 = 1$ and $a_k \in \mathbb{C}$, $k = 1, 2, \ldots, n$. We aim to find coefficients a_k making $\mathcal{P}(q_n(t\Omega)) - e^{t\Omega}$ small for any skew-Hermitian matrix Ω. Applying Lemma 10 with $A = q_n(t\Omega)$ and $\widetilde{U} = e^{t\Omega}$ gives

$$\|q_n(t\Omega)\|^{-1}\|\mathrm{skew}(e^{-t\Omega}q_n(t\Omega))\| \leq \|\mathcal{P}(q_n(t\Omega)) - e^{t\Omega}\| \leq 2\|q_n(t\\Omega)^{-1}\|\|\mathrm{skew}(e^{-t\Omega}q_n(t\Omega))\|, \qquad (29)$$

provided that t is sufficiently small (small enough that $q_n(t\Omega)$ has full rank and $\|q_n(t\Omega) - e^{t\Omega}\| < 1$).

This inequality is of great utility, since $\|q_n(t\Omega)\|^{-1}$ and $\|q_n(t\Omega)^{-1}\|$ are each $O(1)$ as $t \to 0$, and

$$\mathrm{skew}(e^{-t\Omega}q_n(t\Omega)) = \frac{1}{2}\left(e^{-t\Omega}q_n(t\Omega) - q_n(-t\Omega)e^{t\Omega}\right)$$

can be expanded in powers of t. Namely,

$$\begin{aligned}e^{-t\Omega}q_n(t\Omega) - q_n(-t\Omega)e^{t\Omega} &= \sum_{j=0}^{\infty}(-1)^j\frac{(t\Omega)^j}{j!}\sum_{k=0}^{n}a_k(t\Omega)^k - \sum_{k=0}^{n}(-1)^k a_k(t\Omega)^k\sum_{j=0}^{\infty}\frac{(t\Omega)^j}{j!} \\ &= \sum_{l=0}^{\infty} b_l t^l \Omega^l,\end{aligned}$$

where

$$b_l = \sum_{k=0}^{\min(l,n)} \frac{1}{(l-k)!}\left((-1)^{l-k} + (-1)^{k+1}\right) a_k$$
$$= \begin{cases} \sum_{k=0}^{\min(l,n)} \frac{2(-1)^{k+1}}{(l-k)!} a_k, & l \text{ odd}, \\ 0, & l \text{ even}. \end{cases}$$

The quantity $e^{-t\Omega} q_n(t\Omega) - q_n(-t\Omega)e^{t\Omega}$ is thus of the highest order in t when the n coefficients a_k, $k = 1, 2, \ldots, n$, are chosen to make $b_l = 0$ for $l = 1, 3, 5, \ldots, 2n-1$. This is achieved when

$$a_k = \binom{n}{k} \frac{(2n-k)!}{(2n)!} 2^k, \quad k = 1, 2, \ldots, n, \tag{30}$$

as the following lemma shows.

LEMMA 11. *With $a_0 = 1$ and a_k given by (30) for $1 \le k \le n$, we have*

$$\sum_{k=0}^{\min(l,n)} \frac{2(-1)^{k+1}}{(l-k)!} a_k = 0, \quad l = 1, 3, 5, \ldots, 2n-1. \tag{31}$$

Proof. Substitution gives

$$\sum_{k=0}^{\min(l,n)} \frac{2(-1)^{k+1}}{(l-k)!} a_k = \sum_{k=0}^{\min(l,n)} (-2)^{k+1} \binom{n}{k} \frac{(2n-k)!}{(2n)!(l-k)!}$$
$$= \frac{-2(n!)^2}{l!(2n)!} \sum_{k=0}^{\min(l,n)} (-2)^k \binom{l}{k} \binom{2n-k}{n},$$

so it suffices to show that

$$\sum_{k=0}^{\min(l,n)} (-2)^k \binom{l}{k} \binom{2n-k}{n} = 0$$

for each $l = 1, 3, 5, \ldots, 2n-1$. To prove this, consider the polynomial

$$r(z) = \sum_{k=0}^{2n} r_k z^k = (1+z)^{2n-l}(z-1)^l.$$

The coefficient of z^n in this polynomial is precisely

$$r_n = \sum_{k=0}^{\min(l,n)} (-2)^k \binom{l}{k} \binom{2n-k}{n}. \tag{32}$$

Indeed,

$$\begin{aligned} r(z) &= (1+z)^{2n-l}(z-1)^l \\ &= (1+z)^{2n-2l}\left(-2(1+z)+(1+z)^2\right)^l \\ &= (1+z)^{2n-2l}\sum_{k=0}^{l}\binom{l}{k}(-2)^k(1+z)^k(1+z)^{2(l-k)} \\ &= \sum_{k=0}^{l}\binom{l}{k}(-2)^k(1+z)^{2n-k} \\ &= \sum_{k=0}^{l}\binom{l}{k}(-2)^k\sum_{j=0}^{2n-k}\binom{2n-k}{j}z^j, \end{aligned}$$

and taking $j = n$ in the inner summation above gives (32). Now observe that $r(z)$ satisfies the symmetry

$$r(z) = (-1)^l z^{2n} r(z^{-1}).$$

From this it follows that the coefficients r_k satisfy $r_k = (-1)^l r_{2n-k}$, $k = 0, 1, \ldots, 2n$. In particular, $r_n = (-1)^l r_n$, so $r_n = 0$ when l is odd. ∎

This completes the proof of Theorem 1, since by the inequality (29), the unitary factor in the polar decomposition of the polynomial $\sum_{k=0}^{n} a_k t^k \Omega^k$ with coefficients given by (30) delivers an approximation of $e^{t\Omega}$ with error of order t^{2n+1}.

Uniqueness of the solution (30) to (31) is a consequence of the following lemma, which proves that the linear system (31) is nonsingular.

LEMMA 12. *Fixing $a_0 = 1$, the linear system (31) in the n unknowns $a_1, a_2, \ldots, a_n$ is nonsingular.*

Proof. Upon rearrangement, (31) reads

$$Mx = y,$$

where $x, y \in \mathbb{C}^n$ and $M \in \mathbb{C}^{n\times n}$ have entries given by

$$\begin{aligned} x_i &= a_i, \quad i = 1, 2, \ldots, n, \\ y_i &= \frac{2}{(2i-1)!}, \quad i = 1, 2, \ldots, n, \\ M_{ij} &= \begin{cases} \frac{2(-1)^{j+1}}{(2i-1-j)!}, & \text{if } j \le \min(2i-1, n), \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

An inductive argument shows that the determinant of M is equal to

$$\det M = \frac{(-1)^{\lfloor \frac{n}{2} \rfloor}(-2)^n}{\prod_{k=1}^{n-1}(2k-1)!!},$$

where $l!! = \prod_{j=0}^{\lceil l/2 \rceil - 1}(l-2j)$ denotes the double factorial. In particular, $\det M \neq 0$, showing that M is nonsingular. ∎

3.3. Connections with Padé Approximation. We now present an alternative proof of Theorem 1 that relies not on Lemma 9, but rather on a connection between $\mathcal{P}(\Theta_n(t\Omega))$ and the diagonal Padé approximant of $e^{2t\Omega}$.

Our alternative proof will make use of the fact that if $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ has full rank, then

$$\mathcal{P}(A) = A(A^*A)^{-1/2}, \tag{33}$$

where $C^{-1/2}$ denotes inverse of the principal square root of a square matrix C with no nonpositive real eigenvalues [20, Theorem 8.1].

It will also make use of the following facts: If f and g are two scalar-valued functions defined on the spectrum of a square matrix A, then $f(A)$ and $g(A)$ are well-defined [20, Section 1.2], $f(A)$ commutes with $g(A)$ [20, Theorem 1.13(e)], and the spectrum of $f(A)$ is the image of the spectrum of A under f [20, Theorem 1.13(d)].

LEMMA 13. *Let $q_n(z)$ be a polynomial of degree $n \geq 0$ with $q_n(0) \neq 0$, and let $\Omega \in \mathbb{C}^{m\times m}$ be skew-Hermitian. For each t sufficiently small (small enough so that $q_n(t\Omega)$ is nonsingular), we have*

$$\mathcal{P}(q_n(t\Omega))^2 = q_n(t\Omega)q_n(-t\Omega)^{-1}.$$

Furthermore, $\mathcal{P}(q_n(t\Omega))$ commutes with $e^{t\Omega}$.

Proof. Since Ω is skew-Hermitian, $q_n(t\Omega)^* = q_n(-t\Omega)$. Hence, by (33),

$$\mathcal{P}(q_n(t\Omega)) = q_n(t\Omega)\left[q_n(-t\Omega)q_n(t\Omega)\right]^{-1/2}.$$

This shows that $\mathcal{P}(q_n(t\Omega)) = f(t\Omega)$, where $f(z) = q_n(z)(q_n(-z)q_n(z))^{-1/2}$. If $q_n(t\Omega)$ is nonsingular, then f is defined on the spectrum of $t\Omega$. Indeed, if λ is an eigenvalue of $t\Omega$, then $q_n(\lambda)$, being an eigenvalue of $q_n(t\Omega)$, is nonzero, and $q_n(-\lambda)$, being an eigenvalue of $q_n(-t\Omega) = q_n(t\Omega)^*$, is nonzero. Thus, $\mathcal{P}(q_n(t\Omega))$ commutes with any function of $t\Omega$ defined on the spectrum of $t\Omega$, including $e^{t\Omega}$. By similar reasoning, $[q_n(-t\Omega)q_n(t\Omega)]^{-1/2}$ commutes with $q_n(t\Omega)$, so

$$\begin{aligned}\mathcal{P}(q_n(t\Omega))^2 &= q_n(t\Omega)\left[q_n(-t\Omega)q_n(t\Omega)\right]^{-1/2} q_n(t\Omega)\left[q_n(-t\Omega)q_n(t\Omega)\right]^{-1/2}\\ &= q_n(t\Omega)^2\left[q_n(-t\Omega)q_n(t\Omega)\right]^{-1}\\ &= q_n(t\Omega)q_n(-t\Omega)^{-1}.\end{aligned}$$

∎

The preceding lemma implies that if t is sufficiently small, then

$$(\mathcal{P}(q_n(t\Omega)) + e^{t\Omega})(\mathcal{P}(q_n(t\Omega)) - e^{t\Omega}) = q_n(t\Omega)q_n(-t\Omega)^{-1} - e^{2t\Omega}.$$

Using this identity, it is not hard to see that the polynomial q_n for which $\mathcal{P}(q_n(t\Omega)) - e^{t\Omega}$ is of the highest order in t is precisely that for which $q_n(t\Omega)q_n(-t\Omega)^{-1} - e^{2t\Omega}$ is of the highest order in t. That polynomial is none other than the numerator in the diagonal Padé approximant of $e^{2t\Omega}$, which is precisely $\Theta_n(t\Omega)$ [23, p. 97].

3.4. Exponentiation on the Grassmannian. We now prove Theorem 3. The proof will consist of two parts. First, we prove the identity (4) by exploiting the block structure of the matrix $Z = \begin{pmatrix} 0 & -K^* \\ K & 0 \end{pmatrix}$. Then, we insert the right-hand side of (4) into (3) and expand the result to obtain Theorem 3.

Throughout this subsection, we make use of the identities

$$Z^{2j} = \begin{pmatrix} (-K^*K)^j & 0 \\ 0 & (-KK^*)^j \end{pmatrix} \tag{34}$$

and

$$Z^{2j+1} = \begin{pmatrix} 0 & -K^*(-KK^*)^j \\ K(-K^*K)^j & 0 \end{pmatrix}, \tag{35}$$

which hold for for every nonnegative integer j.

LEMMA 14. *Let $r(z) = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n$ be a polynomial, let $K \in \mathbb{C}^{(m-p)\times p}$, and let $Z = \begin{pmatrix} 0 & -K^* \\ K & 0 \end{pmatrix}$. Then*

$$r(Z)^* r(Z) = \begin{pmatrix} B^*B & 0 \\ 0 & C^*C \end{pmatrix},$$

where $B = r(Z)\begin{pmatrix} I \\ 0 \end{pmatrix}$ and $C = r(Z)\begin{pmatrix} 0 \\ I \end{pmatrix}$.

Proof. The diagonal blocks of $r(Z)^*r(Z)$ are automatically given by B^*B and C^*C, so it suffices to show that the off-diagonal blocks of $r(Z)^*r(Z)$ vanish. To this end, observe that the skew-Hermiticity of Z implies $r(Z)^*r(Z) = r(-Z)r(Z)$. But $r(Z)^*r(Z)$ is Hermitian, so taking the Hermitian part of both sides gives $r(Z)^*r(Z) = \operatorname{sym}(r(-Z)r(Z))$. Since $\operatorname{sym}(Z^j) = 0$ for odd j, it follows that $r(Z)^*r(Z)$ is a linear combination of even powers of Z, all of which are block diagonal by (34). ∎

LEMMA 15. *Let $r(z) = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n$ be a polynomial, let $K \in \mathbb{C}^{(m-p)\times p}$, and define $Z = \begin{pmatrix} 0 & -K^* \\ K & 0 \end{pmatrix}$. If $r(Z)$ has full rank, then*

$$\mathcal{P}(r(Z))\begin{pmatrix} I \\ 0 \end{pmatrix} = \mathcal{P}\left(r(Z)\begin{pmatrix} I \\ 0 \end{pmatrix}\right).$$

Proof. In the notation of Lemma 14,

$$\begin{aligned} \mathcal{P}(r(Z)) &= r(Z)\left(r(Z)^*r(Z)\right)^{-1/2} \\ &= r(Z)\begin{pmatrix} (B^*B)^{-1/2} & 0 \\ 0 & (C^*C)^{-1/2} \end{pmatrix}, \end{aligned}$$

so

$$\mathcal{P}(r(Z))\begin{pmatrix} I \\ 0 \end{pmatrix} = r(Z)\begin{pmatrix} (B^*B)^{-1/2} \\ 0 \end{pmatrix}.$$

On the other hand,

$$\begin{aligned} \mathcal{P}\left(r(Z)\begin{pmatrix} I \\ 0 \end{pmatrix}\right) &= r(Z)\begin{pmatrix} I \\ 0 \end{pmatrix}\left(\begin{pmatrix} I & 0 \end{pmatrix} r(Z)^*r(Z)\begin{pmatrix} I \\ 0 \end{pmatrix}\right)^{-1/2} \\ &= r(Z)\begin{pmatrix} I \\ 0 \end{pmatrix}(B^*B)^{-1/2} \\ &= r(Z)\begin{pmatrix} (B^*B)^{-1/2} \\ 0 \end{pmatrix} \end{aligned}$$

as well. ∎

The preceding lemma establishes the identity (4). We now study the quantity $\Theta_n(tZ)\begin{pmatrix} I \\ 0 \end{pmatrix}$ in more detail.

LEMMA 16. *Let $\alpha_n(z)$ and $\beta_n(z)$ be as in Theorem 3, let $K \in \mathbb{C}^{(m-p)\times p}$, and let $Z = \begin{pmatrix} 0 & -K^* \\ K & 0 \end{pmatrix}$. Then*

$$\Theta_n(tZ)\begin{pmatrix} I \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha_n(t^2K^*K) \\ tK\beta_n(t^2K^*K) \end{pmatrix}.$$

Proof. Using (34-35), we have

$$\begin{aligned}
\Theta_n(tZ)\begin{pmatrix} I \\ 0 \end{pmatrix} &= \left(\sum_{j=0}^{\lfloor m/2 \rfloor} a_{2j}(tZ)^{2j} + \sum_{j=0}^{\lfloor (m-1)/2 \rfloor} a_{2j+1}(tZ)^{2j+1} \right) \begin{pmatrix} I \\ 0 \end{pmatrix} \\
&= \sum_{j=0}^{\lfloor m/2 \rfloor} a_{2j} \begin{pmatrix} (-t^2K^*K)^j \\ 0 \end{pmatrix} + \sum_{j=0}^{\lfloor (m-1)/2 \rfloor} a_{2j+1} \begin{pmatrix} 0 \\ tK(-t^2K^*K)^j \end{pmatrix} \\
&= \begin{pmatrix} \alpha_n(t^2K^*K) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ tK\beta_n(t^2K^*K) \end{pmatrix} \\
&= \begin{pmatrix} \alpha_n(t^2K^*K) \\ tK\beta_n(t^2K^*K) \end{pmatrix}.
\end{aligned}$$

∎

We are now in a position to prove Theorem 3 by substituting the preceding results into (3). Combining Lemmas 15 and 16, we have

$$\begin{aligned}
\begin{pmatrix} Y & Y_\perp \end{pmatrix} \mathcal{P}(\Theta_n(tZ)) \begin{pmatrix} I \\ 0 \end{pmatrix} &= \begin{pmatrix} Y & Y_\perp \end{pmatrix} \mathcal{P} \begin{pmatrix} \alpha_n(t^2K^*K) \\ tK\beta_n(t^2K^*K) \end{pmatrix} \\
&= \mathcal{P}\left(\begin{pmatrix} Y & Y_\perp \end{pmatrix} \begin{pmatrix} \alpha_n(t^2K^*K) \\ tK\beta_n(t^2K^*K) \end{pmatrix} \right) \\
&= \mathcal{P}\left(Y\alpha_n(t^2K^*K) + tY_\perp K\beta_n(t^2K^*K) \right) \\
&= \mathcal{P}\left(Y\alpha_n(t^2H^*H) + tH\beta_n(t^2H^*H) \right),
\end{aligned}$$

where the second line follows from (10), and the last line follows from the fact that $H = Y_\perp K$, and $Y_\perp$ has orthonormal columns. This, together with (3), completes the proof of Theorem 3.

3.5. Exponentiation on the Stiefel Manifold. We now turn to the proof of Theorem 5. Let $q(x,y)$ and $r(x,y)$ be polynomials in non-commutating variables x and y, and define

$$A = Yq\left(t^2H^*H, tY^*H\right) + tHr\left(t^2H^*H, tY^*H\right).$$

For the moment we assume only that $q(0,0) = 1$, but later we will make the additional assumptions (2.3.i-2.3.iii) (the first of which implies $q(0,0) = 1$). Using the identities $Y = \begin{pmatrix} Y & Y_\perp \end{pmatrix}\begin{pmatrix} I \\ 0 \end{pmatrix}$, $H = Y\Omega + Y_\perp K = \begin{pmatrix} Y & Y_\perp \end{pmatrix}\begin{pmatrix} \Omega \\ K \end{pmatrix}$, $H^*H = K^*K - \Omega^2$, and $Y^*H = \Omega$, we can write

$$\begin{aligned}
A &= \begin{pmatrix} Y & Y_\perp \end{pmatrix} \left(\begin{pmatrix} I \\ 0 \end{pmatrix} q\left(t^2(K^*K - \Omega^2), t\Omega\right) + \begin{pmatrix} t\Omega \\ tK \end{pmatrix} r\left(t^2(K^*K - \Omega^2, t\Omega)\right) \right) \\
&= \begin{pmatrix} Y & Y_\perp \end{pmatrix} \begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix},
\end{aligned}$$

where we have suppressed the arguments to q and r in the last line to reduce clutter.

Now let $Z = \begin{pmatrix} \Omega & -K^* \\ K & 0 \end{pmatrix}$ and define

$$\widetilde{U} = \begin{pmatrix} Y & Y_\perp \end{pmatrix} e^{tZ} \begin{pmatrix} I \\ 0 \end{pmatrix} = \mathrm{Exp}_Y^{St}(tH).$$

We aim to bound

$$\|\mathcal{P}(A) - \widetilde{U}\| = \|\mathcal{P}\left(Yq\left(t^2H^*H, tY^*H\right) + tHr\left(t^2H^*H, tY^*H\right)\right) - \mathrm{Exp}_Y^{St}(tH)\|$$

using Lemma 9. Since $A\big|_{t=0} = \begin{pmatrix} Y & Y_\perp \end{pmatrix}$ is unitary, it follows that $\sigma_i(A) = O(1)$ as $t \to 0$ for each $i = 1, 2, \ldots, p$. Thus, it is enough to bound $\|\mathrm{skew}(\widetilde{U}^*A)\|$ and $\|(I - \widetilde{U}\widetilde{U}^*)A\|$. We begin with a lemma.

LEMMA 17. *We have*

$$\|\mathrm{skew}(\widetilde{U}^*A)\| = \left\|\mathrm{skew}\left(\begin{pmatrix} I & 0 \end{pmatrix} e^{-tZ} \begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix}\right)\right\|,$$
$$\|(I - \widetilde{U}\widetilde{U}^*)A\| = \left\|\begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} e^{-tZ} \begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix}\right\|.$$

Proof. The first equality follows from a direct calculation, using the fact that Z is skew-Hermitian and $\begin{pmatrix} Y & Y_\perp \end{pmatrix}$ is unitary. For the second, observe that

$$\begin{aligned}(I - \widetilde{U}\widetilde{U}^*)A &= \left[I - \begin{pmatrix} Y & Y_\perp \end{pmatrix} e^{tZ} \begin{pmatrix} I \\ 0 \end{pmatrix} \begin{pmatrix} I & 0 \end{pmatrix} e^{-tZ} \begin{pmatrix} Y^* \\ Y_\perp^* \end{pmatrix}\right] \begin{pmatrix} Y & Y_\perp \end{pmatrix} \begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix} \\ &= \begin{pmatrix} Y & Y_\perp \end{pmatrix} e^{tZ} \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} e^{-tZ} \begin{pmatrix} Y^* \\ Y_\perp^* \end{pmatrix} \begin{pmatrix} Y & Y_\perp \end{pmatrix} \begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix} \\ &= \begin{pmatrix} Y & Y_\perp \end{pmatrix} e^{tZ} \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} e^{-tZ} \begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix}.\end{aligned}$$

The result follows from the fact that the Frobenius norm is unitarily invariant, and $\begin{pmatrix} Y & Y_\perp \end{pmatrix}$ and e^{tZ} are unitary. ∎

The preceding lemma reveals that the order of accuracy of the approximation (12) can be determined by studying the quantity

$$e^{-tZ} \begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix} = e^{-tZ} \begin{pmatrix} q\left(t^2(K^*K - \Omega^2), t\Omega\right) + t\Omega r\left(t^2(K^*K - \Omega^2), t\Omega\right) \\ tKr\left(t^2(K^*K - \Omega^2), t\Omega\right) \end{pmatrix}. \tag{36}$$

Let us carry out this task in order to determine, as an illustration, the highest order approximation of the form (12) that can be achieved using polynomials $q(x, y)$ and $r(x, y)$ satisfying (2.3.i-2.3.iii) with $n = 2$. The cases $n = 1$ and $n = 3$ are handled similarly; we leave those details to the reader. Collectively, these arguments will prove Theorem 5.

It is a simple exercise to show that when $n = 2$, the only polynomials $q(x, y)$ and $r(x, y)$ satisfying (2.3.i-2.3.ii) are of the form

$$q(x, y) = 1 - \frac{1}{3}x + cy^2,$$
$$r(x, y) = 1 - cy,$$

where c is a constant. Substituting into (36), writing $e^{-tZ} = \sum_{k=0}^{\infty} \frac{(-tZ)^k}{k!}$, and multiplying, one finds after a tedious calculation that

$$e^{-tZ}\begin{pmatrix} q + t\Omega r \\ tKr \end{pmatrix} = \begin{pmatrix} I + \frac{1}{6}t^2(K^*K - \Omega^2) - \left(c + \frac{1}{3}\right)t^3 K^*K\Omega \\ -\left(c + \frac{1}{2}\right)t^2 K\Omega \end{pmatrix} + O(t^4).$$

Hence, by Lemma 17 and the symmetry of K^*K and Ω^2, we have

$$\begin{aligned} \|\mathrm{skew}(\widetilde{U}^*A)\| &= \left|c + \frac{1}{3}\right| t^3 \|K^*K\Omega\| + O(t^4), \\ \|(I - \widetilde{U}\widetilde{U}^*)A\| &= \left|c + \frac{1}{2}\right| t^2 \|K\Omega\| + O(t^4). \end{aligned}$$

The optimal choice of c is $c = -\frac{1}{2}$, giving

$$\begin{aligned} q(x,y) &= 1 - \frac{1}{3}x - \frac{1}{2}y^2 = \gamma_2(x,y), \\ r(x,y) &= 1 + \frac{1}{2}y = \delta_2(x,y), \end{aligned}$$

and

$$\begin{aligned} \|\mathrm{skew}(\widetilde{U}^*A)\| &= \frac{1}{6}t^3 \|K^*K\Omega\| + O(t^4), \\ \|(I - \widetilde{U}\widetilde{U}^*)A\| &= O(t^4). \end{aligned}$$

It follows that

$$\mathrm{Exp}_Y^{St}(tH) = \mathcal{P}\left(Y\gamma_2(t^2H^*H, tY^*H) + tH\delta_2(t^2H^*H, tY^*H)\right) + O(t^3).$$

Clearly, no other choice of c will improve this approximant's order of accuracy, proving (15) for $n = 2$.

If it happens that $Y^*H = \Omega = 0$, then (2) and (11) coincide, and (2.3.i-2.3.ii) and Theorem 3 imply

$$\begin{aligned} \mathcal{P}\left(Y\gamma_2(t^2H^*H, 0) + tH\delta_2(t^2H^*H, 0)\right) &= \mathcal{P}\left(Y\alpha_2(t^2H^*H) + tH\beta_2(t^2H^*H)\right) \\ &= \mathrm{Exp}_Y^{Gr}(tH) + O(t^5) \\ &= \mathrm{Exp}_Y^{St}(tH) + O(t^5). \end{aligned}$$

Likewise, if $m = p$, so that $H = Y\Omega$, $YY^* = I$, and $\mathrm{Exp}_Y^{St}(tH) = Ye^{t\Omega}$, then (2.3.iii), (10), and Theorem 1 imply

$$\begin{aligned} &\mathcal{P}\left(Y\gamma_2(t^2H^*H, tY^*H) + tH\delta_2(t^2H^*H,\ tY^*H)\right) \\ &\qquad = \mathcal{P}\left(Y\gamma_2(-t^2\Omega^2, t\Omega) + tY\Omega\delta_2(-t^2\Omega^2, t\Omega)\right) \\ &\qquad = \mathcal{P}\left(Y\Theta_2(t\Omega)\right) \\ &\qquad = Y\mathcal{P}\left(\Theta_2(t\Omega)\right) \\ &\qquad = Ye^{t\Omega} + O(t^5) \\ &\qquad = \mathrm{Exp}_Y^{St}(tH) + O(t^5). \end{aligned}$$

These observations prove (13-14) for the case $n = 2$. The proof of Theorem 5 is completed by performing analogous arguments for the cases $n = 1$ and $n = 3$.

3.6. Geometric and Arithmetic Means of Unitary Matrices. We now prove Proposition 6 and Theorem 7.

Proof of Proposition 6. Without loss of generality, consider the case in which $U_1 = I$, so that

$$(1-s)U_1 + sU_2 = (1-s)I + se^{t\Omega}.$$

By Lemma 10, it suffices to examine the norm of

$$\operatorname{skew}\left(e^{-st\Omega}\left((1-s)I + se^{t\Omega}\right)\right).$$

The series expansion of $e^{-st\Omega}\left((1-s)I + se^{t\Omega}\right)$ reads

$$e^{-st\Omega}\left((1-s)I + se^{t\Omega}\right) = \sum_{k=0}^{\infty} \frac{(1-s)s^k(-1)^k + s(1-s)^k}{k!}(t\Omega)^k.$$

Since Ω is skew-Hermitian, $\operatorname{skew}(\Omega^k) = 0$ for every even k, showing that

$$\operatorname{skew}\left(e^{-st\Omega}\left((1-s)I + se^{t\Omega}\right)\right) = \sum_{\substack{k=1,\\ k \text{ odd}}}^{\infty} \frac{(1-s)s^k(-1)^k + s(1-s)^k}{k!}(t\Omega)^k.$$

When $s = 1/2$, each term in the series vanishes, giving

$$\operatorname{skew}\left(e^{-st\Omega}\left((1-s)I + se^{t\Omega}\right)\right)\Big|_{s=1/2} = 0.$$

When $s \neq 1/2$, the first non-vanishing term is of order t^3, showing that

$$\operatorname{skew}\left(e^{-st\Omega}\left((1-s)I + se^{t\Omega}\right)\right) = O(t^3).$$

The result follows by applying Lemma 10.

Proof of Theorem 7. Let

$$A(t) = \sum_{i=1}^{n} w_i U_i(t)$$

and

$$\widetilde{U}(t) = \mathbb{G}(U_1(t), \ldots, U_n(t); w).$$

Observe that if $A(t) = U(t)H(t)$ is the polar decomposition of $A(t)$, then, by (19),

$$U(t) = \mathbb{A}(U_1(t), \ldots, U_n(t); w).$$

Moreover, using the fact that $\sum_{i=1}^{n} w_i = 1$, we have

$$A(t) = U_1(t) + \sum_{i=2}^{n} w_i(U_i(t) - U_1(t)).$$

This shows, by (21), that

$$\lim_{t\to 0} A(t) = \lim_{t\to 0} U_1(t) = U_1(0).$$

The latter matrix is unitary, so $\|A(t)\| = O(1)$. This is independent of the $U_i(t)$ we chose to pull out of the sum, since (21) implies that $U_1(0) = U_2(0) = \cdots = U_n(0)$.

Now let $\Omega_i(t) = \frac{1}{t}\log(\widetilde{U}(t)^* U_i(t))$ for each i, so that

$$\widetilde{U}(t)^* U_i(t) = e^{t\Omega_i(t)}.$$

Note that $\Omega_i(t) = O(1)$ by (21). In addition, by (20),

$$\sum_{i=1}^{n} w_i \Omega_i(t) = 0.$$

Suppressing the dependencies on t for ease of reading, it follows that

$$\begin{aligned}
\widetilde{U}^* A &= \sum_{i=1}^{m} w_i \widetilde{U}^* U_i \\
&= \sum_{i=1}^{m} w_i e^{t\Omega_i} \\
&= \sum_{i=1}^{m} w_i I + \sum_{i=1}^{m} w_i t\Omega_i + \sum_{i=1}^{m} w_i \left(e^{t\Omega_i} - I - t\Omega_i\right) \\
&= I + \sum_{i=1}^{m} w_i \left(e^{t\Omega_i} - I - t\Omega_i\right).
\end{aligned}$$

The skew-Hermitian part of $\widetilde{U}^* A$ is thus given by

$$\operatorname{skew}(\widetilde{U}^* A) = \sum_{i=1}^{m} w_i \left(\frac{e^{t\Omega_i} - e^{-t\Omega_i}}{2} - t\Omega_i\right).$$

Since

$$\frac{e^{t\Omega_i} - e^{-t\Omega_i}}{2} - t\Omega_i = O(t^3)$$

and $\|A\| = O(1)$, it follows from Lemma 10 that

$$U - \widetilde{U} = O(t^3),$$

i.e.,

$$\mathbb{A}(U_1(t), \ldots, U_n(t); w) = \mathbb{G}(U_1(t), \ldots, U_n(t); w) + O(t^3).$$

4. Numerical Examples. In this section, we discuss how the projected polynomials proposed in Theorems 1, 3, and 5 can be efficiently computed, focusing on iterative methods for computing the polar decomposition. We then present numerical examples that illustrate their order of accuracy.

4.1. Iterative Methods for Computing the Polar Decomposition. The cost of computing a projected polynomial is largely dominated by the cost of evaluating the map $\mathcal{P}$. This map can be computed efficiently via a number of different iterative methods. The most widely known, applicable when $m = p$, is the Newton iteration

$$X_{k+1} = \frac{1}{2}(X_k + X_k^{-*}), \quad X_0 = A. \tag{37}$$

	$n=1$		$n=2$		$n=3$	
t_0/t	Error	Order	Error	Order	Error	Order
1	$1.607 \cdot 10^{0}$		$6.945 \cdot 10^{-2}$		$1.312 \cdot 10^{-3}$	
2	$2.433 \cdot 10^{-1}$	2.723	$2.444 \cdot 10^{-3}$	4.828	$1.109 \cdot 10^{-5}$	6.887
4	$3.223 \cdot 10^{-2}$	2.916	$7.860 \cdot 10^{-5}$	4.959	$8.830 \cdot 10^{-8}$	6.972
8	$4.091 \cdot 10^{-3}$	2.977	$2.474 \cdot 10^{-6}$	4.990	$6.932 \cdot 10^{-10}$	6.993

TABLE 1

Errors in approximating the exponential of a skew-Hermitian matrix Ω with the projected polynomials of Theorem 1. Shown above are the errors $\|\mathcal{P}(\Theta_n(t\Omega)) - e^{t\Omega}\|$ versus t for $n = 1, 2, 3$, where $t_0 = 0.01$ and Ω is a random 1000×1000 skew-Hermitian matrix.

	$n=1$		$n=2$		$n=3$	
t_0/t	Error	Order	Error	Order	Error	Order
1	$5.030 \cdot 10^{-1}$		$9.383 \cdot 10^{-3}$		$7.849 \cdot 10^{-5}$	
2	$6.951 \cdot 10^{-2}$	2.855	$3.090 \cdot 10^{-4}$	4.924	$6.352 \cdot 10^{-7}$	6.949
4	$8.933 \cdot 10^{-3}$	2.960	$9.781 \cdot 10^{-6}$	4.981	$5.006 \cdot 10^{-9}$	6.987
8	$1.125 \cdot 10^{-3}$	2.989	$3.066 \cdot 10^{-7}$	4.995	$3.919 \cdot 10^{-11}$	6.997

TABLE 2

Errors in approximating the Riemannian exponential map on the Grassmannian manifold with the projected polynomials of Theorem 3. Shown above are the errors $\|\mathcal{P}(Y\alpha_n(t^2H^*H) + tH\beta_n(t^2H^*H)) - \mathrm{Exp}_Y^{Gr}(tH)\|$ versus t for $n = 1, 2, 3$, where $t_0 = 0.01$, Y is a random 2000×400 matrix with orthonormal columns, and H is a random 2000×400 matrix satisfying $Y^*H = 0$.

The iterates X_k so defined converge quadratically to the unitary factor $\mathcal{P}(A) = U$ in the polar decomposition $A = UH$ for any nonsingular square matrix A [20, Theorem 8.12]. A closely related iteration, applicable when $m \geq p$, is given by

$$X_{k+1} = 2X_k(I + X_k^*X_k)^{-1}, \quad X_0 = A. \tag{38}$$

These iterates converge quadratically to $\mathcal{P}(A)$ for any full-rank $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ [20, Corollary 8.14(b)]. Finally, the Newton-Schulz iteration

$$X_{k+1} = \frac{1}{2}X_k(I - 3X_k^*X_k), \quad X_0 = A \tag{39}$$

provides an inverse-free iteration whose iterates X_k converge quadratically to $\mathcal{P}(A)$ for any $A \in \mathbb{C}^{m\times p}$ $(m \geq p)$ whose singular values all lie in the interval $(0, \sqrt{3})$ [20, Problem 8.20]. For further information, including other iterations for computing $\mathcal{P}(A)$, see [20, Chapter 8].

4.2. Numerical Convergence. We tested the accuracy of the projected polynomials detailed in Theorems 1, 3, and 5 by applying them to randomly generated inputs. To calculate $\mathcal{P}$, we used (37) for square matrices and (39) for rectangular matrices. The results of the tests, detailed in Tables 1-3, corroborate the convergence rates predicted by the theory.

5. Conclusion. This paper has presented a family of high-order retractions on the unitary group, the Grassmannian manifold, and the Stiefel manifold. All of these retractions were constructed by projecting certain matrix polynomials onto the set of matrices with orthonormal columns using the polar decomposition, or, in the case of the Grassmannian, using either the polar decomposition or the QR decomposition. There are several interesting applications and extensions of this strategy that seem worthwhile to pursue. On quadratic Lie groups other than the unitary group, one

	$n=1$		$n=2$		$n=3$	
t_0/t	Error	Order	Error	Order	Error	Order
1	$1.336 \cdot 10^{0}$		$1.416 \cdot 10^{-1}$		$1.654 \cdot 10^{-2}$	
2	$3.013 \cdot 10^{-1}$	2.149	$1.931 \cdot 10^{-2}$	2.874	$9.724 \cdot 10^{-4}$	4.089
4	$7.195 \cdot 10^{-2}$	2.066	$2.479 \cdot 10^{-3}$	2.961	$5.930 \cdot 10^{-5}$	4.035
8	$1.774 \cdot 10^{-2}$	2.020	$3.120 \cdot 10^{-4}$	2.990	$3.681 \cdot 10^{-6}$	4.010

TABLE 3

Errors in approximating the Riemannian exponential map on the Stiefel manifold with the projected polynomials of Theorem 5. Shown above are the errors $\|\mathcal{P}(Y\gamma_n(t^2H^*H, tY^*H) + tH\delta_n(t^2H^*H, tY^*H)) - \mathrm{Exp}_Y^{St}(tH)\|$ *versus* t *for* $n = 1, 2, 3$*, where* $t_0 = 0.01$*,* Y *is a random* 2000×400 *matrix with orthonormal columns, and* H *is a random* 2000×400 *matrix satisfying* $Y^*H = -H^*Y$*.*

might consider adopting the same strategy, replacing the polar decomposition with the generalized polar decomposition [32, 22]. It might also be worthwhile to consider projecting rational functions, rather than polynomials, to achieve higher accuracy for comparable cost. It may also be possible to leverage these retractions, together with methods for computing their derivatives [14], to construct high-order approximations of parallel transport operators on matrix manifolds; see [2, Section 8.1.2].

It is worth noting that many of the constructions in this paper might generalize nicely to infinite dimensions. For instance, replacing the matrices Y and H in Theorem 3 with quasi-matrices in the sense of [35], one obtains a method for approximating geodesics between finite-dimensional function spaces, with H^*H playing the role of a Gramian, and with $\mathcal{P}$ interpreted as the map sending an ordered basis of functions to the nearest ordered, orthonormal basis of functions.

6. Acknowledgements. We wish to thank the developers of the non-commutative algebra package *NCAlgebra* [19], which we used to carry out some of the calculations that appeared/were mentioned in Section 3.5. The first author is supported by NSF grants CMMI-1334759, DMS-1345013, and DMS-1703719. The second author is supported by NSF grants CMMI-1334759, DMS-1345013, DMS-1411792.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of grassmann manifolds with a view on algorithmic computation*, Acta Applicandae Mathematica, 80 (2004), pp. 199–220.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2009.

[3] P.-A. Absil and J. Malick, *Projection-like retractions on matrix manifolds*, SIAM Journal on Optimization, 22 (2012), pp. 135–158.

[4] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub, *Newton's method on riemannian manifolds and a geometric model for the human spine*, IMA Journal of Numerical Analysis, 22 (2002), pp. 359–390.

[5] N. Bou-Rabee and J. E. Marsden, *Hamilton–Pontryagin integrators on Lie groups part i: Introduction and structure-preserving properties*, Foundations of Computational Mathematics, 9 (2009), pp. 197–219.

[6] E. Celledoni and A. Iserles, *Approximating the exponential from a Lie algebra to a Lie group*, Mathematics of Computation, 69 (2000), pp. 1457–1480.

[7] E. Celledoni and A. Iserles, *Methods for the approximation of the matrix exponential in a Lie-algebraic setting*, IMA Journal of Numerical Analysis, 21 (2001), pp. 463–488.

[8] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM Journal on Matrix Analysis and Applications, 20 (1998), pp. 303–353.

[9] K. Fan and A. J. Hoffman, *Some metric inequalities in the space of matrices*, Proceedings of the American Mathematical Society, 6 (1955), pp. 111–116.

[10] S. Fiori, T. Kaneko, and T. Tanaka, *Tangent-bundle maps on the grassmann manifold:*

Application to empirical arithmetic averaging, IEEE Transactions on Signal Processing, 63 (2015), pp. 155–168.

[11] K. A. Gallivan, A. Srivastava, X. Liu, and P. Van Dooren, *Efficient algorithms for inferences on Grassmann manifolds*, in 2003 IEEE Workshop on Statistical Signal Processing, IEEE, 2003, pp. 315–318.

[12] E. S. Gawlik and M. Leok, *Embedding-based interpolation on the special orthogonal group*, (Preprint), (2016).

[13] E. S. Gawlik and M. Leok, *Interpolation on symmetric spaces via the generalized polar decomposition*, (Preprint), (2016).

[14] E. S. Gawlik and M. Leok, *Iterative computation of the Fréchet derivative of the polar decomposition*, (Preprint), (2016).

[15] E. S. Gawlik, P. Mullen, D. Pavlov, J. E. Marsden, and M. Desbrun, *Geometric, variational discretization of continuum theories*, Physica D: Nonlinear Phenomena, 240 (2011), pp. 1724–1760.

[16] P. Grohs, *Quasi-interpolation in riemannian manifolds*, IMA Journal of Numerical Analysis, 33 (2013), pp. 849–874.

[17] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, vol. 31, Springer Science & Business Media, 2006.

[18] J. Hall and M. Leok, *Lie group spectral variational integrators*, Foundations of Computational Mathematics, 17 (2017), pp. 199–257.

[19] J. W. Helton, M. C. de Oliveira, B. Miller, and M. Stankus, *The ncalgebra suite - version 5.0.* http://math.ucsd.edu/~ncalg/, 2017.

[20] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, 2008.

[21] N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur, *Functions preserving matrix groups and iterations for the matrix square root*, SIAM Journal on Matrix Analysis and Applications, 26 (2005), pp. 849–877.

[22] N. J. Higham, C. Mehl, and F. Tisseur, *The canonical generalized polar decomposition*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2163–2180.

[23] A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna, *Lie-group methods*, Acta Numerica 2000, 9 (2000), pp. 215–365.

[24] A. Iserles and A. Zanna, *Efficient computation of the matrix exponential by generalized polar decompositions*, SIAM Journal on Numerical Analysis, 42 (2005), pp. 2218–2256.

[25] T. Kaneko, S. Fiori, and T. Tanaka, *Empirical arithmetic averaging over the compact stiefel manifold*, IEEE Transactions on Signal Processing, 61 (2013), pp. 883–894.

[26] M. Kobilarov, K. Crane, and M. Desbrun, *Lie group integrators for animation and control of vehicles*, ACM Transactions on Graphics (TOG), 28 (2009), p. 16.

[27] H. L. Krall and O. Frink, *A new class of orthogonal polynomials: The Bessel polynomials*, Transactions of the American Mathematical Society, 65 (1949), pp. 100–115.

[28] W. Li and W. Sun, *Perturbation bounds of unitary and subunitary polar factors*, SIAM Journal on Matrix Analysis and Applications, 23 (2002), pp. 1183–1193.

[29] Y. M. Lui, *Advances in matrix manifolds for computer vision*, Image and Vision Computing, 30 (2012), pp. 380–388.

[30] E. Lundström and L. Eldén, *Adaptive eigenvalue computations using newton's method on the grassmann manifold*, SIAM Journal on Matrix Analysis and Applications, 23 (2002), pp. 819–839.

[31] M. Moakher, *Means and averaging in the group of rotations*, SIAM Journal on Matrix Analysis and Applications, 24 (2002), pp. 1–16.

[32] H. Z. Munthe-Kaas, G. Quispel, and A. Zanna, *Generalized polar decompositions on lie groups with involutive automorphisms*, Foundations of Computational Mathematics, 1 (2001), pp. 297–324.

[33] O. Sander, *Geodesic finite elements on simplicial grids*, International Journal for Numerical Methods in Engineering, 92 (2012), pp. 999–1025.

[34] O. Sander, *Geodesic finite elements of higher order*, IMA J. Numer. Anal., 36 (2016), pp. 238–266.

[35] A. Townsend and L. N. Trefethen, *Continuous analogues of matrix factorizations*, Proc. R. Soc. A, 471 (2015), p. 20140585.

[36] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, *Statistical computations on grassmann and stiefel manifolds for image and video-based recognition*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 33 (2011), pp. 2273–2286.

[37] H. Yoshida, *Construction of higher order symplectic integrators*, Physics Letters A, 150 (1990), pp. 262–268.