J. Austral. Math. Soc. (Series A) **42** (1987), 196–222

THE FINSLER GEOMETRY OF GROUPS OF ISOMETRIES OF HILBERT SPACE

C. J. ATKIN

(Received 17 June 1985)

Communicated by K. Mackenzie

Abstract

The paper deals with six groups: the unitary, orthogonal, symplectic, Fredholm unitary, special Fredholm orthogonal, and Fredholm symplectic groups of an infinite-dimensional Hilbert space. When each is furnished with the invariant Finsler structure induced by the operator-norm on the Lie algebra, it is shown that, between any two points of the group, there exists a geodesic realising this distance (often, indeed, a unique geodesic), except in the full orthogonal group, in which there are pairs of points that cannot be joined by minimising geodesics, and also pairs that cannot even be joined by minimising paths. A full description is given of each of these possibilities.

1980 *Mathematics subject classification* (*Amer. Math. Soc.*): primary 47 D 10; secondary 22 E 65.

This paper investigates the Finsler geometry (the metric, minimising paths, and minimising geodesics) of the orthogonal, unitary, and symplectic groups of Hilbert space, in the Finsler structure given by the operator-norm on the Lie algebra. The methods used derive ultimately from [8] (their extension to the unitary and symplectic cases was noted in [4], but not published).

Notations and definitions are given in Sections 1 and 2. Then in Sections 3 and 4 we discuss the relation between spectral theory and the metric. In Section 5, we review the results of [8]; this enables us to describe the Finsler metric explicitly in Section 6. The existence and uniqueness of minimising paths and minimising geodesics in the unitary and symplectic groups are settled in Section 7 (rather easily); the far more difficult case of the orthogonal group is resolved in Section 8.

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

1. Hilbert spaces: notations

For this material, see [1, Chapter 3], [2, Chapter 1], and [4].

(1.1) Throughout the paper, E denotes a complex Hilbert space with complex Hermitian inner product $\langle\ ,\ \rangle$. By a *conjugate-isometry* of E, we mean a conjugate-linear bijection J: $E \to E$ such that, for any x, $y \in E$,

$$\langle J(x), J(y)\rangle = \langle y, x\rangle.$$

(1.2) A real Hilbert space may be identified, by complexification, with a pair (E, J), where J is an involutive ($J^2 = I$) conjugate-isometry of the complex Hilbert space E. In effect, J is complex conjugation.

(1.3) A left-quaternionic Hilbert space is identified with a pair (E, J) of a complex Hilbert space E and a conjugate-isometry J which satisfies $J^2 = -I$. In effect, J is left-multiplication by the quaternion j. The quaternionic inner product $\langle\ ,\ \rangle_{\mathbf{H}}$ and scalar multiplication are given by

$$(\alpha + \beta i + \gamma j + \delta k)x = (\alpha + \beta i)x + (\gamma + \delta i)J(x),$$
$$\langle x, y\rangle_{\mathbf{H}} = \langle x, y\rangle + \langle x, J(y)\rangle j$$

(where $\alpha, \beta, \gamma, \delta$ are real, and i is both complex and quaternionic.)

(1.4) For uniformity, we shall also describe the pair (E, J) when $J = I$ as a complex Hilbert space. In each case, the (real, quaternionic or complex) subspaces of the original space correspond to the J-invariant complex subspaces of E, and the (real, quaternionic or complex) linear operators of the original space correspond to the complex linear operators of E which commute with J. We shall often speak of Hilbert spaces, subspaces, and operators, leaving further precision to the context. The correspondence preserves adjoints, operator-norms, and orthogonal complements. By the spectrum, eigenvalues, or eigenvectors of an operator, one means those of the associated complex operator in E.

(1.5) In a left-quaternionic Hilbert space, one may construct orthonormal bases (over the quaternions $\mathbf{H}$) in the usual way. For such a basis $\{e_\nu\colon \nu \in A\}$, and for any x in the space, there is a square-summable indexed class $\{\tau_\nu\colon \nu \in A\}$ in $\mathbf{H}$ such that $\sum \tau_\nu e_\nu$ sums unconditionally to x with respect to the topology of E. Define *right-multiplication* by $\sigma \in \mathbf{H}$ relative to the basis $\{e_\nu\}$ by

$$x \cdot \sigma = \sum (\tau_\nu \sigma) e_\nu.$$

The map $x \mapsto x \cdot \sigma$ is left-linear over $\mathbf{H}$, and so is an operator in (E, J). Its operator-norm is $|\sigma|$, and its adjoint is right-multiplication by $\bar{\sigma}$ relative to $\{e_\nu\}$.

(1.6) Let (E, J) be a Hilbert space. Let $\mathbf{L}(E, J)$ denote the real Banach algebra (in operator-norm) of its bounded operators, and $\mathbf{U}(E, J)$ the group of regular isometric operators in $\mathbf{L}(E, J)$; that is, $T \in \mathbf{U}(E, J)$ if and only if $T \in \mathbf{L}(E, J)$

and $T^*T = TT^* = I$. (The isometry group $\mathbf{U}(E, J)$ is called the *orthogonal*, *unitary*, or *symplectic group* of (E, J) in the real, complex, and quaternionic cases, respectively.) The subgroup of $\mathbf{U}(E, J)$ consisting of those isometric operators which differ from the identity only by a compact operator is called the *Fredholm isometry group*, and is denoted $\mathbf{UC}(E, J)$. Each of the six groups thus introduced is a closed subset in the appropriate $\mathbf{L}(E, J)$, and is a Banach Lie group (see [3], [5]). The Lie algebra of $\mathbf{U}(E, J)$ is naturally identified with the algebra $\mathbf{u}(E, J)$ of skew-adjoint operators in $\mathbf{L}(E, J)$, the Lie bracket being the commutator and the exponential map being operator exponentiation; and the Lie algebra of $\mathbf{UC}(E, J)$ is similarly identified with the algebra $\mathbf{uc}(E, J)$ of compact skew-adjoint operators.

2. The Finsler structure

(2.1) Let $\mathfrak{G}$ be a Banach Lie group, with identity element e. The tangent space at $x \in \mathfrak{G}$ is $T_x\mathfrak{G}$; left translation by $y \in \mathfrak{G}$ is L_y: $x \mapsto yx$, and T_xL_y: $T_x\mathfrak{G} \to T_{yx}\mathfrak{G}$ is its tangent map. The Lie algebra $\mathfrak{g}$ of $\mathfrak{G}$ is identified with $T_e\mathfrak{G}$. Now suppose $\|\cdot\|$ is a norm on $\mathfrak{g}$ defining the correct topology. Define a norm $\|\cdot\|_x$ on $T_x\mathfrak{G}$ by

$$(\forall \xi \in T_x\mathfrak{G})\|\xi\|_x = \left\|(T_eL_x)^{-1}\xi\right\|.$$

The norms $\|\cdot\|_x$ constitute a Finsler structure on $\mathfrak{G}$ (as in [7]), called the *left-invariant Finsler structure on* $\mathfrak{G}$ *induced by* $\|\cdot\|$. There is an analogous *right-invariant Finsler structure*, and the two agree if and only if the adjoint representation of $\mathfrak{G}$ on $\mathfrak{g}$ consists entirely of isometries with respect to $\|\cdot\|$.

(2.2) Henceforth, $\mathfrak{G}$ will denote either one of the groups of (1.6), or the principal component thereof. Its Lie algebra $\mathfrak{g}$ is identified with an algebra of operators as in (1.6), and $\mathfrak{G}$ is furnished with the left-invariant Finsler structure induced from the operator-norm on $\mathfrak{g}$ (which is also right-invariant, by (2.1)). This Finsler structure is a natural one to use, especially when E is of infinite dimension; in each group $\mathfrak{G}$, the operator-norm on $\mathfrak{g}$ may be characterised in terms of the algebraic structure of $\mathfrak{g}$.

The *Finsler metric* on $\mathfrak{G}$, denoted by d, is defined as follows: when x, y are in different components of G, let $d(x, y) = 2\pi$. Otherwise, let $d(x, y)$ be the minimum of 2π and the infimum of the Finsler lengths (see [7]) of the piecewise C^1 paths in G joining x and y. A path in $\mathfrak{G}$ is said to be *rectifiable* if it is rectifiable in the obvious sense with respect to d. Thus, a piecewise C^1 path is rectifiable. Any rectifiable path p has a well-defined length $\ell(p)$ which is not less than the distance between its ends; when p is C^1, then $\ell(p)$ is the same as the Finsler length of p.

A path in $\mathfrak{G}$ is, by inclusion, equally a path in $\mathbf{L}(E, I)$. If it is rectifiable in $\mathfrak{G}$, it is also rectifiable in $\mathbf{L}(E, I)$, and its length in $\mathfrak{G}$ is the same as its length in $\mathbf{L}(E, I)$.

(2.3) By a *geodesic* in $\mathfrak{G}$, I mean a geodesic of any left-invariant connection on $\mathfrak{G}$: that is, a left translate of a continuous one-parameter subgroup (which is, therefore, a geodesic of any right-invariant connection too). A *minimising path* between points x, $y \in \mathfrak{G}$ is a rectifiable path of length $d(x, y)$ between these points; a *minimising geodesic* between x and y is a geodesic which is a minimising path between x and y. These definitions do not presuppose any explicit relation between the connections and the Finsler structure.

A geodesic in $\mathfrak{G}$ is automatically parametrised proportionally to arc length, and any rectifiable path may be so reparametrised; then we shall describe it as *uniformly parametrised*.

3. Spectral theory

(3.1) Let $T \in \mathbf{L}(E, I)$ be normal (see (1.6)). We write $\sigma(T)$ for the spectrum of T, and $\sigma'(T)$ for the essential spectrum of T: that is, the set of points of $\sigma(T)$ which are either cluster points of $\sigma(T)$, or are isolated in $\sigma(T)$ but are eigenvalues of infinite multiplicity. Evidently $\sigma'(T)$ is null if E is of finite dimension.

(3.2) LEMMA. *Let E be of infinite dimension*; *let T be a bounded normal operator on E, with associated spectral measure $\mathbf{P}_T$. Then*

(a) $\sigma'(T)$ *is nonnull and compact*;

(b) $\mu \in \sigma'(T)$ *if and only if*, *for every positive β, the spectral projection $\mathbf{P}_T(\{z \in \mathbf{C}\colon |z - \mu| \leqslant \beta\})$ is of infinite rank.*

PROOF. (b) is clear, and $\sigma'(T)$ is manifestly a closed subset of the compact set $\sigma(T)$. If $\sigma'(T) = \varnothing$, then $\sigma(T)$ consists only of isolated points which are eigenvalues of finite multiplicity, and, by compactness, there are at most finitely many. Therefore E is of finite dimension, which contradicts the hypothesis.

(3.3) The following notations will be standard: S will denote the unit circle $\{z \in \mathbf{C}\colon |z| = 1\}$. The metric δ on S will be the Finsler metric of $\mathbf{U}(\mathbf{C}, I)$ (see (2.2)), which is given explicitly by

$$\delta(\exp(i\theta), \exp(i\varphi)) = \min\{|\theta - \varphi + 2n\pi|\colon n \in \mathbf{Z}\}.$$

If $T \in \mathbf{L}(E, I)$ is normal, then the associated spectral measure will be denoted by $\mathbf{P}_T$, unless some other symbol is specially introduced. If $T \in \mathbf{U}(E, I)$ (and, in particular, if T belongs to any of the groups $\mathfrak{G}$ of (2.2)), then $\sigma(T) \subseteq S$. Define, in that case,

$$N(T) = \sup\{|\theta| : \theta \in \mathbf{R}, |\theta| \leqslant \pi, \exp(i\theta) \in \sigma(T)\},$$
$$N'(T) = \sup\{|\theta| : \theta \in \mathbf{R}, |\theta| \leqslant \pi, \exp(i\theta) \in \sigma'(T)\}.$$

For $z \in S$, denote by $\log z$ that value of the logarithm which lies in the interval $(-i\pi, i\pi]$ of the imaginary axis. Thus $\cos(\log z) = \operatorname{Re} z$ for any $z \in S$.

(3.4) LEMMA. *Let $T \in \mathrm{U}(E, I)$, and suppose that F is a T-invariant subspace of E (so that $F^\perp$ is also T-invariant). Then*

$$N(T) = \max\left(N(T \mid F), N(T \mid F^\perp)\right)$$

and

$$N'(T) = \max\left(N'(T \mid F), N'(T \mid F^\perp)\right).$$

PROOF. Indeed, $\sigma(T) = \sigma(T \mid F) \cup \sigma(T \mid F^\perp)$ and $\sigma'(T) = \sigma'(T \mid F) \cup \sigma'(T \mid F^\perp)$.

4. Transformations of the sphere

(4.1) The Hilbert space E over $\mathbf{C}$ becomes a Hilbert space over $\mathbf{R}$ if one takes the real (Euclidean) inner product to be the real part of the complex (Hermitian) inner product given in E. The unit sphere Σ of E, which is the same with respect to either inner product, is a closed C^ω submanifold of E, and so inherits a C^ω Riemannian structure from the real inner product in E. The Riemannian manifold Σ has the following properties, all of which are intuitively obvious and straightforward to prove in an ad hoc fashion.

(a) If x, $y \in \Sigma$, then the Riemannian distance $D(x, y)$ between x and y in Σ is $\cos^{-1}(\operatorname{Re}\langle x, y\rangle) \in [0, \pi]$.

(b) All geodesics in Σ are uniformly parametrised great circles (intersections of Σ with real two-dimensional subspaces of E). In particular, they are all periodic.

(c) All uniformly parametrised minimising paths in Σ are geodesics.

(d) If x, $y \in \Sigma$, and if $x + y \neq 0$ (that is, x and y are not diametrically opposite), then x and y may be joined, except for linear changes of parameter, by exactly two geodesic arcs (injective geodesics), of which one and one only is minimising. If $x + y = 0$, there are infinitely many geodesic arcs (not identifiable by change of parameter) from x to y in Σ, unless E is of complex dimension 1; and all of them are minimising.

The notations Σ and D will be maintained.

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

(4.2) LEMMA. *Suppose* $T \in \mathbf{U}(E, I)$, $\mu \in S$, *and* $\gamma \geqslant 0$; *let* H *denote the image of* $\mathbf{P}_T(\{z \in S\colon \delta(z, \mu) \leqslant \gamma\})$. *Then*

(a) *if* $x \in H^\perp \cap \Sigma$, $D(Tx, \mu x) > \gamma$;

(b) *if* $x \in H \cap \Sigma$, $D(Tx, \mu x) \leqslant \gamma$.

Furthermore, equality holds if and only if x *lies in the image of* $\mathbf{P}_T(\{z \in S\colon \delta(z, \mu) = \gamma\})$.

PROOF. If $\gamma \geqslant \pi$, then $H = E$ and there is nothing to prove. So assume that $\gamma < \pi$. By (4.1)(a),

$$\cos D(Tx, \mu x) = \operatorname{Re}\langle Tx, \mu x\rangle = \operatorname{Re}\int_S z\langle \mathbf{P}_T(dz)x, \mu x\rangle \tag{1}$$

$$= \operatorname{Re}\int_S z\bar{\mu}\langle \mathbf{P}_T(dz)x, x\rangle$$

$$= \int_S \operatorname{Re}(z\bar{\mu})\langle \mathbf{P}_T(dz)x, x\rangle,$$

since $\langle \mathbf{P}_T(dz)x, x\rangle$ is a nonnegative real-valued measure. It is also of total mass 1, and, in case (a), is concentrated on $\{z \in S\colon \delta(z, \mu) > \gamma\}$, where $\operatorname{Re}(z\bar{\mu}) < \cos\gamma$. Hence, in case (a),

$$\cos D(Tx, \mu x) < \cos\gamma.$$

But $\gamma < \pi$ by assumption, and $D(Tx, \mu x) \leqslant \pi$ by (4.1)(a). Thus (a) follows. Similarly, in case (b), the measure of (1) is concentrated on $\{z \in S\colon \delta(z, \mu) \leqslant \gamma\}$, where $\operatorname{Re}(z\bar{\mu}) \geqslant \cos\gamma$. So $\cos D(Tx, \mu x) \geqslant \cos\gamma$; equality will occur if and only if the measure is concentrated on the set where the integrand of (1) takes the value $\cos\gamma$, namely the set $\{z \in S\colon \delta(z, \mu) = \gamma\}$. This is equivalent to saying that x lies in the image of $\mathbf{P}_T(\{z \in S\colon \delta(z, \mu) = \gamma\})$. The remaining assertions now follow, as (a) did previously.

By the same method (or by taking $\pi - \gamma$ in place of γ and orthogonal complements) one obtains

(4.3) LEMMA. *Suppose* $T \in \mathbf{U}(E, I)$, $\mu \in S$, *and* $\gamma > 0$; *let* H *denote the image of* $\mathbf{P}_T(\{z \in S\colon \delta(z, \mu) < \gamma\})$. *Then*

(a) *if* $x \in H \cap \Sigma$, $D(Tx, \mu x) < \gamma$;

(b) *if* $x \in H^\perp \cap \Sigma$, $D(Tx, \mu x) \geqslant \gamma$.

Furthermore, equality holds if and only if x *lies in the image of* $\mathbf{P}_T(\{z \in S\colon \delta(z, \mu) = \gamma\})$.

(4.4) LEMMA. *Let* $T \in \mathbf{U}(E, I)$. *Then (see* (3.3))

$$N(T) = \sup\{D(Tx, x)\colon x \in \Sigma\}.$$

If $y \in \Sigma$, *then* $N(T) = D(Ty, y)$ *if and only if* y *lies in the image of* $\mathbf{P}_T(\{\exp(iN(T)), \exp(-iN(T))\})$.

PROOF. By definition (3.3), $\sigma(T) \subseteq \{z \in S\colon \delta(z,1) \leqslant N(T)\}$, so $\mathbf{P}_T(\{z \in S\colon \delta(z,1) \leqslant N(T)\}) = I$. Take $\mu = 1$, $\gamma = N(T)$ in (4.2)(b); hence, for any $x \in \Sigma$,

$$D(Tx, x) \leqslant N(T), \tag{1}$$

with equality if and only if x lies in the image of $\mathbf{P}_T(\{z \in S\colon \delta(z,1) = N(T)\})$, which is as asserted.

Now suppose that $\varepsilon > 0$, and set

$$Z = \{z \in S\colon \delta(z,1) > N(T) - \varepsilon\}.$$

Then $\mathbf{P}_T(Z) \neq 0$ by the definition (3.3) of $N(T)$, and

$$\mathbf{P}_T(Z)E = \left(\mathbf{P}_T(\{z \in S\colon \delta(z,1) \leqslant N(T) - \varepsilon\})E\right)^{\perp}.$$

Take $y \in \Sigma \cap \mathbf{P}_T(Z)E$. For this y, (4.2)(a) gives

$$D(Ty, y) > N(T) - \varepsilon. \tag{2}$$

Together, (1) and (2) prove the lemma's first assertion.

(4.5) COROLLARY. *For any* T, $U \in \mathbf{U}(E, I)$,

$$N(TU) \leqslant N(T) + N(U).$$

PROOF. For any $x \in \Sigma$,

$$D(TUx, x) \leqslant D(TUx, Ux) + D(Ux, x),$$

and the result follows from (4.4).

(4.6) LEMMA. *Let* $\mathfrak{G}$ *be as in* (2.3), *and suppose that* p *is a rectifiable path in* $\mathfrak{G}$. *For any* $x \in E$, *define a path* p_x *in* E *by setting* $p_x(t) = p(t)x$ *for all* t *in the domain of* p. *Then* p_x *is a rectifiable path in* E *whose length* $\ell(p_x)$ *does not exceed* $\|x\| \cdot \ell(p)$.

PROOF. The proof is mechanical.

(4.7) PROPOSITION. *Suppose* $\mu \in S$ *and* $\beta \geqslant 0$. *Let* p *be a rectifiable path from* T *to* U *in* $\mathbf{U}(E, I)$. *Denote by* K *the image of* $\mathbf{P}_T(\{z \in S\colon \delta(z,\mu) \leqslant \beta\})$, *and by* L *the image of* $\mathbf{P}_U(\{z \in S\colon \delta(z,\mu) \leqslant \beta + \ell(p)\})$. *Then* $K \cap L^{\perp} = 0$.

PROOF. If possible, suppose that $x \in \Sigma \cap K \cap L^{\perp}$. From (4.2)(b), $D(Tx, \mu x) \leqslant \beta$, and from (4.2)(a), $D(Ux, \mu x) > \beta + \ell(p)$. However, from (4.6), $D(Tx, Ux) \leqslant \ell(p_x) \leqslant \ell(p)$. These three inequalities are clearly contradictory.

(4.8) PROPOSITION. *The statement of* (4.7) *remains valid if each weak inequality* ($\leqslant$ *or* $\geqslant$) *is replaced by the corresponding strong inequality* ($<$ *or* $>$).

PROOF. This follows as before, with (4.3) in place of (4.2).

(4.9) COROLLARY. *In each of* (4.7), (4.8), *the dimension of* K *cannot exceed the dimension of* L.

PROOF. If it did, then $K \cap L^{\perp}$ would be nonzero.

5. One-parameter subgroups

(5.1) For $T \in \mathbf{U}(E, J)$, define, recalling (3.3),

$$\log T = \int_{\mathbf{C}} \log z \, \mathbf{P}_T(dz).$$

Then $\log T$ is a bounded skew-adjoint complex-linear operator in E, that is, an element of $\mathbf{u}(E, I)$; it is compact if and only if $T \in \mathbf{UC}(E, J)$; and, most importantly,

$$\exp(\log T) = T.$$

(5.2) For any spectral measure $\mathbf{P}$, define the conjugate spectral measure $\overline{\mathbf{P}}$ by setting, for each Borel set Q in $\mathbf{C}$,

$$\overline{\mathbf{P}}(Q) = \mathbf{P}(\overline{Q}),$$

where $\overline{Q} = \{\bar{z} \colon z \in Q\}$. If $T \in \mathbf{U}(E, J)$, where J is a conjugate-isometry, then

$$T = J^{-1}TJ = \int_{\mathbf{C}} \bar{z} J^{-1} \mathbf{P}_T(dz) J,$$

so that, by the uniqueness of spectral decomposition,

$$\overline{\mathbf{P}}_T(dz) = J^{-1} \mathbf{P}_T(dz) J.$$

Therefore,

$$\begin{aligned} J^{-1}(\log T) J &= \int (\log z)^{-} J^{-1} \mathbf{P}_T(dz) J \\ &= \int (\log z)^{-} \, \overline{\mathbf{P}}_T(dz) \\ &= \int (\log \bar{z})^{-} \, \mathbf{P}_T(dz), \end{aligned}$$

and, as $\log \bar{z} = (\log z)^{-}$ for all $z \in S$ except -1, it follows that

$$J^{-1}(\log T) J - \log T = -2\pi i \mathbf{P}_T(\{-1\}).$$

Consequently, $\log T \in \mathbf{u}(E, J)$ if and only if -1 is not an eigenvalue of T.

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

(5.3) Suppose next that $J^2 = -I$, which is the quaternionic case (see (1.2)). Then $F = \ker(T + I)$ and $F^\perp$ are both subspaces (see (1.4)). Construct a quaternionic orthonormal basis in F, and define $V \in \mathbf{L}(E, I)$ to be $\log(T \mid F^\perp)$ on $F^\perp$, and right-multiplication (with respect to the chosen basis) by a pure imaginary quaternion of length π on F. As -1 is not an eigenvalue of $T \mid F^\perp$, it follows by (5.2) that $V \in \mathbf{L}(E, J)$. It is evidently bounded and skew-adjoint; moreover, if $T \in \mathbf{UC}(E, J)$, then F is finite-dimensional, and so V is compact (since $\log(T \mid F^\perp)$ is). Also, $\exp V = T$.

(5.4) Finally, suppose that $J^2 = +I$, the real case. Let $F = \ker(T + I)$ as before. If F is of infinite or even finite dimension, construct a real orthonormal basis (that is, one consisting of J-invariant vectors) $\{e_\nu, f_\nu : \nu \in A\}$ in F, for some index set A. Define $V \in \mathbf{L}(E, I)$ to be $\log(T \mid F^\perp)$ on $F^\perp$, and, on F, extend from $Ve_\nu = \pi f_\nu$, $Vf_\nu = -\pi e_\nu$. Then, as in (5.3), V is skew-adjoint, bounded, and exponentiates to T. In addition, when $T \in \mathbf{UC}(E, J)$, then V is compact.

(5.5) If $J^2 = I \neq J$ and $\ker(T + I)$ is of odd finite dimension, then T cannot be the exponential of any element of $\mathbf{L}(E, J)$. Indeed, if $T = \exp V$, where $V \in \mathbf{L}(E, J)$, then $\det(T \mid \ker(T + I)) = \exp(\operatorname{tr}(V \mid \ker(T + I))$, which must be real and positive.

(5.6) Let $J^2 = I \neq J$ and let $T \in \mathbf{U}(E, J)$.

(a) Suppose that -1 is isolated and of finite multiplicity in $\sigma(T)$ (see (3.1)), and let $\varepsilon > 0$ be such that

$$\{z \in S\colon \delta(z, -1) < 2\varepsilon\} \cap \sigma(T) = \{-1\}.$$

If $U \in \mathbf{U}(E, J)$ and $d(T, U) < \varepsilon$ (see (2.2)), then there is a rectifiable path p in $\mathbf{U}(E, J)$ from T to U such that $\ell(p) < \varepsilon$, and so, by (4.9),

$$\begin{aligned} \operatorname{rank} \mathbf{P}_T(\{-1\}) &\leqslant \operatorname{rank} \mathbf{P}_U(\{z \in S\colon \delta(z, -1) \leqslant \ell(p)\}) \\ &\leqslant \operatorname{rank} \mathbf{P}_U(\{z \in S\colon \delta(z, -1) \leqslant \varepsilon\}) \\ &\leqslant \operatorname{rank} \mathbf{P}_T(\{z \in S\colon \delta(z, -1) \leqslant \ell(p) + \varepsilon\}) \\ &= \operatorname{rank} \mathbf{P}_T(\{-1\}), \quad \text{by choice of } \varepsilon. \end{aligned}$$

Thus $\{z \in S\colon \delta(z, -1) \leqslant \varepsilon\} \cap \sigma(U)$ contains only finitely many points of $\sigma(U)$, all eigenvalues of finite multiplicity, and their multiplicities sum to the multiplicity of -1 in $\sigma(T)$. Apart from -1, these eigenvalues of U appear in distinct conjugate complex pairs of equal multiplicity; so the parities of $\dim \ker(T + I)$ and of $\dim \ker(U + I)$ must be the same. In particular, if -1 is isolated and of odd finite multiplicity in $\sigma(T)$, then it is isolated and of odd finite multiplicity in $\sigma(U)$.

(b) On the other hand, suppose that $F = \ker(T + I)$ is of odd finite dimension, but that -1 is not isolated in $\sigma(T)$. Then $(T + I) \mid F^\perp$ is one-one, but not onto $F^\perp$; however, its image is dense in $F^\perp$. Let $x \in F$ and $y \in F^\perp$ be real unit

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

vectors, with $y \notin (T+I)(F^{\perp})$. Given $\theta \in \mathbf{R}$, define an operator $M(\theta)$ by linear extension from $M(\theta)|\{x, y\}^{\perp} = 0$, $M(\theta)x = \theta y$, $M(\theta)y = -\theta x$. Then $M(\theta) \in \mathbf{L}(E, J)$; $M(\theta)$ is skew-adjoint; and the operator-norm of $M(\theta)$ is θ. Now assume that $\theta \neq 0$. Then $(I + M(\theta))^{-1}(I - M(\theta))T\xi = -\xi$ if and only if $M(\theta)(I - T)\xi = -(I + T)\xi$; but $(T + I)\xi \in (T + I)F^{\perp}$, which meets the image of $M(\theta)$, namely the span of $\{x, y\}$, only in $\{0\}$. Hence both $(T + I)\xi = 0$ and $M(\theta)(I - T)\xi = 0$, and so $2M(\theta)\xi = 0$. Thus $\xi \in \{x, y\}^{\perp}$, and $\xi \in F$. Since the argument clearly reverses, and since $y \in F^{\perp}$ by construction, $\ker\{(I + M(\theta))^{-1}(I - M(\theta))T + I\} = F \cap \{x, y\}^{\perp} = F \cap \{x\}^{\perp}$, which has even finite dimension $\dim(F) - 1$. As θ may be arbitrarily small, this proves that T may be approximated arbitrarily closely by elements of $\mathbf{U}(E, J)$ which are exponentials of elements of $\mathbf{u}(E, J)$.

(5.7) For $T \in \mathbf{UC}(E, J)$, with $J^2 = I \neq J$ as before, $\sigma(T)$ can contain -1 only as an isolated point of finite multiplicity, and (5.6)(a) in effect shows that the parity of $\dim\ker(T + I)$, as a function of T, is locally constant on $\mathbf{UC}(E, J)$. If $\ker(T + I)$ is even-dimensional, then, by (5.4), $T = \exp V$ for some $V \in \mathbf{uc}(E, J)$; thus the path $\exp(tV)$, $0 \leqslant t \leqslant 1$, joins T to I in $\mathbf{UC}(E, J)$. So the principal component of $\mathbf{UC}(E, J)$, the 'special Fredholm orthogonal group' $\mathbf{SUC}(E, \mathbf{J})$, is precisely the subset of $\mathbf{UC}(E, J)$ consisting of operators T for which $\ker(T + I)$ is of even dimension. It is possible to show, also by operator-theoretic methods, that $\mathbf{SUC}(E, J)$ is of index 2 in $\mathbf{UC}(E, J)$. These results are usually proved by finite-dimensional approximation: see, for instance, [6].

(5.8) Henceforth, $\mathfrak{G}$ will denote the full unitary, orthogonal, or symplectic group $\mathbf{U}(E, J)$ (with $J = I$, $J^2 = I \neq J$, or $J^2 = -I$, respectively), or the Fredholm unitary or symplectic group $\mathbf{UC}(E, J)$ (with $J = I$ or $J^2 = -I$, respectively), or the special Fredholm orthogonal group $\mathbf{SUC}(E, J)$ (with $J^2 = I \neq J$). All of these groups except the full orthogonal group will be described as 'exponential', since in them every element is the exponential of some element of the Lie algebra. In the full orthogonal group, an element is an exponential if and only if it is a square.

6. The Finsler metric

(6.1) LEMMA. *For any $T \in \mathfrak{G}$, $d(I, T) \geqslant N(T)$.*

PROOF. Take any $\varepsilon > 0$. By (4.4), there is $x \in \Sigma$ such that $D(Tx, x) > N(T) - \varepsilon$. Suppose that p is any piecewise C^1 (or merely rectifiable) path from I to T in $\mathfrak{G}$. Then, by (4.6), $\ell(p_x) \leqslant \ell(p)$. However, $\ell(p_x) \geqslant D(p_x(1), p_x(0)) = D(Tx, x) > N(T) - \varepsilon$. Hence $\ell(p) > N(T) - \varepsilon$. As p was arbitrary, we deduce that $d(I, T) \geqslant N(T) - \varepsilon$; and since ε is arbitrary, the result follows.

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

(6.2) LEMMA. *Suppose $T \in \mathfrak{G}$. Then there is a geodesic of length $N(T)$ joining I to T in $\mathfrak{G}$, unless $\mathfrak{G} = \mathrm{U}(E, J)$, where (E, J) is real, and* $\ker(T + I)$ *is of odd finite dimension; except in those circumstances, $d(I, T) \leqslant N(T)$.*

PROOF. First, suppose that $N(T) < \pi$. Then $-1 \notin \sigma(T)$, and, by (5.2), $\log T$ belongs to $\mathfrak{g}$ (see (1.6)). In this case set $V = \log T$. Secondly, suppose that $N(T) = \pi$, so that $-1 \in \sigma(T)$. With the exception specified in the statement, one of (5.2), (5.3), or (5.4) will yield $V \in \mathfrak{g}$ such that $\exp V = T$. (For the special Fredholm orthogonal group, $\ker(T + I)$ is always of even finite dimension, by (5.7); so (5.4) applies.) In either case, V is skew-adjoint, and therefore has norm equal to its spectral radius. In the first case, the spectral radius is $N(T)$ from the definition (3.3), whilst in the second, the spectral radius is π, which remains equal to $N(T)$.

Consider now the path $q(t) = \exp(tV)$, $0 \leqslant t \leqslant 1$. For each t, $\dot{q}(t) = T_I L_{q(t)} \cdot V$ (see (2.1)), and it follows that $\|\dot{q}(t)\|_{q(t)} = \|V\| = N(T)$. Hence $\ell(q) = N(T)$, and, as q joins I to T, this proves that $d(I, T) \leqslant N(T)$, as asserted.

(6.3) THEOREM. *Let $\mathfrak{G}$ be exponential (see* (5.8)*). Then, for any T, $U \in \mathfrak{G}$, $d(T, U) = N(U^{-1}T)$.*

PROOF. Both sides are left-invariant (see (2.2)), so it suffices to prove the equality when $U = I$. As $\mathfrak{G}$ is exponential, both (6.1) and (6.2) apply.

(6.4) LEMMA. *Let T, $U \in \mathfrak{G}$ and $\mu \in \sigma'(U)$. Then*

$$\{z \in S\colon \delta(z, \mu) \leqslant d(T, U)\} \cap \sigma'(T) \neq \varnothing.$$

PROOF. Suppose not. Then, by the compactness of $\sigma'(T)$, there exists $\beta > 0$ such that

$$\{z \in S\colon \delta(z, \mu) \leqslant d(T, U) + 2\beta\} \cap \sigma'(T) = \varnothing.$$

Choose a rectifiable path p from T to U in $\mathfrak{G}$ such that $\ell(p) < d(T, U) + \beta$. Now $\{z \in S\colon \delta(z, \mu) \leqslant \ell(p) + \beta\}$ can contain at most finitely many points of $\sigma(T)$, all isolated and of finite multiplicity (see (3.1), and compare (3.2)); thus $\mathbf{P}_T(\{z \in S\colon \delta(z, \mu) \leqslant \ell(p) + \beta\})$ is of finite rank. By (4.9), so also is $\mathbf{P}_U(\{z \in S\colon \delta(z, \mu) \leqslant \beta\})$. But, as $\mu \in \sigma'(U)$, this contradicts (3.2)(b).

(6.5) THEOREM. *Let (E, J) be a real Hilbert space of infinite dimension, and let $\mathfrak{G} = \mathbf{U}(E, J)$. Then for T, $U \in \mathfrak{G}$,*

$$d(T, U) = N(U^{-1}T),$$

unless -1 *is an isolated point of* $\sigma(U^{-1}T)$ *of odd finite multiplicity, in which case*

$$d(T,U) = 2\pi - N'(U^{-1}T).$$

NOTE. If (E, J) is real and of finite dimension, then $\mathrm{U}(E,J) = \mathrm{UC}(E,J)$; thus, if T and U belong to different components of $\mathrm{U}(E,J)$ (see (5.7)), then $d(T,U) = 2\pi$, by (2.2); otherwise, $d(T,U) = N(U^{-1}T)$, by (6.3).

PROOF. As in (6.3), we may take $U = I$. The proof occupies (6.6)–(6.10).

(6.6) If $\ker(T+I)$ is not of finite odd dimension, then (6.1) and (6.2) both hold, and $d(I,T) = N(T)$, as before. If $\ker(T+I)$ is of finite odd dimension, but -1 is not isolated in $\sigma(T)$, then (5.6)(b) approximates T by elements of $\mathfrak{G}$ whose distance from I does not exceed π (by (6.2)); consequently, $d(I,T) \leqslant \pi$ in this case also, and, from (6.1), $d(I,T) = \pi = N(T)$. In particular, if $N(T) < \pi$, then $-1 \notin \sigma(T)$ and $d(I,T) = N(T) < \pi$.

(6.7) Suppose now that $d(I,T) = \pi$. Given $\varepsilon > 0$, take a piecewise C^1 path p: $[0,1] \to \mathfrak{G}$ such that $p(0) = I$, $p(t) = T$, and $\ell(p) < \pi + \varepsilon/2$. Since $d(I, p(t))$ is continuous in t, there is a point $\tau \in [0,1]$ such that $d(I, p(\tau)) = \pi - \varepsilon/2$. Ergo,

$$\int_0^\tau \|\dot{p}(t)\|_{p(t)}\,dt \geqslant \pi - \varepsilon/2$$

(the length of the path between parameters 0 and τ), and

$$\int_\tau^1 \|\dot{p}(t)\|_{p(t)}\,dt < \varepsilon$$

(the length remaining). Hence, $d(I, p(\tau)) < \pi$ and $d(p(\tau), T) < \varepsilon$. Therefore, T may be approximated in $\mathfrak{G}$ by elements closer than π to I, for which, by (6.1), the spectrum cannot contain -1. In view of (5.6)(a), this cannot occur when -1 is an isolated point of $\sigma(T)$ of odd finite multiplicity. In that case, then, $d(I,T) > \pi$ (indeed, (6.1) ensures that $d(I,T) \geqslant \pi$, and the supposition of equality has led to a contradiction).

(6.8) Together, (6.6) and (6.7) prove that $d(I,T) > \pi$ if and only if -1 is an isolated point of $\sigma(T)$ of odd finite multiplicity, and that otherwise $d(I,T) = N(T)$. Suppose, therefore, that -1 is isolated in $\sigma(T)$ and of odd finite multiplicity. Let $q(t)$, $0 \leqslant t \leqslant 1$, be any piecewise C^1 (or merely rectifiable) path from I to T in $\mathfrak{G}$, and set

$$\rho = \sup\{t\colon 0 \leqslant t \leqslant 1, d(q(t), I) \leqslant \pi\}.$$

By continuity, $d(q(\rho), I) = \pi$, and so $-1 \in \sigma(q(\rho))$. If $-1 \notin \sigma'(q(\rho))$, then -1 is isolated and of finite multiplicity in $\sigma(q(\rho))$ (see (3.1)), and, by (5.6)(a), $\dim\ker(q(t)+I)$ is finite and of constant parity for all t sufficiently close to ρ.

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

For $t = \rho$, the parity is even, as $d(q(\rho), I) = \pi$; but for $t > \rho$, the parity is odd, since $d(q(t), I) > \pi$. Hence $-1 \in \sigma'(q(\rho))$, in fact.

Now take $\mu = -1$ and $U = q(\rho)$ in (6.4). The δ-distance in S from -1 to $\sigma'(T)$ is $\pi - N'(T)$; it follows that

$$d(T, q(\rho)) \geqslant \pi - N'(T).$$

So the segment of q between the parameters 0 and ρ has length not less than π, whilst the remaining segment has length at least $\pi - N'(T)$. In sum, therefore,

$$\ell(q) \geqslant 2\pi - N'(T),$$

and, as q was any piecewise C^1 path from I to T, this shows that

(1) $$d(I, T) \geqslant 2\pi - N'(T).$$

(6.9) It remains to prove the opposite inequality to (6.8)(1) for the same operator T. For (6.9), I shall write $\mathbf{P}$ instead of $\mathbf{P}_T$ and N' in place of $N'(T)$. As $-1 \notin \sigma'(T)$ by hypothesis, certainly $N' < \pi$. Define

$$Z_+ = \{z \in S\colon \operatorname{Im} z > 0 \text{ and } \delta(z, 1) \geqslant N'\},$$
$$Z_- = \{z \in S\colon \operatorname{Im} z < 0 \text{ and } \delta(z, 1) \geqslant N'\} = \{\bar{z}\colon z \in Z_+\},$$
$$Z = Z_+ \cup Z_-.$$

(a) The first case occurs when $N' > 0$ and $\mathbf{P}(Z_+)$, $\mathbf{P}(Z_-)$ are of infinite rank; in other words, $\exp(\pm iN')$ are either eigenvalues of infinite multiplicity or limits of sequences of points further to the left in $\sigma(T)$. Set $V_+ = \int_{Z_+} (i\pi - \log z)\mathbf{P}(dz)$, and $V_- = \int_{Z_-} (-i\pi - \log z)\mathbf{P}(dz)$. Imitating (5.2), one finds that

$$\begin{aligned} J^{-1}V_+J &= \int_{Z_+} \left(-i\pi - (\log z)^{-}\right) J^{-1}\mathbf{P}(dz) J \\ &= \int_{Z_+} (-i\pi - \log \bar{z})\overline{\mathbf{P}}(dz) \quad (\text{as } -1 \notin Z_+) \\ &= \int_{Z_-} (-i\pi - \log z)\mathbf{P}(dz) = V_-, \end{aligned}$$

and symmetrically (as $J^2 = I$), $V_+ = J^{-1}V_-J$. Hence $V = V_+ + V_-$ commutes with J, and is skew-adjoint, since both V_+ and V_- are. Ergo, $V \in \mathbf{u}(E, J)$ (see (1.6)). Also $\sigma(V) = \{i\pi - \log z\colon\ z \in Z_+ \cap \sigma(T)\} \cup \{-i\pi - \log z\colon\ z \in Z_- \cap \sigma(T)\}$, where, by definition, $\log Z_+ \subseteq [iN', i\pi]$, and $\exp(iN') \in \sigma(T)$. Therefore the spectral radius and the norm of V are $\pi - N'$.

Consider the C^ω path in $\mathbf{U}(E, J)$ defined by

$$r(t) = T\exp(tV), \qquad 0 \leqslant t \leqslant 1.$$

As in (6.2), $\|\dot{r}(t)\| = \|V\| = \pi - N'$, for each t. Thus

$$\ell(r) = \pi - N'.$$

It is clear from the construction that

$$\ker(r(1)+I) = \mathbf{P}(\{-1\} \cup Z)(E),$$

which is of infinite dimension, by hypothesis.

(b) The second case occurs when $\mathbf{P}(Z)$ is of odd finite dimension. Then $N' > 0$ necessarily, and $\exp(\pm iN')$ lie in $\sigma'(T)$ only because they are limits of sequences of points further to the right in $\sigma(T)$. Set $\tau = \pi/N' - 1$ and

$$V = \int_{\delta(z,1)<N'} \log z \mathbf{P}(dz).$$

as in (a), $V \in \mathbf{u}(E, J)$ and $\|V\| = N'$. For $0 \leqslant t \leqslant 1$, define

$$r(t) = T\exp(\tau t V).$$

Then $\ell(r) = \|\tau V\| = \pi - N'$, and $r(1)$ may be written as an integral with respect to $\mathbf{P}$. From this expression of $r(1)$, it is clear that -1 is not isolated in $\sigma(r(1))$.

(c) The third case occurs when $N' = 0$, but when 1 is not an eigenvalue of T of infinite multiplicity. Since Z_+, Z_-, as defined, do not contain 1, the formulae and arguments of (a) work in this case without alteration.

(d) Finally, suppose that $N' = 0$ and that $F = \ker(T - I)$ is of infinite dimension. By (6.2), there is a C^ω path $s(t)$, $0 \leqslant t \leqslant 1$, which joins $I \mid F$ to $-I \mid F$ in $\mathrm{U}(F, J\mid)$ and which is of length π. Define the path r in $\mathrm{U}(E, J)$ by

$$r(t) \mid F^\perp = T \mid F^\perp, \qquad r(t) \mid F = s(t).$$

Then r is also C^ω of length π, $r(0) = T$, and $r(1)$ has -1 as an eigenvalue of infinite multiplicity.

(6.10) In each case of (6.9), I have found a path r of length $\pi - N'(T)$ which joins T to a point of $\mathfrak{G}$ at distance π from I in $\mathfrak{G}$ (by (6.6)). Hence

$$d(I, T) \leqslant 2\pi - N'(T),$$

which, with (6.8)(1), completes the proof of (6.5).

(6.11) Remark. (4.4) and (6.3) show that, when $\mathfrak{G}$ is exponential, and when T, $U \in \mathfrak{G}$, then $d(T, U) = \sup\{D(Tx, Ux) \colon x \in \Sigma\}$. However, (6.5) asserts that this is false for the full orthogonal group. Thus the group of all isometries of the unit sphere in infinite-dimensional Hilbert space has two distinct naturally defined invariant metrics.

7. Minimising paths and geodesics

(7.1) Theorem. *The points T, $U \in \mathfrak{G}$ may be joined by a minimising geodesic in $\mathfrak{G}$ if and only if they may be joined by a geodesic, and, specifically, they may be so joined unless $\mathfrak{G} = \mathrm{U}(E, J)$, where (E, J) is real, and $\ker(T + U)$ is of odd finite dimension.*

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

PROOF. Notice that $\ker(T+U) = \ker(I + U^{-1}T)$, and that $U^{-1}p(t)$ is a minimising geodesic from I to $U^{-1}T$ if and only if $p(t)$ is a minimising geodesic from U to T. Thus it suffices to take $U = I$, as in (6.3), (6.5). The result follows from (6.2), (6.1), (5.5), (5.7).

(7.2) NOTE. If T and U may be joined by a geodesic, then, by (5.5) and (6.2), $d(T,U) \leqslant \pi$. Equally, if $d(T,U) < \pi$, then T and U may be joined by a minimising geodesic, by (6.3), (6.5) and (7.1).

(7.3) PROPOSITION. *Suppose that $p(t)$ and $q(t)$ (for $0 \leqslant t \leqslant 1$) are both minimising geodesics from U to T in $\mathfrak{G}$. Then, for each t,*

$$p(t)\,|\,(\ker(T+U)) = q(t)\,|\,(\ker(T+U)).$$

Moreover, there is exactly one minimising geodesic joining U to T and parametrised by $[0,1]$ *if and only if* $\ker(T+U) = 0$.

PROOF. Take $U = I$, as in (7.1). If $V \in \mathfrak{g}$ and $\exp V = T$, then

$$T = \int_S z\mathbf{P}_T(dz) = \int_S \exp(z)\mathbf{P}_V(dz).$$

By the uniqueness of spectral decompositions, this gives

(1) $$\mathbf{P}_T(Q) = \mathbf{P}_V(\exp^{-1}(Q))$$

for any Borel set Q in S. If $\exp(tV)$, $0 \leqslant t \leqslant 1$, is to be a minimising geodesic, its length, which is the spectral radius of V (see (6.2)), cannot exceed π, by (7.2); so $\mathbf{P}_V$ is supported on the segment $[-i\pi, i\pi]$ of the imaginary axis, and (1) determines $\mathbf{P}_V(Q)$, unless Q contains $-i\pi$ or $i\pi$. Of $\mathbf{P}_V(\{i\pi\})$ and $\mathbf{P}_V(\{-i\pi\})$, (1) shows only that their sum must be $\mathbf{P}_T(\{-1\})$. In any case, $V\,|\,(\ker(I+T))^\perp$ is uniquely fixed, which proves the first assertion of the proposition. For the second, note that the constructions (5.2), (5.3), (5.4) of a minimising geodesic (see (6.2)) are all non-unique when $\ker(T+I) \neq 0$. In (5.2), one might take $\log(-1)$ to mean $-i\pi$ (compare (3.3)); in (5.3), one may choose a different basis or a different pure imaginary quaternion; in (5.4), one has a choice of bases.

(7.4) Apart from trivial exceptions when E is of low finite dimension, the operator-norm in $\mathfrak{g}$ is not uniformly convex, and one cannot, therefore, reasonably expect general uniqueness theorems for minimising paths.

Suppose $V \in \mathfrak{g}$, $\|V\| \leqslant \pi$, and $T = \exp V$; let F be a V-invariant closed nonzero subspace of (E, J) with $N(T\,|\,F) = \|V\,|\,F\| < \|V\| = N(T)$, and, in the real case, $\dim F \geqslant 2$. Take $\lambda \in (0,1)$ and write $U = \exp(\lambda V)$. By construction, $N(U) = \lambda N(T) > \lambda N(T\,|\,F) = N(U\,|\,F)$, and $N(U\,|\,F^\perp) = N(T\,|\,F^\perp)$. Hence, by (3.4),

(1) $$N(U) = \max(N(U\,|\,F), N(U\,|\,F^\perp)) = N(U\,|\,F^\perp).$$

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

Now suppose that $W \in \mathfrak{G}$, $W \mid F^{\perp} = I \mid F^{\perp}$ (so F is W-invariant) and

(2) $$N(W \mid F) \leqslant (N(T) - N(T \mid F)) \cdot \min(\lambda, 1 - \lambda).$$

Then, by (4.5), (2), and (3.4) in turn,

(3) $$N(UW \mid F) \leqslant N(U \mid F) + N(W \mid F) \leqslant \lambda N(T),$$

and

$$\begin{aligned} N(UW) &= \max\big(N(UW \mid F), N(UW \mid F^{\perp})\big) \\ &= \max\big(N(UW \mid F), N(U \mid F^{\perp})\big) \\ &= N\big(U \mid F^{\perp}\big) = \lambda N(T), \quad \text{from (3)}. \end{aligned}$$

By the same sort of argument,

$$N(W^{-1}U^{-1}T) = (1 - \lambda)N(T).$$

Consequently, both UW and $W^{-1}U^{-1}T$ may be joined to I by minimising geodesics (see (7.1)) $p(t)$ and $q(t)$, respectively, for $0 \leqslant t \leqslant 1$. Define

$$\begin{aligned} r(t) &= p(t/\lambda) && \text{for } 0 \leqslant t \leqslant \lambda, \\ r(t) &= UWq((t - \lambda)/(1 - \lambda)) && \text{for } \lambda \leqslant t \leqslant 1. \end{aligned}$$

Then, clearly, r is a rectifiable path from I to T in $\mathfrak{G}$, and $\ell(r) = \ell(p) + \ell(q) = \lambda N(T) + (1 - \lambda)N(T) = N(T) = d(I, T)$. However, the minimising path r between I and T cannot be a geodesic, unless $W = I$. For F is invariant under both T and $T^* = \exp(-V)$, and so, from the spectral decomposition of T, it follows that $F \subsetneq (\ker(T + I))^{\perp}$ (since $N(T \mid F) < \pi$); but r has been constructed to differ from $\exp(tV)$ on F. See (7.3).

Provided that a subspace F may be found, this argument provides many minimising paths that are not geodesics. For each of the groups $\mathfrak{G}$, many choices of W are possible, and in addition λ may be arbitrarily selected in $(0, 1)$. The construction may also, in suitable circumstances, be repeated for each of the geodesic segments making up r. In sum, there are very many possible minimising paths between I and a typical point T to which it may be geodesically joined. (An alternative construction is as follows. Suppose that $T = (I + S)^{-1}(I - S)$, where $S \in \mathfrak{g}$. Then

$$s(t) = (I + tS)^{-1}(I - tS), \qquad 0 \leqslant t \leqslant 1,$$

is a minimising path (in general not uniformly parametrised) which cannot usually be reparametrised to a geodesic. This construction would suffice for (7.6).)

(7.5) LEMMA. *Suppose $T \in \mathfrak{G}$, $d(I, T) = N(T)$. Let $p(t)$, $0 \leqslant t \leqslant N(T)$, be a uniformly parametrised minimising path from I to T in $\mathfrak{G}$, and $x \in \Sigma$ an eigenvector of T with corresponding eigenvalue* $\exp(iN(T))$ (*or* $\exp(-iN(T))$). *Then p_x (see (4.6)) is a uniformly parametrised minimising path from x to Tx in Σ.*

PROOF. By (4.1)(a), $D(Tx, x) = N(T)$; so $\ell(p_x) \geqslant N(T)$. By (4.6), $\ell(p_x) \leqslant \ell(p) = N(T)$. Hence $\ell(p_x) = N(T)$. If $0 < \tau < N(T)$, then the length of p_x between the values 0 and τ of the parameter cannot exceed the corresponding length τ of p, nor can its length between τ and $N(T)$ exceed $N(T) - \tau$, by (4.6); hence, in both cases, strict equality is needed to ensure that the total length is $N(T)$.

(7.6) THEOREM. *Suppose that T, $U \in \mathfrak{G}$ may be joined in $\mathfrak{G}$ by a minimising geodesic. Then every uniformly parametrised minimising path from T to U in $\mathfrak{G}$ is a geodesic if and only if, in the complex and quaternionic cases, all elements of $\sigma(U^{-1}T)$ have the same real part; or, in the real case, all elements other than* 1 *of $\sigma(U^{-1}T)$ have the same real part, and* 1, *if it belongs to $\sigma(U^{-1}T)$, is an eigenvalue of multiplicity at most* 1.

PROOF. As usual, take $U = I$, and let $V \in \mathfrak{g}$ be as in (7.4). If the stated condition is not satisfied, then, for at least some $a \in [0, \|V\|)$, the image of $\mathbf{P}_V(\{ix\colon x \in \mathbf{R}, |x| \leqslant a\})$ will be a closed nonzero subspace F of (E, J) such that $N(T \mid F) \leqslant a < N(T)$; in the real case the dimension of F must be at least 2. Thus (7.4) ensures that not all uniformly parametrised minimising paths from I to T in $\mathfrak{G}$ are geodesics.

Now suppose the stated condition holds, and set $E^+ = \ker(T - \exp(iN(T))I)$, $E^- = \ker(T - \exp(-iN(T))I)$, and $D = E^+ + E^-$. Let $p(t)$, $0 \leqslant t \leqslant N(T)$, be a uniformly parametrised minimising path from I to T. Suppose $x \in E^+ \cap \Sigma$. By (7.5) and (4.1)(c), p_x is a geodesic in Σ, and, by (4.1)(b),

$$(1) \qquad p_x(t) = (\cos t)\, p_x(0) + (\sin t)\, \dot{p}_x(0)$$

for all t. The same argument applies when $x \in E^- \cap \Sigma$, and so, by linearity, (1) holds for any $x \in D$. In particular, $\dot{p}_x(0)$, the right derivative of p_x at 0, exists for each $x \in D$. Thus $p(t) \mid D$ is differentiable on the right at 0 in the strong operator topology; the derivative $A\colon D \to E$ is a bounded linear operator (by the Banach-Steinhaus theorem), and $A(J \mid D) = JA$.

Rewrite (1), recalling that $p(0) = I$, as

$$(2) \qquad p(t) \mid D = (\cos t) I \mid D + (\sin t) A.$$

When $N(T) = 0$, there is nothing to prove. Otherwise, the right-hand side of (2) is isometric on D for each $t \in [0, N(T)]$ if and only if, for every x, $y \in D$,

$$(3) \qquad \langle Ax, Ay \rangle = \langle x, y \rangle, \quad \langle Ax, y \rangle = -\langle x, Ay \rangle.$$

Suppose first that $A(D) \subseteq D$. Consider A as an operator in the closed subspace D; by (3), $A^*A = I \mid D$, $A^* = -A$, and therefore $A^2 = -I \mid D$ and A is a bijection of D with itself. Substituting $A^2 = -I \mid D$ in (2), one finds

$$(4) \qquad p(t) \mid D = \exp(tA) \mid D \quad \text{for } 0 \leqslant t \leqslant N(T).$$

In the complex and quaternionic cases, $D = E$ by hypothesis and (4) completes the proof. In the real case, it is also possible that D be of codimension 1; in that case, let ξ denote a real unit vector perpendicular to D.

If $A(D) \subseteq D$, (4) shows $p(t)D = D$ for each t. Since $p(t)$ is real and orthogonal, it must follow that, for each t, $p(t)\xi = \pm\xi$. But $p(t)$ is continuous and $p(0) = I$; hence $p(t)\xi = \xi$ for all t. Define A_1 by $A_1 | D = A$, $A_1\xi = 0$; then A_1 is real skew-adjoint (by (3)), and, in view of (4), $p(t) = \exp(tA_1)$ for all t.

Finally, suppose that $A(D) \not\subseteq D$. Then $A(D) + D = E$ necessarily. Choose real elements a, $b \in D$ such that $\xi = Aa + b$, and define $A_2 | D = A$, $A_2\xi = Ab - a$. It now follows trivially from (3) that, for all x, $y \in E$,

$$(5) \qquad \langle A_2x, A_2y\rangle = \langle x, y\rangle, \qquad \langle A_2x, y\rangle = -\langle x, A_2y\rangle.$$

As before, it follows that $A_2^2 = -I$ and therefore that

$$(6) \qquad \exp(tA_2) = (\cos t)I + (\sin t)A_2 \quad \text{for all } t.$$

By (2), then, $\exp(tA_2)$ and $p(t)$ agree on D. Both are orthogonal operators (by (5), A_2 is skew-adjoint), are continuous in t, and are the identity for $t = 0$. Thus they must agree on ξ as well, for all t. This completes the proof. (Notice that (6) now shows $D = E$, so that this last case cannot in fact occur. This is assumed below.)

(7.7) COROLLARY. *If T, $U \in \mathfrak{G}$ may be joined by a geodesic, then every uniformly parametrised minimising path from T to U in $\mathfrak{G}$ is a geodesic if and only if, in the complex and quaternionic cases, $-T \in \mathfrak{G}$ and U lies on a minimising geodesic from T to $-T$; in the real case, there is the additional possibility that there exist a real subspace D of codimension 1 in E and that U lies on a minimising geodesic p from T to T_1, where $T_1 | D = -T | D$ and $T_1 | D = T | D = p(t) | D$ for $0 \leqslant t \leqslant \pi$.*

(7.8) If $\mathfrak{G} = \mathbf{UC}(E, J)$, where E is infinite-dimensional, the conditions of (7.7) cannot hold for any T, $U \in \mathfrak{G}$.

8. Minimising paths in the orthogonal group

(8.1) THEOREM. *Let $\mathfrak{G} = \mathbf{U}(E, J)$, where (E, J) is a real Hilbert space of infinite dimension. The points T, $U \in \mathfrak{G}$ cannot be joined by any minimising path in $\mathfrak{G}$ if and only if -1 is a non-isolated point of $\sigma(U^{-1}T)$ which is an eigenvalue of odd finite multiplicity. In all other cases there is a minimising path between T and U in $\mathfrak{G}$ which consists of at most two geodesic segments.*

Proof. Take $U = I$ as usual. Then (7.1) settles the case when -1 is not an eigenvalue of T of odd finite multiplicity. If -1 is isolated and of odd finite multiplicity in $\sigma(T)$, but if $\mathbf{P}_T(\{z \in S\colon \delta(z,1) \geqslant N'(T)\}$ is not of finite rank, then one of (6.9)(a), (c), or (d) will construct a minimising path from I to T which consists of two geodesic segments.

The two remaining cases are discussed below. In (8.2) we prove that, when -1 is not isolated in $\sigma(T)$ but is an eigenvalue of odd finite multiplicity, then there is no minimising path joining T to I in $\mathfrak{G}$. In (8.3)–(8.12), on the other hand, we construct minimising paths from T to I when $\mathbf{P}_T(\{z \in S\colon \delta(z,1) \geqslant N'(T)\})$ is of finite odd rank and $N'(T) < \pi$. Certainly, then, $N'(T) > 0$, so that these paths are of length less than 2π (by (6.5)). Now, if $p(t)$, $0 \leqslant t \leqslant d(T,I)$, is a minimising path from I to T which is uniformly parametrised by length, then by (7.2), there are minimising geodesics between I and $p(d(T,I)/2)$, and from $p(d(T,I)/2)$ to T; so our construction will complete the proof.

(8.2) Assume first that $F = \ker(T + I) \neq 0$, $d(T,I) = \pi$. Let $p(t)$, $0 \leqslant t \leqslant \pi$, be (if possible) a uniformly parametrised minimising path from I to T in $\mathfrak{G}$. By (6.6),

$$(1) \qquad d(p(t), I) = t = N(p(t)) \quad \text{for } 0 \leqslant t \leqslant \pi.$$

Now take $x \in F \cap \Sigma$. By (7.5), p_x is uniformly parametrised in Σ, so that

$$(2) \qquad D(p(t)x, x) = t = d(p(t), I) \quad \text{for } 0 \leqslant t \leqslant \pi,$$

and p_x is a minimising path from x to $Tx = -x$ in Σ. Applying (4.1)(c), (b), we find that, for $0 \leqslant t \leqslant \pi$,

$$p_x(t) = (\cos t)p_x(0) + (\sin t)\dot{p}_x(0),$$

where $p_x(0) = x$ and $\dot{p}_x(0)$, the derivative on the right at zero, must exist. Since this is so for each $x \in F \cap \Sigma$, it holds, by linearity, for all $y \in F$; moreover, $\dot{p}_y(0) = Ay$, where $A\colon F \to E$ is a real linear map such that $A(J \mid F) = JA$. Thus, for all $y \in F$, and for $0 \leqslant t \leqslant \pi$,

$$(3) \qquad p(t)y = (\cos t)y + (\sin t)Ay.$$

Now suppose that $x \in F \cap \Sigma$ is real, $Jx = x$. From (1) and (2), $D(p(t)x, x) = N(p(t))$; so, if we set $\mu = 1$, $T = p(t)$, $\gamma = N(p(t))$ in (4.2)(b) (where, in view of the definition (3.3) of $N(p(t))$, $H = E$), then we deduce that

$$(4) \qquad x \in \mathbf{P}_{p(t)}(\{\exp(it), \exp(-it)\})E \quad \text{for } 0 \leqslant t \leqslant \pi.$$

If $M_t = \mathbf{P}_{p(t)}(\{\exp(it)\})E$, then $J(M_t) = \mathbf{P}_{p(t)}(\{\exp(-it)\})E$, and (4) may be rewritten as $x \in M_t + J(M_t)$. However, x is real; so there exists $y(t) \in M_t + J(M_t)$, also real, such that $x + iy(t) \in M_t$. (Explicitly, if $a(t) \in M_t$ and $x = a(t) + Ja(t)$, take $iy(t) = a(t) - Ja(t)$.) Thus

$$(5) \qquad p(t)(x + iy(t)) = \exp(it) \cdot (x + iy(t)).$$

As $p(t) \in \mathbf{U}(E, J)$, it preserves real and imaginary parts. Therefore, equality of the real parts in (5) yields

(6) $$p(t)x = (\cos t)x - (\sin t)y(t).$$

Comparison of (6) and (3) demonstrates that

$$y(t) = -Ax \quad \text{for } 0 < t < \pi.$$

(Note, however, that (6) and (3) do not determine $y(0)$ or $y(\pi)$.) Now substitute back into (5) and take the imaginary parts to obtain:

(7) $$p(t)Ax = (\cos t)Ax - (\sin t)x,$$

for $0 < t < \pi$. Let $t \uparrow \pi$; then (7) shows that in the limit

$$TAx = -Ax, \quad \text{or} \quad Ax \in F.$$

Since F is spanned by real unit vectors, it follows that in fact $A(F) \subseteq F$, and therefore, by (3), that F is $p(t)$-invariant for $0 \leqslant t \leqslant \pi$.

To complete the argument, suppose that F is of finite dimension. Then $p(t) \mid F$, $0 \leqslant t \leqslant \pi$, is a path joining $I \mid F$ to $-I \mid F$ in $\mathbf{UC}(F, J \mid F)$; by (5.7), or by determinants, the dimension of F must be even. We conclude that, if F is of odd finite dimension, but if -1 is not isolated in $\sigma(T)$ (so that, by (6.5), $d(I, T) = \pi$) then there cannot be any minimising path from I to T in G.

(8.3) We shall construct a minimising path in the remaining case of (8.1) in three steps: some technical lemmas, (8.4)–(8.7); a minimising path in a special situation, (8.8)–(8.10); and the general construction, (8.11)–(8.12).

The following notation will be used: $(f_n)_{n=0}^{\infty}$ denotes a real orthonormal sequence in E (so that $Jf_n = f_n$ for each n), whose closed span is F; $(\alpha_n)_{n=1}^{\infty}$ and $(\beta_n)_{n=0}^{\infty}$ denote bounded sequences of nonnegative real numbers; α denotes a positive real number; $\alpha_0 = \beta_{-1} = 0$; and A and B denote the skew-adjoint elements of $\mathbf{L}(E, J)$ defined by

$A \mid F^{\perp} = 0, \quad Af_0 = 0, \quad$ and, $\quad$ for $n \geqslant 1$,

(1) $$Af_{2n} = -\alpha_n f_{2n-1}, \qquad Af_{2n-1} = \alpha_n f_{2n};$$

$B \mid F^{\perp} = 0, \quad$ and, $\quad$ for $n \geqslant 0$,

(2) $$Bf_{2n} = \beta_n f_{2n+1}, \qquad Bf_{2n+1} = -\beta_n f_{2n}.$$

(8.4) LEMMA. *Suppose that, for all* $n \geqslant 0$,

(1) $$\beta_n^2 + \alpha_n^2 + \alpha_{n+1}\beta_n + \alpha_n\beta_{n-1} \leqslant \alpha^2$$

and

(2) $$\beta_n^2 + \alpha_{n+1}^2 + \alpha_{n+1}\beta_{n+1} + \alpha_n\beta_n \leqslant \alpha^2.$$

Then $\|A + B\| \leqslant \alpha$.

PROOF. Let $x = \sum_{n=0}^{\infty} \xi_n f_n \in F$. Then

$$(A + B)x = -\beta_0 \xi_1 f_0 + \sum_{n=1}^{\infty} (\xi_{2n-1}\alpha_n - \beta_n \xi_{2n+1}) f_{2n} + \sum_{n=0}^{\infty} (\xi_{2n}\beta_n - \xi_{2n+2}\alpha_{n+1}) f_{2n+1};$$

$$\begin{aligned}
\|(A + B)x\|^2 &= \beta_0^2 |\xi_1^2| + \sum_{n=1}^{\infty} |\xi_{2n-1}\alpha_n - \beta_n \xi_{2n+1}|^2 + \sum_{n=0}^{\infty} |\xi_{2n}\beta_n - \xi_{2n+2}\alpha_{n+1}|^2 \\
&= \beta_0^2 |\xi_1^2| + \sum_{n=1}^{\infty} \left(\alpha_n^2 |\xi_{2n-1}^2| + \beta_n^2 |\xi_{2n+1}^2| - 2\operatorname{Re}(\xi_{2n-1}\xi_{2n+1})\alpha_n\beta_n\right) \\
&\quad + \sum_{n=0}^{\infty} \left(\beta_n^2 |\xi_{2n}^2| + \alpha_{n+1}^2 |\xi_{2n+2}^2| - 2\operatorname{Re}(\xi_{2n}\xi_{2n+2})\alpha_{n+1}\beta_n\right) \\
&\leqslant \beta_0^2 |\xi_1^2| + \sum_{n=1}^{\infty} \left(\alpha_n^2 |\xi_{2n-1}^2| + \beta_n^2 |\xi_{2n+1}^2| + \alpha_n\beta_n\left(|\xi_{2n-1}^2| + |\xi_{2n+1}^2|\right)\right) \\
&\quad + \sum_{n=0}^{\infty} \left(\beta_n^2 |\xi_{2n}^2| + \alpha_{n+1}^2 |\xi_{2n+2}^2| + \alpha_{n+1}\beta_n\left(|\xi_{2n}^2| + |\xi_{2n+2}^2|\right)\right) \\
&= \sum_{n=0}^{\infty} |\xi_{2n}^2| \left(\beta_n^2 + \alpha_n^2 + \alpha_{n+1}\beta_n + \alpha_n\beta_{n-1}\right) \\
&\quad + \sum_{n=0}^{\infty} |\xi_{2n+1}^2| \left(\beta_n^2 + \alpha_{n+1}^2 + \alpha_{n+1}\beta_{n+1} + \alpha_n\beta_n\right) \\
&\leqslant \alpha^2 \sum_{n=0}^{\infty} |\xi_n|^2 = \alpha^2 \|x\|^2.
\end{aligned}$$

As $(A + B) \mid F^{\perp} = 0$, the result follows.

(8.5) LEMMA. *Suppose that* $\alpha > \alpha_1 > 0$, *and, for all* $n \geqslant 1$, *that*

(1) $$\sqrt{(\alpha_n \alpha)} \leqslant \alpha_{n+1} \leqslant (\alpha + \alpha_n)/2.$$

Set $\eta_n = \alpha - \alpha_n$. *Then*

(a) *the sequence* $(2^n \eta_n)$ *is bounded above and away from* 0;

(b) *if, for all* $n \geqslant 0$, $\beta_n \leqslant \eta_{n+1}$, *then* $\|A + B\| = \|A\| = \alpha$.

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

PROOF. Induction on (1) proves that, for $n \geqslant 1$,

$$\left(\alpha^{2^n-1}\alpha_1\right)^{2^{-n}} \leqslant \alpha_{n+1} \leqslant 2^{-n}\alpha_1 + (1-2^{-n})\alpha,$$

and, trivially, $(\alpha^{2^n-1}\alpha_1)^{2^{-n}} \geqslant \alpha/\{1 + 2^{-n}(\alpha/\alpha_1 - 1)\}$. Hence (a) follows. For (b), notice that (α_n) is strictly increasing, by (1), and therefore, for each $n \geqslant 1$,

$$\eta_{n+1}^2 + \alpha_n^2 + \alpha_{n+1}\eta_{n+1} + \alpha_n\eta_n = \alpha^2 - \alpha_{n+1}\alpha + \alpha_n\alpha < \alpha^2.$$

On the other hand,

$$\begin{aligned}\eta_{n+1}^2 + \alpha_{n+1}^2 + \alpha_{n+1}\eta_{n+2} + \alpha_n\eta_{n+1} & \\ &= 2\alpha_{n+1}^2 + \alpha_n\alpha - \alpha_{n+1}(\alpha + \alpha_{n+1} + \alpha_{n+2}) + \alpha^2 \\ &\leqslant 3\alpha_{n+1}^2 - 3\alpha_{n+1}\alpha_{n+2} + \alpha^2 \quad \text{(by (1))} \\ &< \alpha^2.\end{aligned}$$

Thus, if $0 \leqslant \beta_n \leqslant \eta_{n+1}$ for each $n \geqslant 0$, then (8.4) applies, and consequently $\|A + B\| \leqslant \alpha$. However,

$$\|(A+B)f_{2n+1}\| = \|\alpha_{n+1}f_{2n+2} - \beta_n f_{2n}\| \geqslant \alpha_{n+1} - \beta_n \to \alpha$$

as $n \to \infty$. Hence $\|A + B\| \geqslant \alpha$ also. If we take all the β_n to be zero, it follows that $\|A\| = \alpha$.

(8.6) Let $(\gamma_n)_{n \geqslant 1}$ be any strictly increasing sequence of positive real numbers converging to π; set $c_n = \cos\gamma_n$, and $s_n = \sin\gamma_n$. Thus, for each n, $s_n > 0$ and $c_n^2 + s_n^2 = 1$; moreover, as $n \to \infty$, $c_n \to -1$ and $s_n \to 0$. Define $T_0 \in \mathrm{U}(E, J)$ by the formulae

(1) $$T_0 \,|\, F^\perp = I \,|\, F^\perp, \quad T_0 f_0 = -f_0, \quad \text{and,} \quad \text{for } n \geqslant 1,$$
$$T_0 f_{2n-1} = c_n f_{2n-1} + s_n f_{2n},$$
$$T_0 f_{2n} = -s_n f_{2n-1} + c_n f_{2n}.$$

(Thus, by (6.5) and (8.2), T_0 cannot be joined to I by a minimising path in $\mathrm{U}(E, J)$.)

(8.7) LEMMA. *Suppose that, in* (8.3), $\beta_n > 0$ *for all* $n \geqslant 0$, *and the infinite product* $\prod_{n=1}^{\infty}\{\beta_n(1 - c_{n+1})/s_{n+1}\}$ *converges* (*to a nonzero limit*). *Then* -1 *is not an eigenvalue of* $(I - B)^{-1}(I + B)T_0 \,|\, F$.

PROOF. Let $x = \sum_{n=0}^{\infty}\xi_n f_n \in F$ satisfy $(I - B)^{-1}(I + B)T_0 x = -x$, or equivalently

(1) $$B(I - T_0)x = (I + T_0)x$$

(compare (5.6)(b)). Substituting in (1) and equating coefficients of f_0, of f_1, of f_{2n} (for $n \geqslant 1$), and of f_{2n+1} (for $n \geqslant 1$) in turn, we obtain, respectively,

(2) $$-\beta_0(s_1\xi_2 + (1 - c_1)\xi_1) = 0,$$

(3) $$2\beta_0\xi_0 = -s_1\xi_2 + (1 + c_1)\xi_1,$$

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press

(4) $$-\beta_n\left(s_{n+1}\xi_{2n+2} + (1 - c_{n+1})\xi_{2n+1}\right) = (1 + c_n)\xi_{2n} + s_n\xi_{2n-1},$$

(5) $$\beta_n\left((1 - c_n)\xi_{2n} - s_n\xi_{2n-1}\right) = -s_{n+1}\xi_{2n+2} + (1 + c_{n+1})\xi_{2n+1}.$$

As $\beta_0 > 0$, (2) and (3) yield readily that

(6) $$\xi_1 = \beta_0\xi_0, \qquad s_1\xi_2 + (1 - c_1)\xi_1 = 0.$$

Suppose inductively that, for some $n \geqslant 1$,

(7) $$s_n\xi_{2n} + (1 - c_n)\xi_{2n-1} = 0$$

(the case $n = 1$ is (6)). Then, as $1 - c_n^2 = s_n^2 \neq 0$,

$$(1 + c_n)\xi_{2n} + s_n\xi_{2n-1} = -(1 + c_n)\{(1 - c_n)/s_n\}\xi_{2n-1} + s_n\xi_{2n-1} = 0.$$

Since $\beta_n > 0$, it follows from (4) that

$$s_{n+1}\xi_{2n+2} + (1 - c_{n+1})\xi_{2n+1} = 0.$$

Ergo, equality (7) holds for all $n \geqslant 1$. Use it to substitute for the odd suffixes in (5), and simplify (recalling that $s_n^2 = 1 - c_n^2$). Thus, for $n \geqslant 1$,

(8) $$\xi_{2n+2} = -\beta_n s_{n+1}^{-1}(1 - c_{n+1}) \cdot \xi_{2n}.$$

This equality also holds for $n = 0$, by (6). Now the formulae (8) and (7) express all the coefficients ξ_n as nonzero scalar multiples of ξ_0. If $\xi_0 \neq 0$, then the convergence of $\prod\beta_n s_{n+1}^{-1}(1 - c_{n+1})$ implies that $|\xi_{2n}|$ has a positive limit as $n \to \infty$, by (8). This is absurd, and therefore $\xi_0 = 0$, and so $\xi_n = 0$ for all n; that is, $x = 0$. (Convergence of the infinite product is not really needed; it suffices that the partial products should not form a square-summable sequence.)

(8.8) Suppose now that $(\varepsilon_n)_{n=1}^{\infty}$ is a strictly increasing sequence of positive numbers with limit $\varepsilon < \pi$ such that the series $\sum_{n=1}^{\infty}2^n(\varepsilon - \varepsilon_n)$ converges. Suppose also that $T_1 \in \mathrm{U}(E, J)$ satisfies the following conditions: $N'(T_1) = \varepsilon$, $T_1(F) \subseteq F$, $\ker(T_1 + I)$ is of finite odd dimension, $T_1 f_0 = -f_0$, and, for $n \geqslant 1$,

$$T_1 f_{2n} = -\sin\varepsilon_n \cdot f_{2n-1} + \cos\varepsilon_n \cdot f_{2n},$$
$$T_1 f_{2n-1} = \cos\varepsilon_n \cdot f_{2n-1} + \sin\varepsilon_n \cdot f_{2n}.$$

Choose a strictly increasing sequence $(\alpha_n)_{n \geqslant 0}$ with $\alpha_0 = 0$ and $\alpha = \lim_{n\to\infty}\alpha_n = \pi - \varepsilon$, satisfying (8.5)(1) (for instance, $\alpha_n = \alpha(1 - 2^{-n})$). Define A as in (8.3)(1), and η_n as in (8.5). Suppose, for each n, that $\rho_n \in (0, 1]$, and that $\prod_{n=0}^{\infty}\rho_n$ converges; set $\beta_n = \rho_n\eta_{n+1}/2$ for each n, and define B by (8.3)(2).

If we set $\gamma_n = \alpha_n + \varepsilon_n \uparrow \pi$, then the operator T_0 of (8.6) is precisely $T_1 \cdot \exp A$.

(8.9) LEMMA. *The C^ω path in* $\mathrm{U}(E, J)$ *defined by*

$$p(t) = (I - tB)^{-1}(I + tB)T_1\exp(tA), \qquad 0 \leqslant t \leqslant 1,$$

has length $\alpha = \pi - \varepsilon$.

PROOF. Use (2.2) and differentiate in $\mathbf{L}(E, I)$ to obtain

$$\dot{p}(t) = (I - tB)^{-1}(I + tB)\left\{2B(I - t^2B^2)^{-1} + A\right\}T_1\exp(tA),$$

since $AT_1 = T_1A$, and since all the expressions in B commute. To the left and right of the bracket are isometries, so

$$\|\dot{p}(t)\| = \left\|2B(I - t^2B^2)^{-1} + A\right\|.$$

However, for $n \geqslant 0$,

$$2B(I - t^2B^2)^{-1}f_{2n} = 2\beta_n\left(1 + t^2\beta_n^2\right)^{-1}f_{2n+1}, \text{ and}$$

$$2B(I - t^2B^2)^{-1}f_{2n+1} = -2\beta_n\left(1 + t^2\beta_n^2\right)^{-1}f_{2n},$$

whilst, for $0 \leqslant t \leqslant 1$,

$$2\beta_n\left(1 + t^2\beta_n^2\right)^{-1} \leqslant 2\beta_n \leqslant \eta_{n+1}$$

by construction. Hence, by (8.5) (mutatis mutandis), $\|\dot{p}(t)\| = \alpha$, and the length of p is α, as stated.

(8.10) LEMMA. *T_1 may be joined to I by a minimising path.*

PROOF. Clearly $p(0) = T_1$, $p(1) = (I - B)^{-1}(I + B)T_0$, and $p(1)\,|\,F^{\perp} = p(0)\,|\,F^{\perp}$ (by definition of A and B). Now, in the notation of (8.6), and with $\theta_n = \varepsilon - \varepsilon_n$ for each n,

$$\begin{aligned}\beta_n s_{n+1}^{-1}(1 - c_{n+1}) &= \rho_n\eta_{n+1}/\left\{2\tan\left[(\pi - \gamma_{n+1})/2\right]\right\}\\ &= \rho_n\eta_{n+1}/\left\{(\theta_{n+1} + \eta_{n+1}) + O\left((\theta_{n+1} + \eta_{n+1})^3\right)\right\}\\ &= \rho_n\left\{1 - \eta_{n+1}^{-1}\theta_{n+1} + \eta_{n+1}^{-1}O\left((\theta_{n+1} + \eta_{n+1})^3\right)\right\}.\end{aligned}$$

By our hypotheses in (8.8), $\prod_{n=0}^{\infty}\rho_n$ and $\Sigma 2^n\theta_n$ both converge, whilst, by (8.5)(a), $2^n\eta_n$ is bounded away from 0. Hence $\prod\beta_n s_{n+1}^{-1}(1 - c_{n+1})$ converges; by (8.7), $p(1)\,|\,F$ does not have -1 as an eigenvalue, and consequently

$$\ker(p(1) + I) = F^{\perp} \cap \ker(p(1) + I) = F^{\perp} \cap \ker(T_1 + I)$$

has, because of f_0, dimension less by one than that of $\ker(T_1 + I)$. Thus $p(1)$ may be joined to I by a minimising path, and $d(I, p(1)) \leqslant \pi$ (by (7.1), (7.2)). However, $d(T_1, p(1)) \leqslant \alpha$ by (8.9), and $d(T_1, I) = 2\pi - \varepsilon = \pi + \alpha$ by (6.5). Therefore $d(T_1, p(1)) = \alpha$ and $d(I, p(1)) = \pi$, and a minimising path between T_1 and I may be obtained by taking a minimising segment from $p(1)$ to I and adjoining it to p.

NOTE. That $N'(p(1)) = \pi$, which we have (in effect) proved by distances (compare (6.8)), results also from perturbation theory, since B is compact and $N'(T_0) = \pi$.

(8.11) We can now complete the proof of (8.1). Suppose, then, that $T \in \mathbf{U}(E, J)$, that $N'(T) = \nu \in (0, \pi)$, and that the projection $\mathbf{P}_T(\{z \in S\colon \delta(z,1) \geqslant N'(T)\})$ is of odd finite rank. By the definition of $N'(T)$, $\mathbf{P}_T(\{z \in S\colon \sigma < \delta(z,1) < \nu\})$ is of infinite rank for any $\sigma < \nu$ (see (3.2)). Using this fact, choose inductively a strictly decreasing sequence $(\kappa_n)_{n \geqslant 0}$ of positive real numbers such that

(1) $$\kappa_0 < 2\nu(1 - \nu/\pi),$$

(2) $$\sum_{n=0}^{\infty} 2^n \kappa_n < \infty,$$

(for instance one might take $\kappa_n \leqslant 2^{-2n}\kappa_0$), and such that, for $n \geqslant 0$,

(3) $$\mathbf{P}_T(\{z \in S\colon \nu - \kappa_n \leqslant \delta(z,1) < \nu - \kappa_{n+1}\}) \neq 0.$$

The choice of κ_{n+1}, given κ_n, is always possible, but the spectral decomposition of T may make one of (2), (3) redundant. Now set $\nu_n = \nu - \kappa_n$, and choose τ such that

(4) $$\kappa_0(\nu + \nu_0)^{-1} \leqslant \tau < \pi/\nu - 1,$$

which is possible because of (1).

Define a function $h\colon (-\nu, +\nu) \to \mathbf{R}$ as follows: when $0 \leqslant \xi < \nu_0$, set $h(\xi) = 0$. For $n \geqslant 0$, $\nu_n \leqslant \xi < \nu_{n+1}$, set

(5) $$h(\xi) = \{(1 + \tau)\nu_n - \xi\}/\tau.$$

Finally, set

(6) $$h(-\xi) = -h(\xi).$$

By (4),

(7) $$-\nu \leqslant h(\xi) \leqslant \nu \quad \text{for all } \xi \in (-\nu, +\nu).$$

Set

(8) $$D = \int_{\nu_0 \leqslant \delta(z,1) < \nu} ih(-i \log z)\mathbf{P}_T(dz)$$

(recall (3.3)). Then D is skew-adjoint and real (by (6), cf. (6.9)(a)); by (7), $\sigma(D)$ is included in the interval $[-i\nu, +i\nu]$ of the imaginary axis, so that $\|D\| \leqslant \nu$. Furthermore, (5) gives $h(\nu_n) = \nu_n$ for each n; hence $i\nu_n \in \sigma(D)$, and $\|D\| = \nu$ exactly. Take the geodesic q in $\mathbf{U}(E, J)$ with

(9) $$q(t) = T\exp(tD), \qquad 0 \leqslant t \leqslant \tau.$$

As usual (for instance in (6.9)(a)),

(10) $$\ell(q) = \tau\|D\| = \tau\nu.$$

Write $T_1 = q(\tau)$, and consider its spectral decomposition. If we write F_0 for the image of

$$\mathbf{P}_T(\{z \in S\colon \delta(z,1) \geqslant \nu \text{ or } \delta(z,1) < \nu_0\}),$$

and, for $n \geqslant 0$, if we let F_{n+1} denote the image of

$$\mathbf{P}_T(\{z \in S\colon \nu_n \leqslant \delta(z,1) < \nu_{n+1}\}),$$

then $T_1 \,|\, F_0 = T \,|\, F_0$ (as $D \,|\, F_0 = 0$), and, by (5), (8), (9),

$$\sigma(T_1 \,|\, F_{n+1}) = \{\exp(\pm i(1+\tau)\nu_n)\}.$$

So $\sigma(T_1 \,|\, F_0^{\perp})$ consists only of eigenvalues $\exp(\pm i(1+\tau)\nu_n)$ and their cluster points $\exp(\pm i(1+\tau)\nu)$. By (3.4), then,

$$N'(T_1) = (1+\tau)\nu,$$

and, by (4), $(1+\tau)\nu < \pi$. Thus -1 remains an eigenvalue for T_1 of the same odd finite multiplicity as for T, and in fact with the same eigenvectors (all in F_0). From (6.5),

$$(11) \qquad d(T_1, I) = 2\pi - (1+\tau)\nu = d(T,I) - \ell(q),$$

in view of (10).

(8.12) Take a real unit vector $f_0 \in \ker(T_1 + I) \subseteq F_0$, and, for $n \geqslant 1$, let e_{2n} be a unit $\exp\{i(1+\tau)\nu_n\}$-eigenvector of $T_1 \,|\, F_{n+1}$, so that Je_{2n} is a unit $\exp\{-i(1+\tau)\nu_n\}$-eigenvector. For $n \geqslant 1$, set

$$f_{2n-1} = (e_{2n} + Je_{2n})/\sqrt{2}, \quad f_{2n} = (e_{2n} - Je_{2n})/(i\sqrt{2}),$$

and let F be the closed linear span of $(f_n)_{n \geqslant 0}$. Define $\varepsilon_n = (1+\tau)\nu_n$, and $\varepsilon = (1+\tau)\nu$. Then the data T_1, (f_n), (ε_n), ε satisfy all the conditions of (8.8), by virtue of (8.11)(2); by (8.10), there is a minimising path from T_1 to I, and by (8.11)(11) this path and q together will form a minimising path from T to I. This completes the proof of (8.1). Note that not only were various numerical values to some extent arbitrary, but also the forms of the path p and of the operator B were chosen merely for simplicity.

References

[1] J. F. Adams, *Lectures on Lie Groups* (Benjamin, New York-Amsterdam, 1969).
[2] C. Chevalley, *Theory of Lie Groups* I (Princeton Univ. Press, Princeton, N. J., 1946).
[3] J. Eells, 'A setting for global analysis', *Bull. Amer. Math. Soc* **72** (1966), 751–807.
[4] P. de la Harpe, *Classical Banach-Lie Algebras and Banach-Lie Groups of Operators in Hilbert Space* (Lecture Notes in Mathematics, Vol. 285, Springer-Verlag, Berlin-Heidelberg-New York, 1972).

[5] B. Maissen, 'Lie-Gruppen mit Banachräumen als Parameterräume', *Acta Math.* **108** (1962), 229–270.
[6] R. S. Palais, 'On the homotopy type of certain groups of operators', *Topology* **5** (1966), 1–16.
[7] R. S. Palais, 'Lusternik-Schnirelman theory on Banach manifolds', *Topology* **5** (1966), 115–132.
[8] C. R. Putnam and A. Wintner, 'The orthogonal group in Hilbert space', *Amer. J. Math.* **74** (1952), 52–78.

Department of Mathematics
Victoria University of Wellington
Private Bag
Wellington
New Zealand

https://doi.org/10.1017/S1446788700028202 Published online by Cambridge University Press