Spring, 2011 CIS 610

Advanced Geometric Methods in Computer Science
Jean Gallier

Some Solution for Homework 1

April 5, 2011

"B problems" must be turned in.

Problem B1 (40 pts). (a) Given a rotation matrix

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

where $0 < \theta < \pi$, prove that there is a skew-symmetric matrix B such that

$$R = (I - B)(I + B)^{-1}.$$

(b) If B is a skew-symmetric $n \times n$ matrix, prove that $\lambda I_n - B$ and $\lambda I_n + B$ are invertible for all $\lambda \neq 0$, and that they commute.

(c) Prove that

$$R = (\lambda I_n - B)(\lambda I_n + B)^{-1}$$

is a rotation matrix that does not admit -1 as an eigenvalue. (Recall, a rotation is an orthogonal matrix R with positive determinant, i.e., $\det(R) = 1$.)

(d) Given any rotation matrix R that does not admit -1 as an eigenvalue, prove that there is a skew-symmetric matrix B such that

$$R = (I_n - B)(I_n + B)^{-1} = (I_n + B)^{-1}(I_n - B).$$

This is known as the *Cayley representation* of rotations (Cayley, 1846).

(e) Given any rotation matrix R, prove that there is a skew-symmetric matrix B such that

$$R = \left((I_n - B)(I_n + B)^{-1}\right)^2.$$

Solutions.

B1(a). We are looking for a matrix

$$B = \begin{pmatrix} 0 & -a \\ a & 0 \end{pmatrix}$$

such that

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = (I - B)(I + B)^{-1}.$$

A simple computation shows that

$$(I + B)^{-1} = \begin{pmatrix} 1 & -a \\ a & 1 \end{pmatrix}^{-1} = \frac{1}{1 + a^2} \begin{pmatrix} 1 & a \\ -a & 1 \end{pmatrix}.$$

It follows that

$$(I - B)(I + B)^{-1} = \frac{1}{1 + a^2} \begin{pmatrix} 1 & a \\ -a & 1 \end{pmatrix} \begin{pmatrix} 1 & a \\ -a & 1 \end{pmatrix} = \frac{1}{1 + a^2} \begin{pmatrix} 1 - a^2 & 2a \\ -2a & 1 - a^2 \end{pmatrix},$$

and thus, we must have

$$\cos\theta = \frac{1 - a^2}{1 + a^2}$$
$$\sin\theta = \frac{-2a}{1 + a^2}.$$

In view of the well-known identities

$$\cos\theta = \frac{1 - \tan^2(\theta/2)}{1 + \tan^2(\theta/2)}$$
$$\sin\theta = \frac{2\tan(\theta/2)}{1 + \tan^2(\theta/2)},$$

since $0 \leq \theta < \pi$, we obtain the solution

$$a = -\tan(\theta/2).$$

B1(b). We proved in class that the eigenvalues of a skew-symmetric matrix are pure immaginary or 0, in other words, of the form $i\mu$, where $\mu \in \mathbb{R}$. If e is any eigenvector of B for the eigenvalue $i\mu$, then e is an eigenvector of $\lambda I + B$ for the eigenvalue $\lambda + i\mu$, and e is an eigenvector of $\lambda I - B$ for the eigenvalue $\lambda - i\mu$. Observe that $\lambda + i\mu \neq 0$ and $\lambda - i\mu \neq 0$, since $\lambda \neq 0$. Therefore, all the eigenvalues of $\lambda I + B$ are nonzero, and similarly all the eigenvalues of $\lambda I - B$ are nonzero, which implies that $\lambda I + B$ and $\lambda I - B$ are invertible (if $\lambda \neq 0$).

We have

$$(\lambda I - B)(\lambda I + B) = \lambda^2 I + \lambda B - \lambda B - B^2 = \lambda^2 I - B^2$$
$$(\lambda I + B)(\lambda I - B) = \lambda^2 I - \lambda B + \lambda B - B^2 = \lambda^2 I - B^2,$$

which proves that

$$(\lambda I - B)(\lambda I + B) = (\lambda I + B)(\lambda I - B).$$

(B1c). If e is an eigenvector of $\lambda I + B$ for the eigenvalue $\lambda + i\mu$, then e is an eigenvector of $(\lambda I + B)^{-1}$ for the eigenvalue $(\lambda + i\mu)^{-1}$, and since e is also an eigenvector of $\lambda I - B$ for $\lambda - i\mu$, we deduce that e is an eigenvector of $(\lambda I - B)(\lambda I + B)^{-1}$ for the eigenvalue $(\lambda - i\mu)(\lambda + i\mu)^{-1}$. Now,

$$(\lambda - i\mu)(\lambda + i\mu)^{-1} = \frac{1}{\lambda^2 + \mu^2}(\lambda - i\mu)^2 = \frac{\lambda^2 - \mu^2}{\lambda^2 + \mu^2} + i\frac{2\lambda\mu}{\lambda^2 + \mu^2}.$$

Observe that the above complex number is never equal to -1, since $\lambda \neq 0$, so -1 is not an eigenvalue of $R = (\lambda I - B)(\lambda I + B)^{-1}$.

We have

$$\begin{aligned} R^\top &= \big((\lambda I + B)^{-1}\big)^\top (\lambda I - B)^\top \\ &= (\lambda I + B^\top)^{-1}(\lambda I - B^\top) \\ &= (\lambda I - B)^{-1}(\lambda I + B), \end{aligned}$$

using the fact that B is skew-symmetric. On the other hand,

$$R^{-1} = \big((\lambda I - B)(\lambda I + B)^{-1}\big)^{-1} = (\lambda I + B)(\lambda I - B)^{-1}.$$

In order to prove that $R^\top = R^{-1}$, we have to prove that $\lambda I + B$ and $(\lambda I - B)^{-1}$ commute. However, we know from B1(b) that $\lambda I + B$ and $\lambda I - B$ commute and we can use the following simple fact:

Fact. In a monoid, if $ab = ba$ and b is invertible, then $ab^{-1} = b^{-1}a$.

This is because

$$\begin{aligned} a &= a \\ a &= abb^{-1} \\ a &= bab^{-1} \\ b^{-1}a &= ab^{-1}. \end{aligned}$$

Applying this to $a = \lambda I + B$ and $b = \lambda I - B$, we conclude that $\lambda I + B$ and $(\lambda I - B)^{-1}$ commute, and so $R^\top = R^{-1}$. Therefore, $R = (\lambda I - B)(\lambda I + B)^{-1}$ is an orthogonal matrix, and since we showed that it does not admit -1 as an eigenvalue, it is a rotation matrix.

B1(d). Given a rotation matrix R that does not admit -1 as an eigenvalue, if a skew-symmetric matrix B exists such that

$$R = (I - B)(I + B)^{-1},$$

then we must have

$$R(I + B) = I - B,$$

which yields

$$R + RB = I - B,$$

and then

$$(I + R)B = I - R.$$

Now, since R is a rotation matrix, its eigenvalues are of the form $e^{i\theta}$, and so the eigenvalues of $I + R$ are of the form $1 + e^{i\theta}$. Since we assumed that -1 is not an eigenvalue of R, $1 + e^{i\theta} \neq 0$, and $I + R$ is invertible. It follows that B is uniquely defined by

$$B = (I + R)^{-1}(I - R).$$

Obviously,

$$(I + R)(I - R) = (I - R)(I + R) = I - R^2,$$

and since $I + R$ is invertible, by the Fact proved in B1(c), we get

$$B = (I + R)^{-1}(I - R) = (I - R)(I + R)^{-1}.$$

It remains to prove that B is skew-symmetric. We have

$$\begin{aligned}
B^\top &= (I - R)^\top \big((I + R)^{-1}\big)^\top \\
&= (I - R^\top)(I + R^\top)^{-1} \\
&= (R - I)R^\top \big((R + I)R^\top\big)^{-1} \\
&= (R - I)R^\top R(R + I)^{-1} \\
&= -(I - R)(I + R)^{-1} \\
&= -B,
\end{aligned}$$

as desired. Since we know from B1(c) (with $\lambda = 1$) that $(I-B)(I+B)^{-1}$ is a rotation matrix that does not admit -1 as an eigenvalue, we just proved that for every rotation matrix R that does not admit -1 as an eigenvalue, there is a unique skew-symmetric matrix B such that $R = (I - B)(I + B)^{-1}$, and B is given by the same formula,

$$B = (I - R)(I + R)^{-1}.$$

B1(e). Recall that for every orthogonal matrix $R \in \mathbf{O}(n)$, there is an orthogonal matrix P and a block diagonal matrix D such that $R = PD\,P^\top$, where D is of the form

$$D = \begin{pmatrix} D_1 & & \cdots & \\ & D_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & D_p \end{pmatrix}$$

such that each block D_i is either 1, -1, or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

where $0 < \theta_i < \pi$.

In particular, if R is a rotation matrix ($R \in \mathbf{SO}(n)$), then R has an even number of eigenvalues -1. So, they can be grouped into two-dimensional rotation matrices of the form

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

i.e., we allow $\theta_i = \pi$, and we may assume that D does not contain one-dimensional blocks of the form -1.

Now, for every two-dimensional rotation matrix

$$T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with $0 < \theta \leq \pi$, observe that

$$T^{\frac{1}{2}} = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}$$

does not admit -1 as an eigenvalue (since $0 < \theta/2 \leq \pi/2$) and $T = \left(T^{\frac{1}{2}}\right)^2$. Thus, if we form the matrix $R^{\frac{1}{2}}$ by replacing each two-dimensional block D_i in the above normal form by $D_i^{\frac{1}{2}}$, we obtain a rotation matrix that does not admit -1 as an eigenvalue, $R = \left(R^{\frac{1}{2}}\right)^2$ and the Cayley transform of $R^{\frac{1}{2}}$ is well defined.

Problem B2 (40). (a) Consider the map, $f\colon \mathbf{GL}^+(n) \to \mathbf{S}(n)$, given by

$$f(A) = A^\top A - I.$$

Check that

$$df(A)(H) = A^\top H + H^\top A,$$

for any matrix, H.

(b) Consider the map, $f\colon \mathbf{GL}(n) \to \mathbb{R}$, given by

$$f(A) = \det(A).$$

Prove that $df(I)(B) = \mathrm{tr}(B)$, the trace of B, for any matrix B (here, I is the identity matrix). Then, prove that

$$df(A)(B) = \det(A)\mathrm{tr}(A^{-1}B),$$

where $A \in \mathbf{GL}(n)$.

(c) Use the map $A \mapsto \det(A) - 1$ to prove that $\mathbf{SL}(n)$ is a manifold of dimension $n^2 - 1$.

(d) Let J be the $(n+1) \times (n+1)$ diagonal matrix

$$J = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix}.$$

We denote by $\mathbf{SO}(n,1)$ the group of real $(n+1) \times (n+1)$ matrices

$$\mathbf{SO}(n,1) = \{A \in \mathbf{GL}(n+1) \mid A^\top J A = J \quad \text{and} \quad \det(A) = 1\}.$$

Check that $\mathbf{SO}(n,1)$ is indeed a group with the inverse of A given by $A^{-1} = JA^\top J$ (this is the *special Lorentz group*.) Consider the function $f\colon \mathbf{GL}^+(n+1) \to \mathbf{S}(n+1)$, given by

$$f(A) = A^\top J A - J,$$

where $\mathbf{S}(n+1)$ denotes the space of $(n+1) \times (n+1)$ symmetric matrices. Prove that

$$df(A)(H) = A^\top J H + H^\top J A$$

for any matrix, H. Prove that $df(A)$ is surjective for all $A \in \mathbf{SO}(n,1)$ and that $\mathbf{SO}(n,1)$ is a manifold of dimension $\frac{n(n+1)}{2}$.

Solutions.

B2(a). If $f(A) = A^\top A - I$, we have

$$\begin{aligned} f(A+H) - f(A) - (A^\top H + H^\top A) &= (A+H)^\top (A+H) - I - (A^\top A - I) - A^\top H - H^\top A \\ &= (A^\top + H^\top)(A+H) - I - A^\top A + I - A^\top H - H^\top A \\ &= A^\top A + A^\top H + H^\top A + H^\top H - A^\top A - A^\top H - H^\top A \\ &= H^\top H. \end{aligned}$$

It follows that

$$\epsilon(H) = \frac{f(A+H) - f(A) - (A^\top H + H^\top A)}{\|H\|} = \frac{H^\top H}{\|H\|},$$

and since it is clear that

$$\lim_{X \mapsto 0} \epsilon(H) = 0,$$

we conclude that

$$df(A)(H) = A^\top H + H^\top A,$$

for all A and H.

B2(b). Recall the following result from linear algebra: If A is any square matrix, then

$$\det(\lambda I + A) = \lambda^n + \tau_1(A)\lambda^{n-1} + \cdots + \tau_k(A)\lambda^{n-k} + \cdots + \tau_n(A),$$

with

$$\tau_1(A) = \mathrm{tr}(A) \quad \text{and} \quad \tau_n(A) = \det(A).$$

In fact, $\tau_k(A)$ can be expressed as the sum of $k \times k$ determinants corresponding to principal diagonal $k \times k$ minors of A. For any nonempty subset $I \subseteq \{1, \ldots, n\}$, say $I = \{i_1, \ldots, i_k\}$, let $A_{I,I}$ be the $k \times k$ submatrix of A whose jth column consists of the elements $a_{i_h\, i_j}$, where $h = 1, \ldots, k$. Then, we have

$$\tau_k(A) = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \det(A_{I,I}).$$

Since $\mathbb{R}^{n^2}$ is finite-dimensional, it doesn't matter which norm we pick, so we can pick

$$\|A\|_\infty = \max\{|a_{ij}| \mid 1 \leq i, j \leq n\}.$$

Actually, if $\|A\|_F$ denotes the Froebenius norm (where $\|A\|_F = \sqrt{\mathrm{tr}(A^\top A)}$), observe that

$$\|A\|_F \leq n \|A\|_\infty \quad \text{and} \quad \|A\|_\infty \leq \|A\|_F,$$

so we can use $\|A\|_F$. Then, we see that for $k \geq 2$, since the terms in all the $\det(A_{I,I})$ are at least quadratic,

$$\lim_{A \mapsto 0} \frac{\tau_k(A)}{\|A\|} = 0.$$

Now, setting $\lambda = 1$, we have

$$\begin{aligned} f(I + B) - f(I) - \mathrm{tr}(B) &= \det(I + B) - \det(I) - \mathrm{tr}(B) \\ &= 1 + \sum_{k=1}^{n} \tau_k(B) - 1 - \mathrm{tr}(B) \\ &= \sum_{k=2}^{n} \tau_k(B). \end{aligned}$$

Consequently,

$$\epsilon(B) = \frac{f(I + B) - f(I) - \mathrm{tr}(B)}{\|B\|} = \frac{\sum_{k=2}^{n} \tau_k(B)}{\|B\|},$$

and from pour previous observation,

$$\lim_{B \mapsto 0} \epsilon(B) = 0,$$

which proves that

$$df(I)(B) = \mathrm{tr}(B).$$

If A is invertible, then

$$f(A + B) = \det(A + B) = \det(A(I + A^{-1}B)) = \det(A) \det(I + A^{-1}B)$$

and we get

$$\begin{aligned} f(A+B) - f(A) - \det(A)\mathrm{tr}(A^{-1}B) &= \det(A)\det(I + A^{-1}B) - \det(A) - \det(A)\mathrm{tr}(A^{-1}B) \\ &= \det(A)\Big(1 + \sum_{k=1}^{n} \tau_k(A^{-1}B)\Big) - \det(A) \\ &\quad - \det(A)\mathrm{tr}(A^{-1}B) \\ &= \det(A)\Big(\sum_{k=2}^{n} \tau_k(A^{-1}B)\Big). \end{aligned}$$

If we write

$$\epsilon(B) = \frac{f(A+B) - f(A) - \det(A)\mathrm{tr}(A^{-1}B)}{\|B\|} = \det(A)\frac{\sum_{k=2}^{n} \tau_k(A^{-1}B)}{\|B\|},$$

the same reasoning as before shows that

$$\lim_{B \mapsto 0} \epsilon(B) = 0,$$

which proves that

$$df(A)(B) = \det(A)\mathrm{tr}(A^{-1}B).$$

B2(c). Consider the map $g\colon \mathbf{GL}(n) \to \mathbb{R}$ given by

$$g(A) = \det(A) - 1.$$

Obviously $g^{-1}(0) = \mathbf{SL}(n)$, so to prove that $\mathbf{SL}(n)$ is a manifold is suffices to prove that $g'(A)$ is surjective for every $A \in \mathbf{SL}(n)$. However, because 1 is a constant, by B2(b), we have

$$dg(A)(B) = \det(A)\mathrm{tr}(A^{-1}B),$$

for all $A \in \mathbf{GL}(n)$ and all B. The linear map $dg(A)$ is surjective for all $A \in \mathbf{SL}(n)$ because for every $\lambda \in \mathbb{R}$, if we let $B = (\lambda/(n\det(A))A$, then

$$dg(A)(B) = \det(A)\mathrm{tr}(A^{-1}(\lambda/(n\det(A))A) = \lambda.$$

Since $\mathbf{GL}(n)$ has dimension n^2 and $\mathbb{R}$ has dimension 1, the manifold $\mathbf{SL}(n)$ has dimension $n^1 - 1$.

B2(d). If $A, B \in \mathbf{SO}(n,1)$, then $A^\top JA = J$, $B^\top JB = J$, $\det(A) = 1$, and $\det(B) = 1$, which implies

$$\det(AB) = \det(A)\det(B) = 1$$

and

$$(AB)^\top J(AB) = B^\top A^\top JAB = B^\top JB = 1.$$

Observe that J is symmetric and that $J^2 = I$. If $A \in \mathbf{SO}(n,1)$, then $A^\top JA = J$ and $\det(A) = 1$, so A is invertible and by multiplying both sides of $A^\top JA = J$ on the left by J, since $J^2 = I$, we get

$$JA^\top JA = I,$$

which implies

$$A^{-1} = JA^\top J.$$

Now,

$$(A^{-1})^\top JA^{-1} = (JA^\top J)^\top JA^{-1} = JAJJA^{-1} = JAA^{-1} = J.$$

We also have

$$\det(A^{-1}) = \det(JA^\top J) = \det(J)\det(A^\top)\det(J) = \det(J^2)\det(A) = 1,$$

which proves that $A^{-1} = JA^\top J \in \mathbf{SO}(n,1)$. Therefore, $\mathbf{SO}(n,1)$ is indeed a group.

Consider the function $f\colon \mathbf{GL}^+(n+1) \to \mathbf{S}(n+1)$, given by

$$f(A) = A^\top JA - J,$$

where $\mathbf{S}(n+1)$ denotes the space of $(n+1)\times(n+1)$ symmetric matrices. We have

$$\begin{aligned} f(A+H) - f(A) - (A^\top JH + H^\top JA) &= (A+H)^\top J(A+H) - J - (A^\top JA - J) \\ &\quad - (A^\top JH + H^\top JA) \\ &= (A^\top + H^\top)J(A+H) - A^\top JA - A^\top JH - H^\top JA \\ &= A^\top JA + A^\top JH + H^\top JA + H^\top JH - A^\top JA \\ &\quad - A^\top JH - H^\top JA \\ &= H^\top JH. \end{aligned}$$

If we write

$$\epsilon(H) = \frac{f(A+H) - f(A) - (A^\top JH + H^\top JA)}{\|H\|} = \frac{H^\top JH}{\|H\|},$$

then it is clear that

$$\lim_{H \mapsto 0} \epsilon(H) = 0,$$

which proves that

$$df(A)(H) = A^\top JH + H^\top JA$$

for all $A \in \mathbf{GL}^+(n+1)$ and all H.

Let us prove that $df(A)$ is surjective for all $A \in \mathbf{GL}^+(n+1)$. For any $n \times n$ symmetric matrix S, if we let

$$H = \frac{AJS}{2},$$

then we get

$$\begin{aligned} df(A)(H) &= A^\top JH + H^\top JA \\ &= A^\top J\left(\frac{AJS}{2}\right) + \left(\frac{AJS}{2}\right)^\top JA \\ &= \frac{1}{2}\left(A^\top JAJS + S^\top JA^\top JA\right) \\ &= \frac{1}{2}\left(JJS + S^\top JJ\right) \\ &= \frac{1}{2}\left(S + S^\top\right) \\ &= S, \end{aligned}$$

since S is symmetric and since $A^\top JA = J$ (because $A \in \mathbf{SO}(n,1)$.

Obviously,

$$f^{-1}(0) = \mathbf{SO}(n,1),$$

and since we just proved that $df(A)$ is surjective for all $A \in \mathbf{SO}(n,1)$, it follows that $\mathbf{SO}(n,1)$ is a manifold of dimension

$$(n+1)^2 - \frac{(n+1)(n+2)}{2} = \frac{(n+1)(2(n+1)-(n+2))}{2} = \frac{n(n+1)}{2},$$

since $\mathbf{GL}^+(n+1)$ has dimension $(n+1)^2$ (as an open subset of $\mathbf{GL}(n+1)$, which has dimension $(n+1)^2$), and $\mathbf{S}(n+1)$ has dimension $(n+1)(n+2)/2$.

Problem B3 (40 pts). (a) Given any matrix

$$B = \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \in \mathfrak{sl}(2,\mathbb{C}),$$

if $\omega^2 = a^2 + bc$ and ω is any of the two complex roots of $a^2 + bc$, prove that if $\omega \neq 0$, then

$$e^B = \cosh\omega\, I + \frac{\sinh\,\omega}{\omega}\, B,$$

and $e^B = I + B$, if $a^2 + bc = 0$. Observe that $\mathrm{tr}(e^B) = 2\cosh\,\omega$.

Prove that the exponential map, $\exp\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathbf{SL}(2,\mathbb{C})$, is *not* surjective. For instance, prove that

$$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

is not the exponential of any matrix in $\mathfrak{sl}(2,\mathbb{C})$.

(b) Recall that a matrix, N, is *nilpotent* iff there is some $m \geq 0$ so that $N^m = 0$. Let A be any $n \times n$ matrix of the form $A = I - N$, where N is nilpotent. Why is A invertible?

prove that there is some B so that $e^B = I - N$ as follows: Recall that for any $y \in \mathbb{R}$ so that $|y-1|$ is small enough, we have

$$\log(y) = -(1-y) - \frac{(1-y)^2}{2} - \cdots - \frac{(1-y)^k}{k} - \cdots .$$

As N is nilpotent, we have $N^m = 0$, where m is the smallest integer with this propery. Then, the expression

$$B = \log(I-N) = -N - \frac{N^2}{2} - \cdots - \frac{N^{m-1}}{m-1}$$

is well defined. Use a formal power series argument to show that

$$e^B = A.$$

We denote B by $\log(A)$.

(c) Let $A \in \mathbf{GL}(n, \mathbb{C})$. Prove that there is some matrix, B, so that $e^B = A$. Thus, the exponential map, $\exp\colon \mathfrak{gl}(n, \mathbb{C}) \to \mathbf{GL}(n, \mathbb{C})$, is surjective.

First, use the fact that A has a Jordan form, PJP^{-1}. Then, show that finding a log of A reduces to finding a log of every Jordan block of J. As every Jordan block, J, has a fixed nonzero constant, λ, on the diagonal, with 1's immediately above each diagonal entry and zero's everywhere else, we can write J as $(\lambda I)(I - N)$, where N is nilpotent. Find B_1 and B_2 so that $\lambda I = e^{B_1}$, $I - N = e^{B_2}$, and $B_1 B_2 = B_2 B_1$. Conclude that $J = e^{B_1 + B_2}$.

Solutions. B3(a). Given any matrix

$$B = \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \in \mathfrak{sl}(2, \mathbb{C}),$$

we have

$$B^2 = \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \begin{pmatrix} a & b \\ c & -a \end{pmatrix} = \begin{pmatrix} a^2 + bc & 0 \\ 0 & a^2 + bc \end{pmatrix} = (a^2 + bc) I.$$

If $a^2 + bc = 0$, then

$$e^B = I + B.$$

Otherwise, if we let ω be any complex square root of $a^2 + bc \neq 0$, so that $\omega^2 = a^2 + bc$, then $B^2 = \omega^2 I$ and so

$$e^B = I + \frac{B}{1!} + \frac{\omega^2}{2!} I + \frac{\omega^2}{3!} B + \frac{\omega^4}{4!} I + \frac{\omega^4}{5!} B + \frac{\omega^6}{6!} I + \frac{\omega^6}{6!} B + \cdots .$$

Rearranging the order of the terms, we get

$$e^B = \left(1 + \frac{\omega^2}{2!} + \frac{\omega^4}{4!} + \frac{\omega^6}{6!} + \cdots\right) I + \frac{1}{\omega}\left(\omega + \frac{\omega^3}{3!} + \frac{\omega^5}{5!} + \frac{\omega^7}{7!} + \cdots\right) B.$$

We recognize the power series for $\cosh\omega$ and $\sinh\omega$, so we obtain the equation

$$e^B = \cosh\omega I + \frac{\sinh\omega}{\omega}B.$$

Since $\operatorname{tr}(B) = 0$, we have

$$\operatorname{tr}(e^B) = 2\cosh\omega.$$

Recall that

$$\cosh(x+iy) = \cosh x\cos y + i\sinh x\sin y,$$

where $x, y \in \mathbb{R}$. It follows that $\cosh(x+iy)$ is real if either $x = 0$ or $y = 0$. Since $\cosh x = (e^x + e^{-x})/2$, if $y = 0$, then $\cosh x \geq 1$ for all $x \in \mathbb{R}$ and if $x = 0$, then $\cosh iy = \cos y \geq -1$ for all $y \in \mathbb{R}$. Therefore, if $\cosh\omega$ is real, then its minimum is -1 and it is achieved for $y = \pi + 2k\pi$ ($k \in \mathbb{Z}$). But then, we have

$$\sinh\omega = \sinh i\pi = i\sin\pi = 0.$$

We claim that there is no matrix $B \in \mathfrak{sl}(2,\mathbb{C})$ so that

$$e^B = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix},$$

or equivalently, such that

$$\cosh\omega I + \frac{\sinh\omega}{\omega}B = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

This is because the trace of the matrix on the right-hand side is -2, so we should have

$$\operatorname{tr}(e^B) = 2\cosh\omega = -2,$$

which implies that $\cosh\omega = -1$. However, we just showed that in this case $\sinh\omega = 0$, and then

$$e^B = -2I \neq \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

Therefore, the exponential map $\exp\colon \mathfrak{sl}(2,\mathbb{C}) \to \mathbf{SL}(2,\mathbb{C})$ is *not* surjective.

B3(b). A rigorous solution to this problem turns out to be a lot harder than I thought!

Recall that for every invertible matrix, P, and every matrix, A,

$$e^{PAP^{-1}} = Pe^AP^{-1}$$

and that for every block diagonal matrix,

$$A = \begin{pmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_m \end{pmatrix},$$

we have

$$e^A = \begin{pmatrix} e^{A_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{A_m} \end{pmatrix}.$$

Consequenly, the problem of finding the logarithm of a matrix reduces to the problem of finding the logarithm of a Jordan block $J_r(\alpha)$ with $\alpha \neq 0$, where

$$J_r(\alpha) = \begin{pmatrix} \alpha & 1 & 0 & \cdots & 0 \\ 0 & \alpha & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \alpha \end{pmatrix},$$

where $\alpha \in \mathbb{C}$, with $J_1(\alpha) = (\alpha)$ if $r = 1$. However, every such Jordan block, $J_r(\alpha)$, can be written as

$$J_r(\alpha) = \alpha I + H = \alpha I(I + \alpha^{-1}H),$$

where H is the nilpotent matrix of index of nilpotency, r, given by

$$H = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Furthermore, it is obvious that $N = \alpha^{-1}H$ is also nilpotent of index of nilpotency, r, and we have

$$J_r(\alpha) = \alpha I(I + N).$$

Logarithms of the diagonal matrix, αI, are easily found. If we write $\alpha = \rho e^{i\theta}$ where $\rho > 0$, then $\log \alpha = \log \rho + i(\theta + 2\pi h)$, for any $h \in \mathbb{Z}$, and we can pick a logarithm of αI to be

$$S = \begin{pmatrix} \log \rho + i\theta & 0 & \cdots & 0 \\ 0 & \log \rho + i\theta & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \log \rho + i\theta \end{pmatrix}.$$

Observe that if we can find a logarithm, M, of $I + N$, as S commutes with any matrix and as $e^S = \alpha I$ and $e^M = I + N$, we have

$$e^{S+M} = e^S e^M = \alpha I(I + N) = J_r(\alpha),$$

which means that $S + M$ is a logarithm of $J_r(\alpha)$. Therefore, the problem reduces to finding the logarithm of a unipotent matrix, $I + N$. However, this problem always has a solution. To see this, remember that for $|u| < 1$ (where $u \in \mathbb{C}$), the power series

$$\log(1+u) = u - \frac{u^2}{2} + \frac{u^3}{3} + \cdots + (-1)^{n+1}\frac{u^n}{n} + \cdots$$

is normally convergent. It turns out that the above fact can be generalized to matrices in the following way:

Proposition 0.1 *For every $n \times n$ matrix, A, such that $\|A\| < 1$, the series*

$$\log(I+A) = A - \frac{A^2}{2} + \frac{A^3}{3} + \cdots + (-1)^{n+1}\frac{A^n}{n} + \cdots$$

is normally convergent (here, $\|A\|$ denotes the Frobenius norm). Furthermore, if $\|A\| < 1$, we have

$$e^{\log(I+A)} = I + A.$$

Proof. Since $\|A^n\| \leq \|A\|^n$ and since the power series for $\log(1+z)$ converges for $|z| < 1$, when $z \in \mathbb{C}$, we see that the power series for $\log(I+A)$ is normally convergent if $\|A\| < 1$.

To prove the second part of the lemma, first assume that A can be diagonalized. If $A = PDP^{-1}$ where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then we know that

$$A^n = PD^nP^{-1}.$$

We claim that for any complex matrix A (not necessarily diagonalizable), if $\|A\| < 1$, then $|\lambda_j| < 1$, for $j = 1, \ldots, n$. Let $(u_1, \ldots, u_n)$ be an orthonormal basis and express each vector Au_j in terms of the basis $(u_1, \ldots, u_n)$ as

$$Au_j = \sum_{k=1}^{n} b_{kj}u_k.$$

If we let B be the matrix $B = (b_{ij})$ and U be the orthogonal matrix whose jth column is u_j, then we have

$$AU = UB,$$

which yields

$$A = UBU^\top.$$

It follows that

$$A^\top A = (UBU^\top)^\top UBU^\top = UB^\top U^\top UBU^\top = UB^\top BU^\top,$$

and since

$$\mathrm{tr}(UB^\top BU^\top) = \mathrm{tr}(B^\top B),$$

we get

$$\|A\| = \sqrt{\mathrm{tr}(A^\top A)} = \sqrt{\mathrm{tr}(B^\top B)} = \|B\|\,.$$

On the other hand,

$$\langle u_i, Au_j\rangle = \langle u_i, \sum_{k=1}^n b_{kj}u_k\rangle = \sum_{k=1}^n \overline{b}_{kj}\langle u_i, u_k\rangle = \overline{b}_{ij},$$

and so

$$\|B\|^2 = \sum_{i,j=1}^n |b_{ij}|^2 = \sum_{i,j=1}^n |\langle u_i, Au_j\rangle|^2.$$

It follows that

$$\|A\|^2 = \sum_{i,j=1}^n |\langle u_i, Au_j\rangle|^2.$$

Now, if λ is an eigenvalue of A and if u is a unit eigenvector associated with λ, we can form an orthonormal basis $(u_1, \ldots, u_n)$ where $u_1 = u$, and then

$$\|A\|^2 = \sum_{i,j=1}^n |\langle u_i, Au_j\rangle|^2 = |\lambda|^2 + \sum_{i=1,j=2}^n |\langle u_i, Au_j\rangle|^2,$$

which proves that $|\lambda| \leq \|A\| < 1$, as claimed.

Actually, we can show that for any matrix norm $\|\ \|$ and for any complex $n \times n$ matrix A,

$$\rho(A) = \max_{1\leq i\leq n} |\lambda_i| \leq \|A\|\,,$$

where the λ_i are the eigenvalues of A. Recall that a matrix norm satisfies the property

$$\|AB\| \leq \|A\|\,\|B\|$$

for all A, B.

Let λ be some eigenvalue of A for which $|\lambda| = \rho(A)$ (i.e., $|\lambda|$ is maximum) and let u be a corresponding eigenvector. If U denotes the matrix whose n columns are all u, then $Au = \lambda u$ implies that

$$AU = \lambda U$$

and since

$$|\lambda|\,\|U\| = \|\lambda U\| = \|AU\| \leq \|A\|\,\|U\|$$

and $U \neq 0$, we get

$$\rho(A) = |\lambda| \leq \|A\|\,,$$

as claimed. In particular, since the Frobenius norm $\|\ \|_F$ is a matrix norm, we get another proof of the fact that $\rho(A) \leq \|A\|_F$.

Since $|\lambda_j| < 1$ for $j = 1, \ldots, n$, each power series for $\log(1+\lambda_j)$ converges, and so

$$\begin{aligned}\log(I+A) &= \sum_{n=1}^{\infty}(-1)^{n+1}\frac{A^n}{n!}\\ &= P\mathrm{diag}\left(\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\lambda_1^n}{n!}, \ldots, \sum_{n=1}^{\infty}(-1)^{n+1}\frac{\lambda_n^n}{n!}\right)P^{-1}\\ &= P\mathrm{diag}\left(\log(1+\lambda_1), \ldots, \log(1+\lambda_n)\right)P^{-1}.\end{aligned}$$

It follows that

$$e^{\log(I+A)} = P\mathrm{diag}\left(e^{\log(1+\lambda_1)}, \ldots, e^{\log(1+\lambda_n)}\right)P^{-1} = P\mathrm{diag}\left(1+\lambda_1, \ldots, 1+\lambda_n\right)P^{-1} = I + A,$$

as claimed.

If A can't be diagonalized, it can still be written as

$$A = UTU^{\top},$$

where T is an upper triangular matrix and U is unitary (Schur's lemma). Then, if some of the diagonal entries of T are not distinct, we can find n sequences $(\epsilon_m^1), \ldots, (\epsilon_m^n)$ each converging to 0, so that for every m, the numbers $T_{11} + \epsilon_m^1, \ldots, T_{nn} + \epsilon_m^n$ are all distinct, and if we let T_m be the matrix obtained from T by replacing the diagonal by $T_{11} + \epsilon_m^1, \ldots, T_{nn} + \epsilon_m^n$, then the sequence of matrices $A_m = UT_mU^{\top}$ converges to A and is A_m is diagonalizable, since its eigenvalues are all distinct. If $\|A\| < 1$, then for m large enough we have $\|A_m\| < 1$ and by the previous case, $e^{\log(I+A_m)} = I + A_m$, and by continuity of the exponential and of the log, we get $e^{I+A} = I + A$, as claimed. □

We will now prove that if N is any nilpotent matrix, then

$$e^{\log(I+N)} = I + N,$$

which finishes the proof that every unipotent matrix has a log, and thus, that every complex matrix has a log.

Note that $\|N\|$ is can be arbitary, so a different argument is needed. We argue as follows: For any nilpotent matrix N (with $N^r = 0$), the map

$$t \mapsto e^{\log(I+tN)} - (I + tN), \qquad t \in \mathbb{R}$$

is a polynomial, since $N^r = 0$. Furthermore, for t sufficiently small, $\|tN\| < 1$ and in view of Proposition 0.1, we have $e^{\log(I+tN)} = I + tN$, so the above polynomial vanishes in a neighborhood of 0, which implies that it is identically zero. Therefore, $e^{\log(I+N)} = I + N$, as required.

Problem B4 (60 pts). Recall from Homework 1, Problem B1, the Cayley parametrization of rotation matrices in $\mathbf{SO}(n)$ given by

$$C(B) = (I - B)(I + B)^{-1},$$

where B is any $n \times n$ skew-symmetric matrix. In that problem, it was shown that $C(B)$ is a rotation matrix that does not admit -1 as an eigenvalue and that every such rotation matrix is of the form $C(B)$.

(a) If you have not already done so, prove that the map $B \mapsto C(B)$ is injective.

(b) Prove that

$$dC(B)(A) = D_A((I-B)(I+B)^{-1}) = -[I + (I-B)(I+B)^{-1}]A(I+B)^{-1}.$$

Hint. First, show that $D_A(B^{-1}) = -B^{-1}AB^{-1}$ (where B is invertible) and that $D_A(f(B)g(B)) = (D_A f(B))g(B) + f(B)(D_A g(B))$, where f and g are differentiable matrix functions.

Deduce that $dC(B)$ is injective, for every skew-symmetric matrix, B. If we identify the space of $n \times n$ skew-symmetric matrices with $\mathbb{R}^{n(n-1)/2}$, show that the Cayley map, $C\colon \mathbb{R}^{n(n-1)/2} \to \mathbf{SO}(n)$, is a parametrization of $\mathbf{SO}(n)$.

(c) Now, consider $n = 3$, i.e., $\mathbf{SO}(3)$. Let E_1, E_2 and E_3 be the rotations about the x-axis, y-axis, and z-axis, respectively, by the angle π, i.e.,

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad E_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad E_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Prove that the four maps

$$\begin{aligned} B &\mapsto C(B) \\ B &\mapsto E_1C(B) \\ B &\mapsto E_2C(B) \\ B &\mapsto E_3C(B) \end{aligned}$$

where B is skew-symmetric, are parametrizations of $\mathbf{SO}(3)$ and that the union of the images of C, E_1C, E_2C and E_3C covers $\mathbf{SO}(3)$, so that $\mathbf{SO}(3)$ is a manifold.

(d) Let A be *any* matrix (not necessarily invertible). Prove that there is some diagonal matrix, E, with entries $+1$ or -1, so that $EA + I$ is invertible.

(e) Prove that every rotation matrix, $A \in \mathbf{SO}(n)$, is of the form

$$A = E(I-B)(I+B)^{-1},$$

for some skew-symmetric matrix, B, and some diagonal matrix, E, with entries $+1$ and -1, and where the number of -1 is even. Moreover, prove that every orthogonal matrix $A \in \mathbf{O}(n)$ is of the form

$$A = E(I-B)(I+B)^{-1},$$

for some skew-symmetric matrix, B, and some diagonal matrix, E, with entries $+1$ and -1. The above provide parametrizations for $\mathbf{SO}(n)$ (resp. $\mathbf{O}(n)$) that show that $\mathbf{SO}(n)$ and $\mathbf{O}(n)$

are manifolds. However, observe that the number of these charts grows exponentially with n.

Solutions.

B4(a). We proved in B1(d) that for any rotation matrix R such that R does not admit -1 as an eigenvalue, the skew-symmetric matrix B such that

$$R = C(B) = (I - B)(I + B)^{-1}$$

is uniquely determined by the formula

$$B = (I - R)(I + R)^{-1}.$$

Therefore, the Cayley map $B \mapsto C(B)$ is injective.

B4(b). Observe that the Cayley map $C\colon B \mapsto C(B)$ can be viewed as a map between the normed vector spaces $\mathbb{R}^{n(n-1)/2}$ and $\mathbb{R}^{n^2}$, since the vector space of skew-symmetric matrices is isomorphic to $\mathbb{R}^{n(n-1)/2}$.

Let f and g be the maps defined on $n \times n$ matrices by $f(B) = I - B$ and $g(B) = I + B$. We would like to compute the derivative $d(fg^{-1})(B)$ of $C = fg^{-1}$ on the vector space of skew-symmetric matrices. We will use the product rule and the chain rule.

First, we claim that

$$\begin{aligned} df(B) &= -\mathrm{id} \\ dg(B) &= \mathrm{id}, \end{aligned}$$

for all matrices B. Indeed

$$f(B + H) - f(B) + \mathrm{id}(H) = I - (B + H) - (I - B) + H = 0$$

and

$$g(B + H) - g(B) - \mathrm{id}(H) = I + (B + H) - (I + B) - H = 0,$$

which proves our claim.

Let h be the matrix inverse function, namely,

$$h(B) = B^{-1}.$$

The map h is defined on $\mathbf{GL}(n, \mathbb{R})$, an open subset of $\mathbb{R}^{n^2}$. We claim that $dh(B)(H) = -B^{-1}HB^{-1}$, for all $B \in \mathbf{GL}(n, \mathbb{R})$ and for all H. We have

$$\begin{aligned} h(B + H) - h(B) + B^{-1}HB^{-1} &= (B + H)^{-1} - B^{-1} + B^{-1}HB^{-1} \\ &= (B + H)^{-1}[I - (B + H)B^{-1} + (B + H)B^{-1}HB^{-1}] \\ &= (B + H)^{-1}[B - (B + H) + (B + H)B^{-1}H]B^{-1} \\ &= (B + H)^{-1}[-H + H + HB^{-1}H]B^{-1} \\ &= (B + H)^{-1}HB^{-1}HB^{-1}. \end{aligned}$$

Consequently, we get

$$\epsilon(H) = \frac{h(B+H) - h(B) + B^{-1}HB^{-1}}{\|H\|} = \frac{(B+H)^{-1}HB^{-1}HB^{-1}}{\|H\|},$$

and it is clear that $\lim_{H \mapsto 0} \epsilon(H) = 0$, which proves that

$$dh(B)(H) = -B^{-1}HB^{-1}$$

for all $B \in \mathbf{GL}(n, \mathbb{R})$ and for all H. Now, $g^{-1}(B) = (h \circ g)(B)$, so by the chain rule

$$dg^{-1}(B) = dh(g(B)) \circ dg(B)$$

and since $dg(B) = \mathrm{id}$ and $g(B) = I + B$, we get

$$dg^{-1}(B)(H) = -(I+B)^{-1}H(I+B)^{-1},$$

for all skew-symmetric matrices B and for all H. By the product rule,

$$df g^{-1}(B)(H) = df(B)(H)g^{-1}(B) + f(B)dg^{-1}(B)(H),$$

and since $df(B)(H) = -H$, $f(B) = I - B$, and $g(B) = I + B$, we get

$$\begin{aligned} df g^{-1}(B)(H) &= -H(I+B)^{-1} + (I-B)\big(-(I+B)^{-1}H(I+B)^{-1}\big) \\ &= -[I + (I-B)(I+B)^{-1}]H(I+B)^{-1}, \end{aligned}$$

for all skew-symmetric matrices B and for all H, as claimed. Actually, since $C = fg^{-1}$ is a map from the space of skew-symmetric matrices to $\mathbb{R}^{n^2}$, $dC(B)$ is a linear map from the space of skew-symmetric matrices to $\mathbb{R}^{n^2}$, but it happens to be also defined on $\mathbb{R}^{n^2}$.

Because B is a skew-symmetric matrix, we proved in Problem B1 that $(I-B)(I+B)^{-1}$ is a rotation matrix that does not admit the eigenvalue -1, which implies that $I+(I-B)(I+B)^{-1}$ is invertible. Since $(I+B)^{-1}$ is also invertible and since the linear map $dC(B)$ is given by

$$dC(B)(H) = -[I + (I-B)(I+B)^{-1}]H(I+B)^{-1},$$

it is clear that $dC(B)$ is injective.

The Cayley map $B \mapsto C(B)$ is continuous since its derivative exists for all skew-symmetric matrices B. We already know that it is injective and its derivative is injective for all skew-symmetric B. To be a parametrization, it remains to show that the inverse of C is continuous on the image of C, namely, the set of rotation matrices that do not admit -1 as an eigenvalue. The inverse map if given by the formula

$$B = (I-R)(I+R)^{-1}$$

which involves computing the inverse of a matrix. However, we know that the inverse of a matrix can be computed in terms of ratios of determinants (using the cofactors of the

matrix) and these functions are rational, and thus continuous. Therefore, the Cayley map is a parametrization of $\mathbf{SO}(n)$.

B4(c). Since the matrices E_1, E_2, E_3 are rotation matrices, they are invertible and so, the maps $C_1 \colon C \mapsto E_1C(B)$, $C_2 \colon C \mapsto E_2C(B)$, and $C_3 \colon C \mapsto E_3C(B)$ are injective. Since E_1, E_2, E_3 are constant matrices, by the product rule, we get

$$\begin{aligned}
dC_1(B) &= dE_1(B)C(B) + E_1(B)dC(B) = 0 + E_1dC(B) = E_1dC(B)\\
dC_2(B) &= dE_2(B)C(B) + E_2(B)dC(B) = 0 + E_2dC(B) = E_2dC(B)\\
dC_3(B) &= dE_3(B)C(B) + E_3(B)dC(B) = 0 + E_3dC(B) = E_3dC(B),
\end{aligned}$$

Since E_1, E_2, E_3 are invertible and since $dC(B)$ is injective, the linear maps $dC_1(B) = E_1dC(B)$, $dC_2(B) = E_2dC(B)$, and $dC_3(B) = E_3dC(B)$ are injective. Since $dC_1(B)$, $dC_2(B)$, and $dC_3(B)$ exist for all skew-symmetric B, the maps C_1, C_2, C_3 are continuous.

Observe that $E_i^2 = I$, for $i = 1, 2, 3$. If

$$R = E_i(I - B)(I + B)^{-1},$$

then

$$E_iR = (I - B)(I + B)^{-1},$$

from which we get

$$E_iR(I + B) = I - B,$$

and then

$$(I + E_iR)B = I - E_iR.$$

Because B is skew-symmetric, we know that $(I - B)(I + B)^{-1} = E_iR$ does not admit -1 as an eigenvalue, and thus $I + E_iR$ is invertible and so B is uniquely determined by the formula

$$B = (I - E_iR)(I + E_iR)^{-1},$$

which also shows that the inverse of the map C_i is continuous. Therefore, the maps C, C_1, C_2, C_3 are parametrizations of $\mathbf{SO}(3)$. It remains to show that the images of these maps cover $\mathbf{SO}(3)$. I can't find a direct proof for the 3D case but this follows from B4(e).

B4(d). This is a rather tricky exercise due to Richard Bellman. Actually, we can prove a little more.

Observe that if E is a diagonal matrix whose entries are ± 1, then $E^2 = I$. Consequently, by multiplying by E, we get the following fact:

$$I + EA \text{ is invertible iff } E + A \text{ is.}$$

Thus, we are naturally led to the following problem: If A is any $n \times n$ matrix, is there a way to perturb the diagonal entries of A, i.e., to add some diagonal matrix, $C = \mathrm{diag}(c_1, \ldots, c_n)$, to A so that $C + A$ becomes invertible?

Indeed this can be done, and we will show in the next section that what matters is not the magnitude of the perturbation but the signs of the entries being added.

Proposition 0.2 *For every $n \times n$ matrix (invertible or not), A, and every any diagonal matrix, $C = \mathrm{diag}(c_1, \ldots, c_n)$, with $c_i \neq 0$ for $i = 1, \ldots, n$, there an assignment of signs, $\epsilon_i = \pm 1$, so that if $E = \mathrm{diag}(\epsilon_1 c_1, \ldots, \epsilon_n c_n)$, then $E + A$ is invertible.*

Proof. Let us evaluate the determinant of $C+A$. We see that $\Delta = \det(C+A)$ is a polynomial of degree n in the variables $c_1, \ldots, c_n$ and that all the monomials of Δ consist of products of distinct variables (i.e., every variable occurring in a monomial has degree 1). In particular, Δ contains the monomial $c_1 \cdots c_n$. In order to prove Proposition 0.2, it will suffice to prove

Proposition 0.3 *Given any polyomial, $P(x_1, \ldots, x_n)$, of degree n (in the indeterminates $x_1, \ldots, x_n$ and over any integral domain of characteristic unequal to 2), if every monomial in P is a product of distinct variables, for every n-tuple $(c_1, \ldots, c_n)$ such that $c_i \neq 0$ for $i = 1, \ldots, n$, then there is an assignment of signs, $\epsilon_i = \pm 1$, so that*

$$P(\epsilon_1 c_1, \ldots, \epsilon_n c_n) \neq 0.$$

Clearly, any assignment of signs given by Proposition 0.3 will make $\det(E + A) \neq 0$, proving Proposition 0.2. □

It remains to prove Proposition 0.3.

Proof of Proposition 0.3. We proceed by induction on n (starting with $n = 1$). For $n = 1$, the polynomial $P(x_1)$ is of the form $P(x_1) = a + bx_1$, with $b \neq 0$ since $\deg(P) = 1$. Obviously, for any $c \neq 0$, either $a + bc \neq 0$ or $a - bc \neq 0$ (otherwise, $2bc = 0$, contradicting $b \neq 0$, $c \neq 0$ and the ring being an integral domain of characteristic $\neq 2$).

Assume the induction hypothesis holds for any $n \geq 1$ and let $P(x_1, \ldots, x_{n+1})$ be a polynomial of degree $n + 1$ satisfying the conditions of Proposition 0.3. Then, P must be of the form

$$P(x_1, \ldots, x_n, x_{n+1}) = Q(x_1, \ldots, x_n) + S(x_1, \ldots, x_n)x_{n+1},$$

where both $Q(x_1, \ldots, x_n)$ and $S(x_1, \ldots, x_n)$ are polynomials in $x_1, \ldots, x_n$ and $S(x_1, \ldots, x_n)$ is of degree n and all monomials in S are products of distinct variables. By the induction hypothesis, we can find $(\epsilon_1, \ldots, \epsilon_n)$, with $\epsilon_i = \pm 1$, so that

$$S(\epsilon_1 c_1, \ldots, \epsilon_n c_n) \neq 0.$$

But now, we are back to the case $n = 1$ with the polynomial

$$Q(\epsilon_1 c_1, \ldots, \epsilon_n c_n) + S(\epsilon_1 c_1, \ldots, \epsilon_n c_n)x_{n+1},$$

and we can find $\epsilon_{n+1} = \pm 1$ so that

$$P(\epsilon_1 c_1, \ldots, \epsilon_n c_n, \epsilon_{n+1} c_{n+1}) = Q(\epsilon_1 c_1, \ldots, \epsilon_n c_n) + S(\epsilon_1 c_1, \ldots, \epsilon_n c_n)\epsilon_{n+1} c_{n+1} \neq 0,$$

establishing the induction hypothesis. □

Note that in Proposition 0.2, the c_i can be made arbitrarily small or large, as long as they are not zero. Thus, we see as a corollary that any matrix can be made invertible by a very small perturbation of its diagonal elements. What matters is the signs that are assigned to the perturbation.

B4(e). We prove the following proposition:

Proposition 0.4 *For any orthogonal matrix, $R \in \mathbf{O}(n)$, there is some diagonal matrix, E, whose entries are $+1$ or -1, and some skew-symmetric matrix, S, so that*

$$R = E(I - S)(I + S)^{-1}.$$

Proof of Proposition 0.4. Let $R \in \mathbf{O}(n)$ be any orthogonal matrix. By Proposition 0.2, we can find a diagonal matrix, E (with diagonal entries ± 1), so that $I + ER$ is invertible. But then, as E is orthogonal, ER is an orthogonal matrix that does not admit the eigenvalue -1 and so, by the Cayley representation theorem, there is a skew-symmetric matrix, S, so that

$$ER = (I - S)(I + S)^{-1}.$$

However, notice that $E^2 = I$, so we get

$$R = E(I - S)(I + S)^{-1},$$

as claimed. □

We have $\det(R) = \det(E)\det((I - S)(I + S)^{-1}) = \det(E)$, because $(I - S)(I + S)^{-1}$ is a rotation matrix. If R is a rotation matrix, $\det(R) = +1$, and so $\det(E) = +1$. Therefore, when R is a rotation matrix, E must have an even number of -1.

The fact that the maps $B \mapsto EC(B)$ parametrize the manifold $\mathbf{SO}(n)$ when E has an even number of -1 and the manifold $\mathbf{O}(n)$ in general is an immediate generalization of part B4(c).