Exercise Guide for *Linear Algebra* by Peter Lax

Muthu Chidambaram

Last Updated: June 24, 2019

Contents

About

> *"A modern mathematical proof is not very different from a modern machine, or a modern test setup: the simple fundamental principles are hidden and almost invisible under a mass of technical details."*
> - Hermann Weyl

These notes contain short summaries of my proof ideas for exercises from the book *Linear Algebra* by Peter Lax. Note that this is not the same book as Lax's more recent *Linear Algebra And Its Applications*. Unfortunately, I only had access to a hard copy of the former.

1 Fundamentals

1.1 Exercise 1

$x + z = x = x + z' \implies z = z'$.

1.2 Exercise 2

$0x + x = (0 + 1)x = x$.

1.3 Exercise 3

Coefficients can be represented as row vectors.

1.4 Exercise 4

Function can be represented as row vector by letting $a_i = f(s_i)$ for each $s_i \in S$.

1.5 Exercise 5

Follows from exercises 3 and 4.

1.6 Exercise 6

$y_1 + z_1 + y_2 + z_2 = (y_1 + y_2) + (z_1 + z_2)$ and $k(y_1 + z_1) = ky_1 + kz_1$.

1.7 Exercise 7

$a \in Y \cap Z \implies ka \in Y, ka \in Z \implies ka \in Y \cap Z$.

1.8 Exercise 8

$k0 = 0, 0 + 0 = 0$.

1.9 Exercise 9

If S contains x_i then it must contain kx_i.

1.10 Exercise 10

If $x_i = 0$, k_i can be anything.

1.11 Exercise 11

$x = \sum_{i=1}^{m} \sum_{j=1}^{\dim Y_i} y_j^{(i)}$.

1.12 Exercise 12

Complete basis for W to U and V. Use W basis vectors and additional U and V basis vectors to get $\dim X = \dim U - \dim W + \dim V - \dim W + \dim W$.

1.13 Exercise 13

Send i^{th} basis vector to e_i, where e_i is vector of all zeroes except a one in the i^{th} place. Can permute mapping to get different isomorphisms.

1.14 Exercise 14

$x_1 - x_2 + x_2 - x_3 = x_1 - x_3$.

1.15 Exercise 15

$x' = x + z_x, y' = y + z_y \implies x' + y' = x + y + (z_x + z_y)$.

1.16 Exercise 16

$x \in X_1 \bigoplus X_2 \implies x = (x_1, x_2) = (x_1, 0) + (0, x_2)$.

1.17 Exercise 17

Construct a basis for X from Y: $y_1, ..., y_j, x_{j+1}, ..., x_n$.
Then $X/Y = \text{span}\{x_{j+1}, ..., x_n\}$.

2 Duality

Theorem 1

$x = \sum_{i=1}^{n} a_i x_i \implies k_i(x) = a_i.$

2.1 Exercise 1

$l_1, l_2 \in Y^{\perp} \implies l_1(y) + l_2(y) = 0 = (l_1 + l_2)(y).$

2.2 Exercise 2

$\forall \xi \in Y^{\perp\perp} \implies \forall l \in Y^{\perp},\, \xi(l) = 0 = l(y)\, \forall y \in Y.$

3 Linear Mappings

3.1 Exercise 1

(a) $x \in X \implies x = \sum_{i=1}^{n} k_i x_i \implies T(x) = \sum_{i=1}^{n} k_i T(x_i) \in U$.

(b) $T(x), T(y) \in U \implies T(x+y) \in U \implies x + y \in X$.

Theorem 1

$x \in X, y \in N_T \implies T(x+y) = T(x) + T(y) = T(x)$.

3.2 Exercise 2

(a) Differentiation constant and sum rules imply linearity, and multiplication by s is distributive. Take $p(s) = 1$ to see that $ST \neq TS$.

(b) Rotation by 90 degrees amounts to swapping and negating coordinates, which is linear. Take $p = (1, 1, 0)$ to see that $ST \neq TS$.

3.3 Exercise 3

(i) $T^{-1}(T(a+b)) = T^{-1}(T(a)+T(b)) = a+b = T^{-1}(T(a))+T^{-1}(T(b))$.

(ii) Composition of isomorphisms is an isomorphism, hence ST is invertible.

3.4 Exercise 4

(i) Let $T : X \to U$, $S : U \to V$ and $l_v \in V'$. Then $(ST)'(l_v) = l_v(ST) = (l_v S)T = (S' l_v)T = T'S' l_v$, since $S' l_v \in U'$.

(ii) Follows from linearity of transpose (definition).

(iii) Let $T : X \to U$ be an isomorphism. Then $l_x = l_u T \implies l_x T^{-1} = l_u$ for $l_u \in U'$, $l_x \in X'$.

3.5 Exercise 5

$T''(l_{x'}) = l_{x'} T'$ where $l_{x'} \in X''$ and $l_{x'} T' \in U''$. Since we can identify elements in X'' and U'' with elements in X and U respectively, we have that T'' assigns elements of U to X.

Theorem 2'

Since $T' : U' \to X'$ we have $l_u \in N_{T'} \implies T'(l_u) = l_u T = 0$. $N_{T'}^{\perp}$ consists of elements $l_{u'} | l_{u'}(l_u) = 0$. From $l_u T x = 0$ we have that each $l_{u'}$ is identified with a $u \in R_T$.

3.6 Exercise 6

The first two elements of x are already 0 after applying P, so $P^2 = P$. Linearity follows from linearity of vector addition.

3.7 Exercise 7

P is linear since function addition is linear. $P^2 f = \frac{f(x)+f(-x)}{4} + \frac{f(x)+f(-x)}{4} = Pf$.

4 Matrices

4.1 Exercise 1

$(P+T)_{ij} = ((P+T)e_j)_i = (Pe_j + Te_j)_i = P_{ij} + T_{ij}$.

4.2 Exercise 2

Represent A as a column of row vectors A_i and B as a row of column vectors B_i. Denote blocks by parenthesized subscripts. Then the first block of AB looks like:

$$
\begin{aligned}
(AB)_{(11)} &= \begin{pmatrix} A_1 B_1 & & \\ & \ddots & \\ & & A_k B_k \end{pmatrix} \\
&= \begin{pmatrix} A_{1,:(k+1)} B_{1,:(k+1)} & & \\ & \ddots & \\ & & A_{k,:(k+1)} B_{k,:(k+1)} \end{pmatrix} \\
&+ \begin{pmatrix} A_{1,(k+1):} B_{1,(k+1):} & & \\ & \ddots & \\ & & A_{k,(k+1):} B_{k,(k+1):} \end{pmatrix} \\
&= A_{(11)} B_{(11)} + A_{(12)} B_{(21)}
\end{aligned}
$$

Where:

$$
\begin{aligned}
A_{i,:(k+1)} B_{i,:(k+1)} &= \sum_{j=1}^{k} A_{i,j} B_{i,j} \\
A_{i,(k+1):} B_{i,(k+1):} &= \sum_{j=k+1}^{n} A_{i,j} B_{i,j}
\end{aligned}
$$

The rest follow similarly.

5 Determinant and Trace

5.1 Exercise 1

(a) The discriminant already has ordered versions of all the (i,j) difference terms. Applying a permutation only changes the signs of some of the difference terms, hence $\sigma(p) = 1, -1$.

(b) $\sigma(p_1 \circ p_2) = \text{sign}(P(p_1 \circ p_2(x_1, ..., x_n))) = \sigma(p_1)\text{sign}(P(p_2(x_1, ..., x_n)))$.

5.2 Exercise 2

(c) A transposition swaps two indices, and hence flips the sign of their associated difference term in the discriminant.

(d) If $p(i) = j$, then we can start with the permutation $(i\ j)$. Next, if $p(j) = k$, we can compose with $(i\ k)$ to get $(i\ k) \circ (i\ j)$. We can do this until we have completely reconstructed the permutation using transpositions.

5.3 Exercise 3

By starting with a different i in Exercise 2 (d), we can obtain a different decomposition of transpositions. However, the parity of the decomposition must be the same, as otherwise $\sigma(p)$ will take on two different values for the same p.

5.4 Exercise 4

(Property II): Each term in $D(a_1, ..., a_n)$ contains exactly one element from each of the a_i. Thus, scaling any of the a_i by k scales the entire determinant by k. Similar logic for vector addition.

(Property III): The only non-zero term in $D(e_1, ..., e_n)$ is associated with the identity permutation, hence $D(e_1, ..., e_n) = 1$.

(Property IV): Swapping two arguments is the same as applying a transposition to each of the terms in $D(a_1, ..., a_n)$, which flips the sign of D.

5.5 Exercise 5

Suppose $a_1 = a_2$. Then:

$$\begin{aligned} D(a_1, a_2, ..., a_n) &= -D(a_2, a_1, ..., a_n) \\ D(a_1, a_2, ..., a_n) + D(a_1, a_2, ..., a_n) &= 0 \end{aligned}$$

5.6 Exercise 6

We can swap rows and columns until A is in the same form as in Lemma 2. Since each row and column swap is equivalent to applying a transposition, we get that $\det A = (-1)^{i+j} \det A_{ij}$.

5.7 Exercise 7

Each term in the sum $D(a_1, ..., a_n) = \sum \sigma(p) a_{p_1 1} ... a_{p_n n}$ consists of exactly one element from each column and each row; swapping rows and columns does not change the terms in the sum. However, the permutation associated with each term is changed. The permutation p that sends $1 \to p_1$ becomes p' sending $p_1 \to 1$. This p' is exactly p^{-1}. Since $\sigma(1) = \sigma(p^{-1} \circ p) = \sigma(p^{-1})\sigma(p)$, $\sigma(p^{-1}) = \sigma(p)$ we are done.

5.8 Exercise 8

P is the linear transformation such that $P(e_j) = e_i$; in other words, P rearranges the representation of x by applying p to the components of x. We also have that $PQx = Pq(x) = p \circ q(x)$, since Qx permutes the components of x to produce $q(x)$, and $Pq(x)$ permutes the components of $q(x)$ to produce $p \circ q(x)$.

5.9 Exercise 9

$$\begin{aligned}
\operatorname{Tr} AB &= \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} \\
\operatorname{Tr} BA &= \sum_{j=1}^{n} (BA)_{jj} = \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} a_{ij}
\end{aligned}$$

5.10 Exercise 10

$$\begin{aligned}
\operatorname{Tr} AA^\top &= \sum (AA^\top)_{ii} = \sum \sum a_{ij} a^\top_{ji} \\
&= \sum \sum a_{ij}^2
\end{aligned}$$

6 Spectral Theory

6.1 Exercise 1

(a) We can re-express h as a linear combination of its eigenvectors, from which we can see that A^n causes all of these components to go to 0.

(b) Same as in part (a), except all of the components now go to ∞.

6.2 Exercise 2

$A(A^N f) = A(a^N f + Na^{N-1}h) = a^{N+1}f + a^N h + Na^N h$.

6.3 Exercise 3

Suppose $q(A) = \sum_{i=0}^{N} q_i A^i$. Then $q_i A^i f = q_i a^i f + q_i i a^{i-1} h$ by Exercise 2. From the linearity of the derivative it then follows that $q(A)f = q(a)f + q'(a)h$.

6.4 Exercise 4

Applying Lemma 9 to $p_1 ... p_k$ and p_{k+1} gives the desired result.

6.5 Exercise 5

$(A - aI)^d x = 0 \implies A(A - aI)^d x = 0 \implies (A - aI)^d (Ax) = 0 \implies Ax \in N_d$.

6.6 Exercise 6

Since m_A divides the characteristic polynomial of A (by definition of d_i), $m_A(A) = 0$ by Cayley-Hamilton. Suppose there is some polynomial $q(A) = 0$ with $\deg(q) < \deg(m_A)$. The roots of q must contain all of the eigenvalues of A, since for any eigenvector h we have that $q(A)h = q(a_h)h$ (where a_h denotes the eigenvalue associated with h). Thus, the roots of q can only differ in multiplicity from m_A. However, if any root of q has multiplicity $d_i' < d_i$, then we can choose an element $x \in N_{d_i}$ such that $q(A)x \neq 0$, which is a contradiction (since $N_{d_i'} \subset N_{d_i}$).

6.7 Exercise 7

The columns of A are Ax_i.

6.8 Exercise 8

By induction.

6.9 Exercise 9

The minimal polynomial of A divides the minimal polynomial of $A^\top$, and vice versa, so they must be the same. Thus, the indices of each eigenvalue of A and $A^\top$ must be the same, which means we can apply Theorem 12 to see that they are similar.

6.10 Exercise 10

$(\xi^{(i)}, x) = \sum k_j(\xi^{(i)}, x^{(j)}) = k_i(\xi^{(i)}, x^{(i)})$ since $(\xi^{(i)}, x^{(j)}) = 0$ by Theorem 17.

7 Euclidean Structure

7.1 Exercise 1

Applying Cauchy-Schwarz gives that $(x, y) \leq \|x\|\|y\| = \|x\|$, which yields the desired result.

7.2 Exercise 2

Both of these just follow from the fact that a linear space with Euclidean structure is isomorphic to R^k, which has the desired properties. The following feel like cop-out solutions, but I feel they're fair given this isn't supposed to be a real analysis text.

(i) Let $x_k = \sum_i a_i^{(k)} x^{(i)}$ and $x_j = \sum_i a_i^{(j)} x^{(i)}$. Then we have that $\|x_k - x_j\| \to 0 \implies \left|a_i^{(k)} - a_i^{(j)}\right| \to 0$. Since $\mathbb{R}$ is complete, $a_i^{(n)} \to a_i$ for some $a_i \in \mathbb{R}$. Thus $x_k \to x = \sum a_i x^{(i)}$.

(ii) Same logic as (i): the individual $a_i^{(n)}$ have convergent subsequences.

7.3 Exercise 3

We also need to assume X is finite dimensional (I think).

(i) Since X is finite dimensional, we have that $x = \sum a_i e^{(i)}$ for some basis $e^{(i)}$. Let v be the vector whose components are $v_i = \left\|Ae^{(i)}\right\|$. Then

$$\begin{aligned}
\|Ax\| &= \left\|\sum a_i A e^{(i)}\right\| \\
&= \|(x, v)\| \\
&\leq \|x\|\|v\|
\end{aligned}$$

Since $\|v\|$ is a constant, $\|Ax\|$ is bounded on the unit sphere.

(ii) We have that $\|A\| = \max_x \frac{\|Ax\|}{\|x\|} = \max_x \|Ax\|$. The result then follows from $(Ax, y) \leq \|Ax\|$.

(iii) Let v be as in (i). Then $\|Ax_i - Ax_j\| \leq \|A\|\|x_i - x_j\|$. Since $\|A\| \leq \|v\|$ from (i), we are done (we can make $\|x_i - x_j\|$ as small as we'd like).

7.4 Exercise 4

Follows immediately from $(Ax, y) = (x, A^*y)$ and Exercise 3 (ii).

7.5 Exercise 5

Let $x = y_1 + y_1^{\perp}$ and $z = y_2 + y_2^{\perp}$. Then

$$\begin{aligned}(P_Y x, z) &= (y_1, y_2 + y_2^{\perp}) \\ &= (y_1, y_2) + 0 \\ &= (y_1, y_2) + (y_1^{\perp}, y_2) \\ &= (x, P_Y z)\end{aligned}$$

7.6 Exercise 6

Reflection across $x_3 = 0$ sends $(x_1, x_2, x_3) \to (x_1, x_2, -x_3)$. Hence it can be represented as

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

Which has determinant -1 (product of diagonal terms).

7.7 Exercise 7

(a) If A is orthogonal, then $A^*A = I$. Since $A^* = A^\top$ in this case, we immediately get that the column vectors of A are pairwise orthogonal unit vectors. Similarly, a matrix A with pairwise orthogonal unit vectors satisfies $A^*A = I$, implying that it is orthogonal.

(b) $A^*A = I \implies A^* = A^{-1} \implies AA^* = I$ so A orthogonal implies A^* orthogonal. The result then follows from plugging A^* into (a).

7.8 Exercise 8

The proof is almost identical to the non-complex case.

$$\begin{aligned}(x + ty, x + ty) &= (x, x) + (ty, x) + (x, ty) + \|t\|^2 (y, y) \\ &= (x, x) + t(y, x) + \bar{t}(x, y) + \|t\|^2 (y, y)\end{aligned}$$

Plugging in $t = \frac{(x,y)}{(y,y)}$ and using the fact that $(x + ty, x + ty) \geq 0$ for all complex t gives the desired inequality.

7.9 Exercise 9

7.9.1 Theorem 4

The proof of Theorem 4 is the same, except we use $y = \sum \bar{b_k} x^{(k)}$ instead.

7.9.2 Theorem 5

No changes need to be made to the proof of Theorem 5.

7.9.3 Theorem 6

Again, no changes need to be made.

7.9.4 Theorem 7

No changes need to be made here either, since $(y, y^\perp) = 0 \implies \overline{(y, y^\perp)} = 0$.

7.10 Exercise 10

7.10.1 Theorem 8

Parts (i)-(iii) remain the same. For part (iv), we see that

$$(Ax, y) = (x, A^*y) = \overline{(A^*y, x)} = \overline{(y, A^{**}x)} = (A^{**}x, y)$$

7.10.2 Theorem 9

(i) $\|kA\| = \max_x \frac{\sqrt{(kAx, kAx)}}{\|x\|} = \sqrt{k\bar{k}}\|A\| = \|k\|\|A\|$.

(ii) Applying triangle inequality to the definition of norm gives the result.

(iii) Comes immediately from $\|A(Bx)\| \leq \|A\|\|Bx\|$.

7.11 Exercise 11

Follows from $\|M(x) - M(y)\| = (\bar{x} - \bar{y})(x - y) = \|x - y\|$.

7.12 Exercise 12

Same idea as proof of Theorem 10.

7.13 Exercise 13

As in Exercise 7, $M^*M = I \implies M^* = M^{-1} \implies MM^* = I$. Similarly, $M^* = M^{-1} \implies (M^*)^{-1}M^{-1} = I$.

7.14 Exercise 14

Associativity follows from associativity of composition. From Exercise 13, if M is unitary, so is M^{-1}. Finally, I is also unitary, so the unitary maps are a group with unit I.

7.15 Exercise 15

Again, same idea as in Theorem 10: $\det M^* \det M = 1$.

7.16 Exercise 16

$$(Mf, Mg) = \int_{-1}^{1} m^2(s) f(s) \bar{g}(s) = \int_{-1}^{1} f(s) \bar{g}(s)$$

8 Spectral Theory of Selfadjoint Mappings of a Euclidean Space Into Itself

8.1 Exercise 1

This is only true when (y, My) is real, since

$$\begin{aligned}(y, \frac{M+M^*}{2}y) &= \frac{1}{2}(y, My) + \frac{1}{2}\overline{(y, My)} \\ &= \mathrm{Re}(y, My)\end{aligned}$$

8.2 Exercise 2

Algorithm is described by Lax; I'll pass on coding.

8.3 Exercise 3

Let S be the max subspace such that $q \geq 0$ and S_+ be the max subspace such that $q > 0$. Then the nullspace of the mapping $P : S \to S_+$ defined as setting the first p_+ elements of a vector to 0 consists of those vectors for which $q = 0$. Since the dimension of this nullspace is p_0, we have the desired result. Showing the same for S_- is analogous.

8.4 Exercise 4

This exercise is almost identical to the second proof of Theorem 4 shown in the book.

(a) Let $q(x) = (x, Hx)$ and $p(x) = (x, Mx)$. Then $R_{H,M}(x) = \frac{q(x)}{p(x)}$ is real since H and M are both self-adjoint. Furthermore, $R_{H,M}(kx) = R_{H,M}(x)$, so it suffices to optimize $R_{H,M}$ over the unit sphere. Since the unit sphere is compact, and $R_{H,M}$ is the quotient of two continuous functions (with $p(x) > 0$ due to positivity), $R_{H,M}$ has a minimum on the unit sphere (which is non-zero).

(b) Let the minimum in question be f. Then we can do exactly as Lax did and consider $R_{H,M}(f + tg)$ as a function of a single real variable t. The derivative of this function at $t = 0$ is 0, from which we get

$$\begin{aligned}\frac{p(f)\dot{q}(0) - q(f)\dot{p}(0)}{p^2(f)} &= 0 \\ \dot{q}(0) - b\dot{p}(0) = 0 \quad b &= \frac{q(f)}{p(f)} \\ (g, Hf) = b(g, Mf) &\implies Hf = bMf\end{aligned}$$

Where $\dot{q}$ and $\dot{p}$ refer to the derivatives of q, p with respect to the variable t.

(c) We are done if we can show that both H and M map the space Y consisting of $(y, Mf) = 0$ back to itself. To do so, we appeal to the representation of

M as a sum of projection operators, $M = \sum a_i P_i$ (this is possible since M is self-adjoint). Using this representation, we have that

$$(My, Mf) = (y, M^2 f) = (y, \sum a_i^2 P_i f) = \sum a_i (y, a_i P_i f) = 0$$

The analogous result follows for H since $HMf = bM^2 f$.

(d) This follows immediately from solving the constrained optimization problem in (c) in the same manner as (b).

8.5 Exercise 5

Repeatedly applying the procedure outlined in Exercise 4 allows us to construct the desired f_i. These f_i all satisfy $(Mf_i, Mf_j) = 0$ for $i \neq j$, so Mf_i forms a basis for X. As such, f_i must also form a basis for X, as otherwise the Mf_i would have a nontrivial linear relation sending them to 0.

8.6 Exercise 6

The equivalent minmax is

$$b_j = \min_{\dim S = j} \max_{x \in S, x \neq 0} \frac{(x, Hx)}{(x, Mx)}$$

The logic to show this is the same as in Theorem 10 in the book. For showing $\geq b_j$ we use the linear conditions $(x, f_i) = 0$ to get an x with no contribution from $f_1, ..., f_{j-1}$. For $\leq b_j$ we use the space spanned by $f_1, ..., f_j$.

The only difference is that instead of using Equation (36) from the chapter, we instead use

$$\begin{aligned} x &= z_1 f_1 + ... + z_n f_n \\ \frac{(x, Hx)}{(x, Mx)} &= \frac{\sum b_i z_i^2 (f_i, Mf_i)}{\sum z_i^2 (f_i, Mf_i)} \end{aligned}$$

8.7 Exercise 7

We have that $Hf_i = bMf_i \implies M^{-1}Hf_i = b_i f$ and all of the b_i are real per Exercises 4 and 5. If H is also positive, then $R(x)$ as defined in Exercise 4 is always positive, so the b_i will be positive as well.

8.8 Exercise 8

Since N is normal, it has an orthonormal basis consisting of eigenvectors f_i. Thus, we have

$$\begin{aligned} x &= z_1 f_1 + ... + z_n f_n \\ \frac{(Ax, Ax)}{(x, x)} &= \frac{\sum n_i^2 z_i^2}{\sum z_i^2} \\ &\leq \max n_j^2 \end{aligned}$$

Equality holds when x is selected to be the eigenvector corresponding to the largest (in absolute value) eigenvalue. Thus, $\|N\| = \max |n_j|$.

8.9 Exercise 9

(a) S is clearly an isometry, since all it does is rearrange terms in the expression for the distance between two vectors.

(b) Since $S^n = I$, the eigenvalues of S must be n^{th} roots of unity. Each n^{th} root a_i can be used to construct an eigenvector by letting $x_1 = \bar{a}_i, x_2 = \bar{a_i}^2, ..., x_n = 1$.

(c) Let x and y be eigenvectors corresponding to the eigenvalues a_i and a_j and let $(x, y) = \alpha$. Then $a_i \bar{a}_j (x, y) = \alpha$ since the $a_i \bar{a}_j$ powers are cyclic. Since $1 - a_i \bar{a}_j \neq 0$ for all $i \neq j$, $(x, y) = 0$ when $i \neq j$ and the eigenvectors are therefore orthogonal.

9 Calculus of Vector and Matrix Valued Functions

9.1 Exercise 1

That $\dot{x}(t) = 0 \implies x(t) = c$ follows immediately from the mean value inequality for vector-valued functions, $\|x(b) - x(a)\| \leq (b - a)\|x'(t)\|$ for some $t \in (a, b)$. My guess is that Lax was hinting at applying the mean value theorem for real-valued functions to $(x(b) - x(a), y)$.

9.2 Exercise 2

We have that

$$\frac{\mathrm{d}}{\mathrm{d}x}A^{-1}A = 0 = \left(\frac{\mathrm{d}}{\mathrm{d}x}A^{-1}\right)A + A^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}A\right) \implies \frac{\mathrm{d}}{\mathrm{d}x}A^{-1} = -A^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}A\right)A^{-1}$$

9.3 Exercise 3

The matrix $A + B$ satisfies $(A + B)^2 = I$, so we have that

$$\begin{aligned} e^{A+B} &= \sum_{k=0}^{\infty} \frac{(A+B)^k}{k!} \\ &= \left(\sum_{k=0}^{\infty} \frac{1}{(2k)!}\right)I + \left(\sum_{k=0}^{\infty} \frac{1}{(2k+1)!}\right)(A + B) \\ &= \cosh(1)I + \sinh(1)(A + B) \end{aligned}$$