The exercises of the textbook are not yet too interesting at this early stage, so I’m adding some of my own. I’ll have more to say about them in class.

1. Recall that the Lie bracket (or commutator) of smooth vector fields \( X, Y \) defined on some open set \( U \subset \mathbb{R}^n \) is the vector field \([X, Y]\) on \( U \) such that \([X, Y]f = XYf - YXf\) for every smooth function \( f : U \to \mathbb{R} \).

   (a) Show that the Lie bracket satisfies the Jacobi identity:
   \[ [X, [Y, Z]] = [[X, Y], Z] + [Y, [X, Z]] \]
   for all vector fields \( X, Y, Z \). (If you are wondering why I wrote the identity in this particular form, note the following: defining \( \mathcal{L}_X Y : = [X, Y] \), then the Jacobi identity is equivalent to Leibniz’s rule for \( \mathcal{L}_X \): \( \mathcal{L}_X [Y, Z] = [\mathcal{L}_X Y, Z] + [Y, \mathcal{L}_X Z] \). This operation on vector fields is called the Lie derivative along \( X \). We will have more to say about it later; keep it in the back of your mind because it is an important and useful gadget.)

   (b) Let \( A = (a_{ij}) \) be any \( n \times n \) matrix. Define the smooth vector field \( X^A \) on \( \mathbb{R}^n \) by
   \[ X^A = \sum_{i,j} x_i a_{ij} \frac{\partial}{\partial x_j} \]
   We call \( X^A \) a linear vector field. Show that linear vector fields are closed under taking Lie brackets. More precisely, show that
   \[ [X^A, X^B] = X^{[A, B]} \]
   where \( [A, B] = AB - BA \).

   (c) Consider the vector fields \( X_1, X_2 \) and \( X_3 \) on \( \mathbb{R}^3 \) defined by
   \[ X_1(x) = x_3 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_3}, \quad X_2(x) = x_3 \frac{\partial}{\partial x_3} - x_3 \frac{\partial}{\partial x_1}, \quad X_3(x) = x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2} \]
   Show that
   \[ [X_1, X_2] = X_3, \quad [X_2, X_3] = X_1, \quad [X_3, X_1] = X_2. \]

2. Spectral theorem for symmetric matrices. Let \( f : U \to \mathbb{R} \) be a differentiable function defined on an open subset of \( \mathbb{R}^n \). We assume that \( \mathbb{R}^n \) is equipped with the standard inner product (i.e., the dot product), which we denote by \( \langle \cdot, \cdot \rangle \). We define the gradient of \( f \) at \( x \) as the vector \( \text{grad}_x f \in T_x \mathbb{R}^n \) such that
   \[ \langle \text{grad}_x f, v \rangle = vf \]
for all \( v \in T_x \mathbb{R}^n \). (Convince yourself that this set of equations, one for each \( v \), indeed defines uniquely a vector at \( x \).) Recall the various notations we have been using for the directional derivative of \( f \):

\[
v f = df_x v = D_v f = \left. \frac{d}{dt} \right|_{t=0} f(\gamma(t))
\]

where \( \gamma(t) \) is any differentiable path such that \( \gamma(0) = x \) and \( \gamma'(0) = v \). The denote the sphere of radius 1 centered at the origin by \( S^{n-1} = \{ x \in \mathbb{R}^n : \|x\| = 1 \} \).

(a) Show that \( \text{grad}_x f \) is orthogonal the kernel of \( d f_x : T_x \mathbb{R}^n \to \mathbb{R} \).

(b) Show that \( \| \text{grad}_x f \| \) is the maximum rate of change of \( f \) along any direction:

\[
\max_{\|v\|=1} \| df_x v \| = \| \text{grad}_x f \|
\]

and that the maximum is achieved when \( v = \text{grad}_x f / \| \text{grad}_x f \| \). (Schwarz inequality may be useful.)

(c) Let \( A \in M(n, \mathbb{R}) \) be a symmetric matrix and define \( f : \mathbb{R}^n \to \mathbb{R} \) by

\[
f(x) = \frac{1}{2} \langle Ax, x \rangle.
\]

Show that if \( x \in S^n \) is a point where \( f \) achieves its maximum or its minimum valued, then \( x \) is an eigenvector of \( A \).

(d) Show that there exists an orthonormal basis of \( \mathbb{R}^n \) consisting of eigenvectors of \( A \). For this, use a finite induction starting from the existence of one eigenvector, then restricting the function to the intersection of the sphere with the subspace orthogonal to the previously obtained eigenvectors.

3. **Differential of the determinant function.** Let us denote by \( GL(n, \mathbb{R}) \) the set of all invertible \( n \times n \) real matrices. This is an open set in the vector space of all \( n \times n \) real matrices \( M_n(\mathbb{R}) \). It is not hard to check (you need not do it here) that \( GL(n, \mathbb{R}) \) is, in fact, a group under matrix multiplication, known as the general linear group.

Let \( D \) be the determinant function, which we restrict here to \( GL(n, \mathbb{R}) \). You are going show that \( D \) is everywhere differentiable and that its directional derivative at \( A \) in direction \( W \) is

\[
dD_A W = D(A) \text{tr}(WA^{-1})
\]

for all \( A \in GL(n, \mathbb{R}) \) and all \( W \in M(n, \mathbb{R}) \). I suggest breaking up the proof into a few subproblems:

(a) It is not hard to prove (you need not do it here) that the determinant, regarded as a multilinear function on the columns of the matrix argument, is uniquely characterized by the following properties: Let \( A_1, \ldots, A_n \) be the columns of the matrix \( A \). Then

\begin{itemize}
  \item \( A_i \in \mathbb{R}^n \to D(A_1, \ldots, A_n) \in \mathbb{R} \) is a linear functional for each \( i \);
  \item \( D(A_{\sigma(1)}, \ldots, A_{\sigma(n)}) = \text{sign}(\sigma) D(A_1, \ldots, A_n) \) for all permutations \( \sigma \in S_n \);
  \item \( D(e_1, \ldots, e_n) = 1 \) where \( \{e_1, \ldots, e_n\} \) is the standard basis of \( \mathbb{R}^n \). In other words, \( D(I) = 1 \).
\end{itemize}

By only using these three properties, show that

\[
D \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.
\]

(b) Explain why \( D \) is a smooth function.
(c) Using the above characterization of determinant, show that

\[ D(AB) = D(A)D(B) \]

for all matrices \( A, B \in M_n(\mathbb{R}) \). I suggest the following argument: First assume that \( A \in GL(n, \mathbb{R}) \) and define a function

\[ D_A(B) := \frac{D(AB_1, \ldots, AB_n)}{D(A)} \]

where \( B_1, \ldots, B_n \) are the columns of \( B \). Argue that \( D_A \) also satisfies the above three properties, so \( D_A = D \). Now use the fact that \( GL(n, \mathbb{R}) \) is a dense subset of \( M_n(\mathbb{R}) \). (Why is \( GL(n, \mathbb{R}) \) dense in \( M_n(\mathbb{R}) \)? Recall that a square matrix is not invertible if and only if its columns are linearly dependent vectors. Convince yourself that if one column of \( A \) is a linear combination of the others, then \( D(A) = 0 \).)

(d) Show that \( D(A + tW) = D(A)D(I + tWA^{-1}) \).

(e) Show that \( dD_IW = \text{tr}(W) \).

(f) Finally show that \( dD_AW = D(A)\text{tr}(WA^{-1}) \).

4. Using the facts of the previous problem and the definition of wedge product, show the following.

(a) The determinant of \( A \) may also be uniquely characterized by the identity

\[ (Ae_1) \wedge \cdots \wedge (Ae_n) = D(A)(e_1 \wedge \cdots \wedge e_n) \]

where \( \{e_1, \ldots, e_n\} \) is the standard basis of \( \mathbb{R}^n \).

(b) If \( \{f_1, \ldots, f_n\} \) is any other basis of \( \mathbb{R}^n \), show that

\[ (Af_1) \wedge \cdots \wedge (Af_n) = D(A)(f_1 \wedge \cdots \wedge f_n) \]

(Note that we can write \( f_i = Be_i \) for some (invertible) matrix \( B \). Apply what you proved in one of the items of the previous problem.)