1. (a) The Chain Rule for differentiable functions $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^p$ states that $f \circ g : \mathbb{R}^n \to \mathbb{R}^p$ is also differentiable, and that at each point $x \in \mathbb{R}^n$,

$$[D(f \circ g)(x)] = [Df(g(x))] \circ [Dg(x)].$$

(b) Writing $x = (x_1, x_2)$, we have that $h(x_1, x_2) = g(x_1^2 + x_2^2)$. To find the partial derivatives of $h$, we compute using $g$ and the chain rule:

$$D_1 h \left( \begin{array}{c} x_1 \\ x_2 \end{array} \right) = D_1 g(x_1^2 + x_2^2) = 2x_1 g'(x_1^2 + x_2^2)$$

$$D_2 h \left( \begin{array}{c} x_1 \\ x_2 \end{array} \right) = D_2 g(x_1^2 + x_2^2) = 2x_2 g'(x_1^2 + x_2^2).$$

The equation $x_2 D_1 h(x) = x_1 D_2 h(x)$ is now seen to be true by inspection.

2. (a) A $k$-dimensional manifold in $\mathbb{R}^n$ is a subset $M$ of $\mathbb{R}^n$ such that, at each point $x \in M$, there is an open set $U$ containing $x$ with the property that $M \cap U$ is the graph of a $C^1$ function of $k$ variables.

The Implicit Function Theorem can be used to prove that a set is a manifold in the following way: suppose that $f : \mathbb{R}^n \to \mathbb{R}^{n-k}$ is a $C^1$ function, and $M \subset \mathbb{R}^n$ is defined by $M = f^{-1}(0)$. Suppose, moreover, that at each point $x \in M$, the derivative map $Df(x) : \mathbb{R}^n \to \mathbb{R}^{n-k}$ is onto. Then the Implicit Function Theorem says precisely that at each point $x \in M$, there is a neighborhood $U$ of $x$ and a function $g : \mathbb{R}^k \to \mathbb{R}^{n-k}$ such that the set $U \cap M$ is the graph of $g$, where the domain variables of $g$ are the variables corresponding to the non-pivotal columns of $Df(x)$. Therefore $M$ is a manifold, by the above definition.

(b) Define $f : \text{Mat}_{2 \times 3} \to \text{SMat}_{2 \times 2}$ by

$$f(A) = AA^\top - \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix};$$

then $U = f^{-1}(0)$. The derivative of this map is

$$Df(A) : H \mapsto AH^\top + HA^\top.$$

In coordinates this becomes

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \end{pmatrix} + \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix} = \begin{pmatrix} 2a_{11}h_{11} + 2a_{12}h_{12} + 2a_{13}h_{13} \\ a_{11}h_{21} + a_{12}h_{22} + a_{13}h_{23} + a_{21}h_{11} + a_{22}h_{12} + a_{23}h_{13} + 2a_{21}h_{21} + 2a_{22}h_{22} + 2a_{23}h_{23} \end{pmatrix}.$$ 

We can write this as a map $\mathbb{R}^6 \to \mathbb{R}^3$ (making use of the fact that the image of $f$, hence of $Df(A)$, lies in the space of symmetric matrices):

$$\begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \end{pmatrix} \mapsto \begin{pmatrix} 2a_{11} & 2a_{12} & 2a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 2a_{21} \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \end{pmatrix}.$$

The first and third rows are clearly linearly independent when at least one element of each row of $A$ is non-zero; but this is necessary, because the equation $AA^T = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ says, in particular, that the length of the first row of $A$ is 2 and the length of the second row of $A$ is 3. The second row is in the span of the first and third rows iff the first and second rows of $A$ are linearly dependent; but the equation $AA^T = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ also implies that the rows of $A$ are orthogonal, which means that they are linearly independent (because they are non-zero). Therefore $Df(A)$ is onto at every $A \in U$, and $U$ is a manifold.

(c) At $A_0$, the tangent space is the set of all $H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \\ h_{13} & h_{23} \end{pmatrix} + \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} h_{11} - h_{12} + h_{11} - h_{12} & h_{21} - h_{22} + h_{11} + h_{12} + h_{13} \\ h_{11} + h_{12} + h_{13} + h_{21} - h_{22} & h_{21} + h_{22} + h_{23} + h_{21} + h_{22} + h_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$ 

The $(1,2)$ and $(2,1)$ entries of this matrix are identical, so we ignore one of them. This leaves us with the equations $h_{11} = h_{22}$, $h_{21} + h_{12} + h_{13} = 0$, and $h_{21} + h_{22} + h_{23} = 0$. By the last two equations, $h_{12} + h_{13} = h_{22} + h_{23}$. Therefore a basis for the tangent space $T_{A_0}U$ is

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

from the equations $h_{11} = h_{22}$, $h_{21} = -h_{22} - h_{23}$, and $h_{12} = h_{22} - h_{23} - h_{13}$. 


3. (a) Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a $C^1$ function and let $a_0 \in \mathbb{R}^n$. Suppose $Df(a_0)$ is invertible, and set

$$a_1 = a_0 - [Df(a_0)]^{-1} f(a_0).$$

Suppose further that $Df(x)$ satisfies a Lipschitz condition with Lipschitz constant $M$ on $B_r(a_1)$, where $r = |a_1 - a_0| = |[Df(a_0)]^{-1} f(a_0)|$, i.e.,

$$|Df(x) - Df(y)| \leq M |x - y| \quad \text{for all } x, y \in B_r(a_1).$$

Then Kantorovich’s Theorem says that the equation $f(x) = 0$ has a unique solution in $B_r(a_1)$, provided that the following inequality is satisfied:

$$M \cdot |[Df(a_0)]^{-1}|^2 \cdot |f(a_0)| \leq \frac{1}{2}.$$

Moreover, Newton’s method starting at $a_0$ converges to this solution.

(b) The derivative of $p$ is $7x^6 - 1$. At $a_0 = 2$, this derivative becomes $p'(2) = 7 \cdot 2^6 - 1 = 7 \cdot 64 - 1 = 447$. Thus the first step of Newton’s method gives the next guess as

$$a_1 = 2 - \frac{1}{447}(-2) = 2 + \frac{2}{447} = \frac{896}{447}.$$

Now we look for a Lipschitz constant on $[2, 2.01]$ (since $2/447 < 2/400 = .005$, this interval contains the ball we’re interested in). We have:

$$|7x^6 - 1 - (7y^6 - 1)| = 7|x^6 - y^6|$$
$$= 7|x - y||x^5 + x^4y + x^3y^2 + x^2y^3 + xy^4 + y^5|$$
$$\leq 42|x - y||2.01|^5 \leq 42 \cdot 33 \cdot |x - y|.$$

Now we check the Kantorovich inequality:

$$(42)(33) \left(\frac{1}{447^2}\right)(2) \leq \frac{42^2}{447^2} \cdot 2 \leq \frac{2}{10^2} \leq \frac{1}{2}.$$

Therefore Kantorovich’s Theorem implies that Newton’s method will converge to a root in the interval $(2, 2.01)$.

[Note: We could have computed a Lipschitz constant for $p'$ by finding a bound for the second derivative, instead of using the factorization of $x^6 - y^6$. In this case, we even get the same constant. The second derivative of $p$ is $p''(x) = 42x^5$, which on $[2, 2.01]$ is bounded by $42 \cdot |2.01|^5 \approx 42 \cdot 33.$]
4. (a) The fact that $I$ is linear follows immediately from the properties of integrals:

$$I(p + q) = \int_{-1}^{1} (p(x) + q(x)) \, dx = \int_{-1}^{1} p(x) \, dx + \int_{-1}^{1} q(x) \, dx = I(p) + I(q)$$

$$I(cp) = \int_{-1}^{1} cp(x) \, dx = c \int_{-1}^{1} p(x) \, dx = cI(p).$$

(b) Let $A$ be a $k \times n$ matrix, with rank $r$. Then the Fundamental Theorem of Linear Algebra states that

$$null(A) = (row(A))^\perp$$

and

$$col(A) = (null(A^T))^\perp.$$

$null(A)$ is the nullspace of $A$, with dimension $n - r$. $col(A)$ is the column space of $A$, and $row(A)$ is the row space of $A$ (i.e., the column space of $A^T$); these both have dimension $r$. $null(A^T)$ therefore has dimension $k - r$.

(c) Let $bx \in W$. We compute directly:

$$I(bx) = \int_{-1}^{1} bx \, dx = \frac{b}{2} x^2 \bigg|_{-1}^{1} = \frac{b}{2} - \frac{b}{2} = 0,$$

and therefore $W \subset \ker I$. It is not the full kernel of $I$, however: $I$ maps from a three-dimensional space to a one-dimensional space, so its kernel must have at least dimension 2, by the rank-nullity formula. But $W$ is only one-dimensional.
5. (a) The partial derivatives of \( f \) are

\[
D_1 f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = y - 1, \quad D_2 f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x + 1, \quad \text{and} \quad D_3 f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2z.
\]

These all vanish only at the point \( x_0 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \), and therefore this is the only critical point of \( f \). \( f \) is a quadratic polynomial, so its Taylor polynomial of degree two at \( x_0 \) is

\[
P^2_{f,x_0} \begin{pmatrix} -1 + h_x \\ 1 + h_y \\ 0 + h_z \end{pmatrix} = f \begin{pmatrix} 1 + h_x \\ 1 + h_y \\ 0 + h_z \end{pmatrix} = (-1 + h_x)(1 + h_y) - (-1 + h_x) + (1 + h_y) + (h_z)^2
\]

\[= -1 + h_x - h_y + h_x h_y + 1 - h_x + 1 + h_y + h_z^2
\]

\[= 1 + h_x h_y + h_z^2.
\]

The quadratic terms yield the quadratic form

\[
\frac{1}{4} ((h_x + h_y)^2 - (h_x - h_y)^2) + h_z^2,
\]

which has signature (2, 1). Therefore \( x_0 \) is a saddle of \( f \).

(b) \( F \) is our constraint function. Its Jacobian is

\[
DF \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 2(x + 1) & 2(y - 1) & 2z \end{bmatrix}.
\]

By the Lagrange multiplier theorem, a constrained critical point on \( S_{\sqrt{2}}(x_0) \) occurs at a point where

\[
Df \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda DF \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \text{i.e.,} \quad \begin{cases} y - 1 = 2\lambda(x + 1) \\ x + 1 = 2\lambda(y - 1) \\ 2z = 2\lambda z \end{cases}.
\]

The final equation says either \( \lambda = 1 \) or \( z = 0 \).

If \( z \neq 0 \), solving the other equations yields \( x = -1, y = 1 \). Then the constraint function \( F \) shows that \( z = \pm \sqrt{2} \). Thus \( \begin{pmatrix} -1 \\ 1 \\ \pm \sqrt{2} \end{pmatrix} \) are constrained critical points of \( f \).

If \( z = 0 \), then the first two equations imply that \( \lambda = \pm 1/2 \) or \( \lambda = 0 \). But if \( \lambda = 0 \), then \( y = 1 \) and \( x = -1 \), yielding the point \( x_0 \), which is not on \( S_{\sqrt{2}}(x_0) \). Therefore \( \lambda = \pm 1/2 \). In the case \( \lambda = 1/2 \), we get \( y = x + 2 \), while if \( \lambda = -1/2 \),
\[ y = -x. \] Both of these lead to \( x = 0, -2. \) Thus \( \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} , \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix} , \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} , \text{ and} \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix} \) are also constrained critical points of \( f. \)

(c) We showed in part (a) that the only critical point in the interior of the ball is at \( x_0, \) and that this point is a saddle; therefore it cannot be an extremum. We compute \( f \) at the constrained critical points (on the surface of the ball) found in (b):

\[
\begin{align*}
f \begin{pmatrix} -1 \\ 1 \\ \pm \sqrt{2} \end{pmatrix} &= 3 \\
f \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} &= 2 \\
f \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix} &= 2 \\
f \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} &= 0 \\
f \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix} &= 0
\end{align*}
\]

Therefore the maximum of \( f \) on \( B_{\sqrt{2}}(x_0) \) is 3 and the minimum is 0.

6. The vector \( \begin{pmatrix} a \\ b \\ c \end{pmatrix} \) is orthogonal to \( H. \) Hence the normalized vector

\[
v = \frac{1}{\sqrt{a^2 + b^2 + c^2}} \begin{pmatrix} a \\ b \\ c \end{pmatrix}
\]

is a normal vector to \( H \) at every point. (\( v \) is well-defined because \( a, b, \) and \( c \) are not all zero.) Thus the Gauss map \( H \rightarrow S^2 \) is the constant map \( x \mapsto v \) for all \( x \in H. \) The derivative of this map is the zero map at every point, which has determinant zero. Therefore the Gaussian curvature of \( H \) is 0 everywhere.