These solutions are being shared for the benefit of future Math 516 graders; please do not circulate among other students. Note also that these solutions may have typos and they may be incomplete - many problems can be solved with more than one approach.

**Exercise 1.1.** Given a collection of real $m \times n$ matrices $A_1, A_2, \ldots, A_l$, define the linear mapping $A : \mathbb{R}^{m \times n} \to \mathbb{R}^l$ by setting

$$A(X) := (\langle A_1, X \rangle, \langle A_2, X \rangle, \ldots, \langle A_l, X \rangle).$$

Show that the adjoint is the mapping $A^* y = y_1 A_1 + y_2 A_2 + \ldots + y_l A_l$.

**Proof.** Let $X \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^l$. Then

$$\langle A(X), Y \rangle = \langle (\langle A_1, X \rangle, \ldots, \langle A_l, X \rangle), y \rangle$$

$$= \sum_{i=1}^{l} \langle A_i, X \rangle \cdot y_i$$

$$= \langle \sum_{i=1}^{l} y_i A_i, X \rangle$$

$$= \langle A^* y, X \rangle.$$

□

**Exercise 1.2.** Given a positive definite linear operator $A$ on $E$, show that the assignment $\langle v, w \rangle_A := \langle Av, w \rangle$ is an inner product on $E$, with the induced norm $\|v\|_A = \sqrt{\langle Av, v \rangle}$. Show that the dual norm with respect to the original inner product is $\|v\|_{A^{-1}} = \sqrt{\langle A^{-1} v, v \rangle}$.

**Proof.** (Symmetry) Since $A$ is self-adjoint and the original inner product is symmetric, we have

$$\langle v, w \rangle_A = \langle Av, w \rangle = \langle v, A^* w \rangle = \langle v, Aw \rangle = \langle Aw, v \rangle = \langle w, v \rangle_A.$$

(Bilinearity) Follows from the fact that $A$ is a linear operator and the original inner product is bilinear:

$$\langle av_1 + bv_2, w \rangle_A = \langle A(av_1 + bv_2), w \rangle = \langle aAv_1 + bAv_2, w \rangle$$

$$= a\langle Av_1, w \rangle + b\langle Av_2, w \rangle$$

$$= a\langle v_1, w \rangle_A + b\langle v_2, w \rangle_A.$$
(Positive Definiteness) Follows immediately from the positive definiteness of $A$.

The induced norm is as stated (by definition), and the dual norm is
\[
\|v\|_A^* = \max \{ \langle v, x \rangle : \|x\|_A \leq 1 \}
= \max \{ \langle v, x \rangle : \langle Ax, x \rangle \leq 1 \}
\]
Using the Lagrangian, you can show that the $x$ obtaining the maximum above is
\[
x = \frac{A^{-1}v}{\|v\|_{A^{-1}}}
\]
which yields
\[
\|v\|_A^* = \langle v, \frac{A^{-1}v}{\|v\|_{A^{-1}}} \rangle = \|v\|_{A^{-1}}.
\]
Alternatively, do a change of variables with $y = A^{1/2}x$ and apply Cauchy-Schwarz.\qed

**Exercise 1.3.** Equip $\mathbb{R}^n$ and $\mathbb{R}^m$ with the $l_p$-norms. Then for any matrix $A \in \mathbb{R}^{m \times n}$, show the equalities
\[
\|A\|_1 = \max_{j=1,\ldots,n} \|A_{\cdot j}\|_1
\]
\[
\|A\|_\infty = \max_{i=1,\ldots,n} \|A_{i \cdot}\|_1
\]
where $A_{\cdot j}$ and $A_{i \cdot}$ denote the $j$'th column and $i$'th row of $A$, respectively.

**Proof.** First we show the second equality:
\[
\|A\|_\infty = \max_{\|x\|_\infty \leq 1} \|Ax\|_\infty
= \max_{\|x\|_\infty \leq 1} \max_{i=1,\ldots,m} |(Ax)_i|
= \max_{i=1,\ldots,m} \max_{\|x\|_\infty \leq 1} |A_{i \cdot}x|
= \max_{i=1,\ldots,m} \|A_{i \cdot}\|_1
\]
where the last line follows from the definition of the dual norm to $\| \cdot \|_\infty$ in $\mathbb{R}^n$.

Now we consider the first equality in the exercise. Using the fact that $\| \cdot \|_\infty$ and $\| \cdot \|_\infty$ are dual in $\mathbb{R}^n$, we have
\[
\|A\|_1 = \max_{\|x\|_1 \leq 1} \|Ax\|_1
= \max_{\|x\|_1 \leq 1} \max_{\|y\|_\infty \leq 1} \langle Ax, y \rangle
= \max_{\|y\|_\infty \leq 1} \max_{\|x\|_1 \leq 1} \langle x, A^*y \rangle
= \max_{\|y\|_\infty \leq 1} \|A^*y\|_\infty
= \max_{\|y\|_\infty \leq 1} \max_{j=1,\ldots,n} |(A^*y)_j|
= \max_{j=1,\ldots,n} \max_{\|y\|_\infty \leq 1} |A_{\cdot j}y|
= \max_{j=1,\ldots,n} \|A_{\cdot j}y\|_1.
Exercise 1.6. Define the function

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

where \( A : E \to E \) is a linear operator, \( v \) is lies in \( E \), and \( c \) is a real number.

1. Show that if \( A \) is replaced by the self-adjoint operator \( (A + A^*)/2 \), the function values \( f(x) \) remain unchanged.
2. Assuming \( A \) is self-adjoint derive the equations:
   \[ \nabla f(x) = Ax + v \quad \text{and} \quad \nabla^2 f(x) = A. \]
3. Using parts 1 and 2, describe \( \nabla f(x) \) and \( \nabla^2 f(x) \) when \( A \) is not necessarily self-adjoint.

Proof.
1. We have
   \[
   \left( \frac{A + A^*}{2} \right) x, x = \frac{1}{2} \langle Ax, x \rangle + \frac{1}{2} \langle A^* x, x \rangle \\
   = \frac{1}{2} \langle Ax, x \rangle + \frac{1}{2} \langle x, Ax \rangle \\
   = \langle Ax, x \rangle
   \]
   and thus \( f(x) \) is unchanged by replacing \( A \) with \( \frac{A + A^*}{2} \).
2. Assuming \( A = A^* \), we have
   \[
   f(x + h) - f(x) - \langle Ax + v, h \rangle \\
   = \frac{1}{2} \langle A(x + h), x + h \rangle - \frac{1}{2} \langle Ax, x \rangle + \langle v, x + h \rangle - \langle v, x \rangle - \langle Ax + v, h \rangle \\
   = \frac{1}{2} \langle Ax, h \rangle + \frac{1}{2} \langle Ah, x \rangle + \frac{1}{2} \langle Ah, h \rangle - \langle Ax + v, h \rangle \\
   = \frac{1}{2} \langle Ax, h \rangle + \frac{1}{2} \langle h, Ax \rangle + \frac{1}{2} \langle Ah, h \rangle - \langle Ax, h \rangle \\
   = \frac{1}{2} \langle Ah, h \rangle
   \]
   Dividing by \( \|h\| \) and letting \( h \to 0 \), we obtain 0, and thus \( \nabla f(x) = Ax + v \) as desired. Next we consider
   \[
   \nabla f(x + h) - \nabla f(x) - Ah \\
   = \nabla f(x + h) - \nabla f(x) \\
   = \nabla f(x + h) - \nabla f(x) - Ah \\
   = 0
   \]
   which is clearly \( o(\|h\|) \), and thus \( \nabla^2 f(x) = A \).
3. Using parts (1) and (2), when \( A \) is not necessarily self-adjoint we have
   \[
   \nabla f(x) = \left( \frac{A + A^*}{2} \right) x + v
   \]
   and
   \[
   \nabla^2 f(x) = \frac{A + A^*}{2}. \]
Exercise 1.7. Define the function \( f(x) = \frac{1}{2} \|F(x)\|^2 \), where \( F: \mathbb{E} \to \mathbb{Y} \) is a \( C^1 \)-smooth mapping. Prove the identity \( \nabla f(x) = \nabla F(x)^* F(x) \).

Proof. We have

\[
\begin{align*}
(f(x+h) - f(x) - \langle \nabla F(x)^* F(x), h \rangle & = \frac{1}{2} \|F(x+h)\|^2 - \frac{1}{2} \|F(x)\|^2 - \langle \nabla F(x)^* F(x), h \rangle \\
& = \frac{1}{2} \|F(x) + \nabla F(x)h + o(\|h\|)\|^2 - \frac{1}{2} \|F(x)\|^2 - \langle \nabla F(x)^* F(x), h \rangle \\
& = \langle F(x), \nabla F(x)h + o(\|h\|) \rangle + \frac{1}{2} \|\nabla F(x)h + o(\|h\|)\|^2 - \langle \nabla F(x)^* F(x), h \rangle \\
& = \langle F(x), o(\|h\|) \rangle + \frac{1}{2} \|\nabla F(x)h + o(\|h\|)\|^2 \\
& = o(\|h\|)
\end{align*}
\]

and thus \( \nabla f(x) = \nabla F(x)^* F(x) \) as desired. \( \square \)

Exercise 1.8. Consider a function \( f: U \to \mathbb{R} \) and a linear mapping \( \mathcal{A}: \mathbb{Y} \to \mathbb{E} \) and define the composition \( h(x) = f(\mathcal{A}x) \).

1. Show that if \( f \) is differentiable at \( \mathcal{A}x \), then

\[
\nabla h(x) = \mathcal{A}^* \nabla f(\mathcal{A}x).
\]

2. Show that if \( f \) is twice differentiable at \( \mathcal{A}x \), then

\[
\nabla^2 h(x) = \mathcal{A}^* \nabla^2 f(\mathcal{A}x) \mathcal{A}.
\]

Proof. (1) We have

\[
\begin{align*}
h(x+\epsilon) - h(x) - \langle \mathcal{A}^* \nabla f(\mathcal{A}x), \epsilon \rangle & = f(\mathcal{A}x + \mathcal{A}\epsilon) - f(\mathcal{A}x) - \langle \mathcal{A}^* \nabla f(\mathcal{A}x), \epsilon \rangle \\
& = f(\mathcal{A}x) + \langle \nabla f(\mathcal{A}x), \mathcal{A}\epsilon \rangle + o(\|\mathcal{A}\epsilon\|) - f(\mathcal{A}x) - \langle \mathcal{A}^* \nabla f(\mathcal{A}x), \epsilon \rangle \\
& = \langle \mathcal{A}^* \nabla f(\mathcal{A}x), \epsilon \rangle + o(\|\mathcal{A}\epsilon\|) - \langle \mathcal{A}^* \nabla f(\mathcal{A}x), \epsilon \rangle \\
& = o(\|\mathcal{A}\epsilon\|) \\
& = o(\|\epsilon\|)
\end{align*}
\]

and thus \( \nabla h(x) = \mathcal{A}^* \nabla f(\mathcal{A}x) \).

(2) Now using part (1), we have

\[
\begin{align*}
\nabla h(x+\epsilon) - \nabla h(x) - \left( \mathcal{A}^* \nabla^2 f(\mathcal{A}x) \mathcal{A} \right) \epsilon & = \mathcal{A}^* \nabla f(\mathcal{A}x + \mathcal{A}\epsilon) - \mathcal{A}^* \nabla f(\mathcal{A}x) - \mathcal{A}^* \nabla^2 f(\mathcal{A}x) \mathcal{A}\epsilon \\
& = \mathcal{A}^* \left( \nabla f(\mathcal{A}x + \mathcal{A}\epsilon) - \nabla f(\mathcal{A}x) \right) - \mathcal{A}^* \nabla^2 f(\mathcal{A}x) \mathcal{A}\epsilon \\
& = \mathcal{A}^* \left( \nabla^2 f(\mathcal{A}x) \mathcal{A}\epsilon + o(\|\mathcal{A}\epsilon\|) \right) - \mathcal{A}^* \nabla^2 f(\mathcal{A}x) \mathcal{A}\epsilon \\
& = o(\|\mathcal{A}\epsilon\|) \\
& = o(\|\epsilon\|)
\end{align*}
\]
and thus $\nabla^2 h(x) = A^* \nabla^2 f(Ax)A$ as desired.

Exercise 1.9. Consider a mapping $F(x) = G(H(x))$ where $H$ is differentiable at $x$ and $G$ is differentiable at $H(x)$. Derive the formula $\nabla F(x) = \nabla G(H(x)) \nabla H(x)$.

Proof. To derive the chain rule, we first use differentiability of $H$ and then differentiability of $G$:

$$F(x + h) = G(H(x) + \nabla H(x)h + o(||h||))$$
$$= G(H(x)) + \nabla G(H(x))(\nabla H(x)h + o(||h||)) + o(||\nabla H(x)h + o(||h||)||)$$
$$= F(x) + \nabla G(H(x))\nabla H(x)h + o(||h||).$$

This shows that $\nabla F(x) = \nabla G(H(x)) \nabla H(x)$ as desired.

Exercise 1.10. Define the two sets

$$R^n_{++} := \{ x \in \mathbb{R}^n : x_i > 0 \text{ for all } i = 1, \ldots, n \},$$
$$S^n_{++} := \{ X \in \mathbb{S}^n : X \succ 0 \}.$$

Consider the two functions $f : R^n_{++} \rightarrow \mathbb{R}$ and $F : S^n_{++} \rightarrow \mathbb{R}$ given by

$$f(x) = -\sum_{i=1}^n \log x_i \quad \text{and} \quad F(X) = -\ln \det(X),$$

respectively. Note, from basic properties of the determinant, the equality $F(X) = f(\lambda(X))$, where we set $\lambda(X) := (\lambda_1(X), \ldots, \lambda_n(X))$.

(1) Find the derivatives $\nabla f(x)$ and $\nabla^2 f(x)$ for $x \in R^n_{++}$.

(2) Using the property $\text{tr}(AB) = \text{tr}(BA)$, prove $\nabla F(X) = -X^{-1}$ and $\nabla^2 F(X)[V] = X^{-1} VX^{-1}$ for any $X \succ 0$.

[Hint: To compute $\nabla F(X)$, justify $F(X + tV) - F(X) + t\langle X^{-1}, V \rangle = -\ln \det(I + X^{-1/2}VX^{-1/2}) + \text{tr}(X^{-1/2} VX^{-1/2})$.

By rewriting the expression in terms of eigenvalues of $X^{-1/2} VX^{-1/2}$, deduce that the right-hand-side is $o(t)$. To compute the Hessian, observe

$$(X + V)^{-1} = X^{-1/2} (I + X^{-1/2} VX^{-1/2})^{-1} X^{-1/2},$$

and then use the expansion

$$(I + A)^{-1} = I - A + A^2 - A^3 + \ldots = I - A + O(||A||_{\text{op}}^2),$$

whenever $||A||_{\text{op}} < 1$.]

(3) Show

$$\langle \nabla^2 F(X)[V], V \rangle = ||X^{-\frac{1}{2}} VX^{-\frac{1}{2}}||_F^2$$

for any $X \succ 0$ and $V \in S^n$. Deduce that the operator $\nabla^2 F(X) : S^n \rightarrow S^n$ is positive definite.
Proof. (1) Straightforward calculations give
\[ \nabla f(x) = \left( -\frac{1}{x_1}, \ldots, -\frac{1}{x_n} \right) \]
and
\[ \nabla^2 f(x) = \text{Diag} \left( \frac{1}{x_1^2}, \ldots, -\frac{1}{x_n^2} \right) \]

(2) To check the formula for \( \nabla F(X) \), we show
\[
F(X + tV) - F(X) + t\langle X^{-1}, V \rangle
= -\ln \det(X + tV) + \ln \det(X) + t \cdot \text{tr} (X^{-1/V}X^{-1/2})
\]
\[
= -\ln \left( \frac{\det(X + tV)}{\det(X)} \right) + t \cdot \text{tr} (X^{-1/2}VX^{-1/2})
\]
\[
= -\ln \det \left( X^{-1/2}(X + tV)X^{-1/2} \right) + t \cdot \text{tr} (X^{-1/2}VX^{-1/2})
\]
\[
= -\ln \det \left( I + tX^{-1/2}VX^{-1/2} \right) + t \cdot \text{tr} (X^{-1/2}VX^{-1/2})
\]
\[
= -\ln \left( \prod_{i=1}^{n} (1 + t\lambda_i) \right) + t \sum_{i=1}^{n} \lambda_i
\]
where \( \lambda_i \) are the eigenvalues of \( X^{-1/2}VX^{-1/2} \). Thus we have
\[
F(X + tV) - F(X) + t\langle X^{-1}, V \rangle = -\sum_{i=1}^{n} \ln(1 + t\lambda_i) + t \sum_{i=1}^{n} \lambda_i
\]

To show that the right hand side is \( o(t) \), we use the Taylor series of \( \ln(1 + x) \):
\[
\lim_{t \to 0} \frac{F(X + tV) - F(X) + t\langle X^{-1}, V \rangle}{t} = \lim_{t \to 0} \sum_{i=1}^{n} \frac{-\ln(1 + t\lambda_i) + t\lambda_i}{t}
\]
\[
= \lim_{t \to 0} \frac{1}{t} \sum_{i=1}^{n} \left[ t\lambda_i - \frac{(t\lambda_i)^2}{2} + \frac{(t\lambda_i)^3}{3} - \ldots \right] + t\lambda_i
\]
\[
= \lim_{t \to 0} \sum_{i=1}^{n} \frac{(t\lambda_i)^2}{2t} - \frac{(t\lambda_i)^3}{3t} + \ldots
\]
\[
= 0.
\]

From this, we conclude that \( \nabla F(X) = -X^{-1} \), as desired.

Now we consider the Hessian. First we note that
\[
(X + V)^{-1} = X^{-1/2} (I + X^{-1/2}VX^{-1/2})^{-1} X^{-1/2}
\]

Assuming \( \|X^{-1/2}VX^{-1/2}\|_F < 1 \), we can expand the middle term as
\[
(I + X^{-1/2}VX^{-1/2})^{-1} = I - X^{-1/2}VX^{-1/2} + (X^{-1/2}VX^{-1/2})^2 - (X^{-1/2}VX^{-1/2})^3 + \ldots
\]

Thus for \( t \) sufficiently small, we have
\[
\nabla F(X + tV) = -(X + tV)^{-1}
\]
\[
= -X^{-1/2} (I - tX^{-1/2}VX^{-1/2} + t^2(X^{-1/2}VX^{-1/2})^2 - t^3(X^{-1/2}VX^{-1/2})^3 + \ldots) X^{-1/2}
\]
\[
= -X^{-1} + tX^{-1}VX^{-1} - t^2(X^{-1}V)^2X^{-1} + t^3(X^{-1}V)^3X^{-1} - \ldots
\]
This gives us
\[
\nabla F(X+tV) - \nabla F(X) - tX^{-1}VX^{-1}
\]
\[
= -t^2(X^{-1}V)^2X^{-1} + t^3(X^{-1}V)^3X^{-1} - \ldots
\]
\[
= t^2(X^{-1}V)^2 (-I + tX^{-1}V - t^2(X^{-1}V)^2 - \ldots) X^{-1}
\]
\[
= -t^2(X^{-1}V)^2 (I + tX^{-1}V)^{-1} X^{-1}
\]
which is \( o(t) \) as desired.

(3) Using part (2), we have
\[
\langle \nabla^2 F(X)[V], V \rangle = \langle X^{-1}VX^{-1}, V \rangle
\]
and since \( X \) is positive definite, we can write \( X^{-1} = X^{-1/2}X^{-1/2} \) with \( X^{-1/2} \) being self-adjoint. Thus we get
\[
\langle \nabla^2 F(X)[V], V \rangle = \langle X^{-1/2}X^{-1/2}VX^{-1/2}X^{-1/2}, V \rangle
\]
\[
= \langle X^{-1/2}VX^{-1/2}, X^{-1/2}VX^{-1/2} \rangle
\]
\[
= \|X^{-1/2}VX^{-1/2}\|^2
\]
This shows that \( \nabla^2 F(X) \) is a positive definite operator on \( S^n \), since the above quantity is always nonnegative and is zero if and only if \( X^{-1/2}VX^{-1/2} = 0 \), which is equivalent to \( V = 0 \).

Exercise 1.11. Consider a function \( f: U \to \mathbb{R} \) and two points \( x, y \in U \). Define the univariate function \( \varphi: [0, 1] \to \mathbb{R} \) given by \( \varphi(t) = f(x + t(y-x)) \) and let \( x_t := x + t(y-x) \) for any \( t \).

(1) Show that if \( f \) is \( C^1 \)-smooth, then equality
\[
\varphi'(t) = \langle \nabla f(x_t), y-x \rangle
\]
holds for any \( t \in (0, 1) \).

(2) Show that if \( f \) is \( C^2 \)-smooth, then equality
\[
\varphi''(t) = \langle \nabla^2 f(x_t)(y-x), y-x \rangle
\]
holds for any \( t \in (0, 1) \).

Proof. Both parts follow from Exercise 1.9 (the chain rule).

(1) Set \( F = \varphi, G = f \) and \( H = x_t \) as a function of \( t \).

(2) Set \( F = \varphi', G = \langle \nabla f(\cdot), y-x \rangle \) and \( H = x_t \) as a function of \( t \).

\[\Box\]

Exercise 1.15. Consider a \( C^1 \)-smooth mapping \( F: U \to \mathbb{R}^m \) and two points \( x, y \in U \). Derive the equations
\[
F(y) - F(x) = \int_0^1 \nabla F(x + t(y-x))(y-x) \, dt.
\]
\[
F(y) = F(x) + \nabla F(x)(y-x) + \int_0^1 (\nabla F(x + t(y-x)) - \nabla F(x))(y-x) \, dt.
\]
Proof. To see the first equation, fix an arbitrary basis and apply Theorem 1.12 coordinate-wise to each \((F(y))_i\) and \((F(x))_i\). To see the second equation, simply add and subtract the term \(\nabla F(x)(y - x)\) and note that this is a constant with respect to \(t\), so we can bring it inside the integral of length 1.

□

Exercise 1.16. Show that a \(C^1\)-smooth mapping \(F: U \to Y\) is \(L\)-Lipschitz continuous if and only if \(\|\nabla F(x)\| \leq L\) for all \(x \in U\).

Proof. First suppose that \(F\) is \(L\)-Lipschitz. Then for any \(y \neq x\) we have

\[
\frac{\|F(y) - F(x)\|}{\|y - x\|} \leq L
\]

which implies by differentiability of \(F\) that

\[
\frac{\|\nabla F(x)(y - x) + o(\|y - x\|)\|}{\|y - x\|} \leq L
\]

Letting \(y \to x\), we obtain

\[
\lim_{y \to x} \frac{\|\nabla F(x)\|}{\|y - x\|} \leq L
\]

which implies \(\|\nabla F(x)\| \leq L\) as desired.

For the reverse direction, suppose we have \(\|\nabla F(x)\| \leq L\) for all \(x \in E\). Then by Exercise 1.12, we have

\[
\|F(y) - F(x)\| = \left\| \int_0^1 \nabla F(x + t(y - x))(y - x) \, dt \right\|
\]

\[
\leq \int_0^1 \|\nabla F(x + t(y - x))\| \cdot \|y - x\| \, dt
\]

\[
\leq \int_0^1 L \cdot \|y - x\| \, dt
\]

\[
= L \cdot \|y - x\|
\]

which shows that \(F\) is \(L\)-Lipschitz as desired.

□