Homework 5

Self grades are due at 11 PM on February 23, 2024.

1. **Matrix Norm Calculations**

Let A have SVD equal to

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{1}$$

(a) Compute $\|A\|_F$, the Frobenius norm of A.

Solution: We have $\|A\|_F = \sqrt{\sigma_1\{A\}^2 + \sigma_2\{A\}^2}$, i.e., the square root of the sum of squared singular values of A, which gives

$$\|A\|_F = \sqrt{4^2 + 2^2} = \sqrt{20} = 2\sqrt{5}. \tag{2}$$

(b) Compute $\|A\|_2$, the spectral norm of A.

Solution: We have $\|A\|_2 = \sigma_1\{A\}$, i.e., the largest singular value of A, which gives $\|A\|_2 = 4$.

2. PCA and low-rank compression

We have a data matrix $X = \begin{bmatrix} \vec{x}_1^\top \\ \vec{x}_2^\top \\ \vdots \\ \vec{x}_n^\top \end{bmatrix}$ of size $n \times d$ containing n data points[1], $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$, with $\vec{x}_i \in \mathbb{R}^d$. Note that $\vec{x}_i^\top$ is the ith row of X. Assume that the data matrix is centered, i.e. each column of X is zero mean. In this problem, we will show equivalence between the following three problems:

(P_1) Finding a line going through the origin that maximizes the variance of the scalar projections of the points on the line. Formally P_1 solves the problem:

$$\underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmax}} \quad \vec{u}^\top C \vec{u} \tag{3}$$

with $C = \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i \vec{x}_i^\top$ denoting the covariance matrix associated with the centered data.

(P_2) Finding a line going through the origin that minimizes the sum of squares of the ℓ^2 distances from the points to their vector projections. Formally P_2 solves the minimization problem:

$$\underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{v_i \in \mathbb{R}} \left\| \vec{x}_i - v_i \vec{u} \right\|_2^2 . \tag{4}$$

Note that the vector projection of $\vec{x}$ on $\vec{u}$ is given by $v^\star \vec{u}$, where

$$v^\star = \underset{v \in \mathbb{R}}{\operatorname{argmin}} \left\| \vec{x} - v \vec{u} \right\|_2^2 , \tag{5}$$

and we will show that $v^\star = \langle \vec{x}, \ \vec{u} \rangle$ in part (a).

(P_3) Finding a rank-one approximation to the data matrix. Formally P_3 solves the minimization problem:

$$\underset{Y : \operatorname{rank}(Y) \leq 1}{\operatorname{argmin}} \quad \left\| X - Y \right\|_F . \tag{6}$$

Note that loosely speaking, two problems are said to be “equivalent" if the solution of one can be “easily" translated to the solution of the other. Some form of “easy" translations include adding/subtracting a constant or some quantity depending on the data points.

Note the significance of these results. P_1 is finding the first principal component of X, the direction that maximizes variance of scalar projections. P_2 says that this direction also minimizes the distances between the points to their vector projections along this direction. If we view the distances as errors in approximating the points by their projections along a line, then the error is minimized by choosing the line in the same direction as the first principal component. Finally P_3 tells us that finding a rank one matrix to best approximate the data matrix (in terms of error computed using Frobenius norm) is equivalent to finding the first principal component as well!

(a) Consider the line $\mathcal{L} = \{\vec{x}_0 \ + a\vec{u} : a \in \mathbb{R}\}$, with $\vec{x}_0 \in \mathbb{R}^d$, $\vec{u}^\top \vec{u} = 1$. Recall that the vector projection of a point $\vec{x} \in \mathbb{R}^d$ on to the line $\mathcal{L}$ is given by $\vec{z} = \vec{x}_0 + a^\star \vec{u}$, where $a^\star$ is given by:

[1]Data matrices are sometimes represented as above, and sometimes as the transpose of the matrix here. Make sure you always check this, and recall that based on the definition of the data matrix, the definition of the covariance matrix also changes.

$$a^\star = \underset{a}{\text{argmin}} \left\| \vec{x}_0 + a\vec{u} - \vec{x} \right\|_2. \tag{7}$$

Show that $a^\star = (\vec{x} - \vec{x}_0)^\top \vec{u}$. Use this to show that the square of the distance between x and its vector projection on $\mathcal{L}$ is given by:

$$\|\vec{x} - \vec{z}\|_2^2 = \|\vec{x} - \vec{x}_0\|_2^2 - ((\vec{x} - \vec{x}_0)^\top \vec{u})^2. \tag{8}$$

Solution: The projection of point $\vec{x}$ on $\mathcal{L}$ corresponds to the following problem:

$$a^\star = \min_a \ \left\| \vec{x}_0 + a\vec{u} - \vec{x} \right\|_2. \tag{9}$$

The squared objective writes

$$\left\| \vec{x}_0 + a\vec{u} - \vec{x} \right\|_2^2 = a^2 - 2a(\vec{x} - \vec{x}_0)^\top \vec{u} + \left\| \vec{x} - \vec{x}_0 \right\|_2^2. \tag{10}$$

By taking the derivative of the above expression with respect to a and setting it to 0, we obtain the optimal value of a as

$$a^\star = (\vec{x} - \vec{x}_0)^\top \vec{u}. \tag{11}$$

The square of the distance between $\vec{x}$ and its vector projection on $\mathcal{L}$ ($\vec{z}$) is given by $\|\vec{z} - \vec{x}\|_2^2$. We have shown that $\vec{z} = \vec{x}_0 + a^\star \vec{u} = \vec{x}_0 + [(\vec{x} - \vec{x}_0)^\top \vec{u}]\vec{u}$. At optimum, the squared objective function, which equals the minimum squared distance $\|\vec{z} - \vec{x}\|_2^2$, takes the desired value:

$$\left\| \vec{x}_0 + [(\vec{x} - \vec{x}_0)^\top \vec{u}]\vec{u} - \vec{x} \right\|_2^2 = \|\vec{x} - \vec{x}_0\|_2^2 - ((\vec{x} - \vec{x}_0)^\top \vec{u})^2. \tag{12}$$

(b) Show that P_2 is equivalent to P_1.

HINT: Start with P_2 and using the result from part (a) show that it is equivalent to P_1.

Solution: From part (a), we have the following decomposition of P_2:

$$\begin{aligned}
\underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\text{argmin}} \sum_{i=1}^n \min_{v_i \in \mathbb{R}} \|\vec{x}_i - v_i u\|_2^2 &= \underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\text{argmin}} \sum_{i=1}^n \|\vec{x}_i\|_2^2 - (\vec{x}_i^\top \vec{u})^2 &(13)\\
&= \underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\text{argmax}} \sum_{i=1}^n \vec{u}^\top \vec{x}_i \vec{x}_i^\top \vec{u} &(14)\\
&= \underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\text{argmax}} \ \vec{u}^\top C \vec{u}. &(15)
\end{aligned}$$

From the above equation, we see that a solution for P_1 constitutes a solution for P_2 and vice-versa.

(c) Show that every matrix $Y \in \mathbb{R}^{n \times d}$ with rank at most 1, can be expressed as $Y = \vec{v}\vec{u}^\top$ for some $\vec{v} \in \mathbb{R}^n$, $\vec{u} \in \mathbb{R}^d$ and $\|\vec{u}\|_2 = 1$.

Solution: First, consider the case where Y is rank-0. If Y is rank 0, all of its all of its singular values must be 0 and hence, Y must be the 0 matrix. Therefore, we can express $Y = \vec{v}\vec{u}^\top$ by setting $\vec{v} = 0$ and $\vec{u}$ being any arbitrary unit-length vector.

Now let Y be a rank 1 matrix. Then its has the following SVD: $Y = \sigma \vec{w} \vec{u}^\top$ where $\sigma \neq 0$. It follows that $Y = \vec{v}\vec{u}^\top$ for $\vec{v} = \sigma \vec{w}$.

(d) Show that P_3 is equivalent to P_2.

HINT: Use the result from part (c) to show that P_3 is equivalent to:

$$\underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1,\vec{v}\in\mathbb{R}^n}{\operatorname{argmin}} \left\|X - \vec{v}\vec{u}^\top\right\|_F^2 \tag{16}$$

Prove that this is equivalent to P_2.

Solution: From the previous part, we have that the set of matrices, Y, with rank at most 1 is equivalent to the set $\{\vec{v}\vec{u}^\top : \|\vec{u}\| = 1, \vec{u} \in \mathbb{R}^d, \vec{v} \in \mathbb{R}^n\}$. Therefore, we may equivalently reformulate P_3 as:

$$\underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1,\vec{v}\in\mathbb{R}^n}{\operatorname{argmin}} \left\|X - \vec{v}\vec{u}^\top\right\|_F^2. \tag{17}$$

X is a matrix with rows $\vec{x}_i^\top$, and $\vec{v}\vec{u}^\top$ is a matrix with rows $v_i\vec{u}^\top$. We expand the Frobenius norm in the objective in the above equation as

$$\left\|X - \vec{v}\vec{u}^\top\right\|_F^2 = \sum_{i=1}^n \|\vec{x}_i - v_i\vec{u}\|_2^2, \tag{18}$$

i.e., express the matrix norm as a sum of vector norms, which follows from the definition of the Frobenius norm.

With this reformulation, we see that any solution $(\vec{u}^\star, \vec{v}^\star)$ must satisfy

$$\vec{v}^\star = \underset{\vec{v}}{\operatorname{argmin}} \sum_{i=1}^n \|\vec{x}_i - v_i\vec{u}\|_2^2, \quad \vec{u}^\star = \underset{\vec{u}}{\operatorname{argmin}} \sum_{i=1}^n \|\vec{x}_i - v_i^\star\vec{u}\|_2^2 \tag{19}$$

i.e., we can minimize it over $\vec{u}, \vec{v}$ sequentially. We separate the minimization over $\vec{u}$ and $\vec{v}$ to get

$$\vec{u}^\star = \underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1}{\operatorname{argmin}} \min_{\vec{v}\in\mathbb{R}^n} \sum_{i=1}^n \|\vec{x}_i - v_i\vec{u}\|_2^2 \tag{20}$$

We now have a minimization of a sum of squares of vector norms $\|\vec{x}_i - v_i\vec{u}\|_2^2$, each of which depends only on a single element of $\vec{v}$, i.e., v_i.

Note: The objective of an optimization problem $\min_{x,y} f(x, y)$ is said to be *separable* when the objective can be written as a sum of two functions- one which depends on x, and one on y, i.e.,

$$\min_{x,y} f(x, y) = \min_{x,y}[g(x) + h(y)]. \tag{21}$$

If the objective is separable, we can solve the problem *separately* across the two variables, and

$$(x^\star, y^\star) = \underset{x,y}{\operatorname{argmin}} f(x, y) = (\underset{x}{\operatorname{argmin}}\, g(x), \underset{y}{\operatorname{argmin}}\, h(y)). \tag{22}$$

We can split the minimization problem in 20 over each individual v_i. We have

$$\vec{u}^\star = \underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1}{\operatorname{argmin}} \sum_{i=1}^n \min_{v_i\in\mathbb{R}} \|\vec{x}_i - v_i\vec{u}\|_2^2. \tag{23}$$

Therefore, $\vec{u}^\star$ is also a solution to P_2.

3. Operator Norms

For a matrix $A \in \mathbb{R}^{m\times n}$, the *induced norm* or *operator norm* $\|A\|_p$ is defined as

$$\|A\|_p \doteq \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p}. \tag{24}$$

In this problem, we provide a characterization of the induced norm for certain values of p. Let a_{ij} denote the (i,j)-th entry of A. Prove the following:

(a) $\|A\|_2 = \sigma_{\max}\{A\}$, the maximum singular value of A. *HINT: Consider connecting $\|A\|_2^2$ to a particular Rayleigh coefficient.*

Solution: Approach 1: Rayleigh Coefficient

We have

$$\begin{aligned}
\|A\|_2^2 &= \left(\max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2}\right)^2 && (25)\\
&= \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_2^2}{\|\vec{x}\|_2^2} && (26)\\
&= \max_{\vec{x}\neq\vec{0}} \frac{(A\vec{x})^\top(A\vec{x})}{\vec{x}^\top\vec{x}} && (27)\\
&= \max_{\vec{x}\neq\vec{0}} \frac{\vec{x}^\top A^\top A\vec{x}}{\vec{x}^\top\vec{x}} && (28)\\
&= \lambda_{\max}\{A^\top A\} && (29)\\
&= \sigma_{\max}\{A\}^2. && (30)
\end{aligned}$$

Taking square roots gives the solution.

Approach 2: SVD

Write $A = U\Sigma V^\top$. Then

$$\begin{aligned}
\|A\|_2 &= \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2} && (31)\\
&= \max_{\vec{x}\neq\vec{0}} \frac{\left\|U\Sigma V^\top\vec{x}\right\|_2}{\|\vec{x}\|_2}. && (32)
\end{aligned}$$

Here we use the fact that multiplying by the square orthonormal matrix U does not change the norm, so we have

$$\|A\|_2 = \max_{\vec{x}\neq\vec{0}} \frac{\left\|\Sigma V^\top\vec{x}\right\|_2}{\|\vec{x}\|_2}. \tag{33}$$

Now we do the change of basis $\vec{z} = V^\top\vec{x}$, i.e., $\vec{x} = V\vec{z}$. Thus, we have

$$\|A\|_2 = \max_{\substack{\vec{x}\neq\vec{0}\\ \vec{z}=V^\top\vec{x}}} \frac{\|\Sigma\vec{z}\|_2}{\|V\vec{z}\|_2}. \tag{34}$$

Since V is square orthonormal, multiplying by it does not change the norm, so we have

$$\|A\|_2 = \max_{\substack{\vec{x}\neq\vec{0}\\ \vec{z}=V^\top\vec{x}}} \frac{\|\Sigma\vec{z}\|_2}{\|\vec{z}\|_2}. \tag{35}$$

Finally, we make the crucial realization that since V is square orthonormal, it is invertible. Since we can always get $\vec{x}$ from $\vec{z}$ and $\vec{z}$ from $\vec{x}$, it is sufficient to optimize directly over $\vec{z}$. Thus,

$$\|A\|_2 = \max_{\vec{z}\neq\vec{0}} \frac{\|\Sigma\vec{z}\|_2}{\|\vec{z}\|_2}. \tag{36}$$

Finally, we evaluate this maximum, or rather its square (and take square roots afterwards). Suppose without loss of generality that $\sigma_1\{A\} \geq \sigma_2\{A\} \geq \cdots$. We have

$$\begin{aligned} \frac{\|\Sigma\vec{z}\|_2^2}{\|\vec{z}\|_2^2} &= \frac{\vec{z}^\top\Sigma^\top\Sigma\vec{z}}{\vec{z}^\top\vec{z}} &(37)\\ &= \frac{\sum_i \sigma_i\{A\}^2 z_i^2}{\sum_i z_i^2} &(38)\\ &= \sum_i \sigma_i\{A\}^2 \cdot \frac{z_i^2}{\sum_j z_j^2}. &(39)\end{aligned}$$

This is maximized when $z_1^2 = 1$ and all other entries are 0, so $\vec{z} = \pm\vec{e}_1$. In this case we have that

$$\frac{\|\Sigma\vec{z}\|_2^2}{\|\vec{z}\|_2^2} = \sigma_1\{A\}^2 = \sigma_{\max}\{A\}^2. \tag{40}$$

Thus we have

$$\|A\|_2 = \max_{\vec{z}\neq\vec{0}} \frac{\|\Sigma\vec{z}\|_2}{\|\vec{z}\|_2} = \sigma_{\max}\{A\} \tag{41}$$

as claimed.

(b) $\|A\|_1$ is the maximum absolute column sum of A,

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|\,. \tag{42}$$

HINT: Write $A\vec{x}$ as a linear combination of the columns of A to obtain $\|A\vec{x}\|_1 = \|\sum_{i=1}^n x_i \cdot \vec{a}_i\|_1$, where $\vec{a}_i$ denotes the i-th column of A. Then apply triangle inequality to terms within the sum.

Solution: We denote the columns of A to be $\begin{bmatrix}\vec{a}_1 & \cdots & \vec{a}_n\end{bmatrix}$. Then, for any $x \in \mathbb{R}^n$, we have that

$$\begin{aligned} \|A\vec{x}\|_1 = \left\|\sum_{i=1}^n x_i \cdot \vec{a}_i\right\|_1 &\leq \sum_{i=1}^n \|x_i \cdot \vec{a}_i\|_1 && \text{(Triangle Inequality)} &(43)\\ &= \sum_{i=1}^n |x_i| \cdot \|\vec{a}_i\|_1 && (|x_i| \text{ can come out of norm} &(44)\\ &\leq \left(\sum_{i=1}^n |x_i|\right) \cdot \max_i \|\vec{a}_i\|_1 && (\|\vec{a}_j\|_1 \leq \max_i \|\vec{a}_i\|_1\,,\ \forall j) &(45)\\ &= \|\vec{x}\|_1 \cdot \max_i \|\vec{a}_i\|_1. && &(46)\end{aligned}$$

We see that $\|A\|_1 = \max_{\vec{x}\neq\vec{0}} \dfrac{\|A\vec{x}\|_1}{\|\vec{x}\|_1} \leq \max_i \|\vec{a}_i\|_1$.

Next we show that we can achieve this upper bound by choosing $\vec{x} = \vec{e_i}$ where i is the index for the column of A such that $\vec{a}_i$ has the maximum column sum, and this vector gives $\dfrac{\|A\vec{x}\|_1}{\|\vec{x}\|_1} = \max_i \|\vec{a}_i\|_1$. Therefore, we obtain $\|A\|_1$ as the maximum absolute column sum of A.

(c) **(OPTIONAL)** $\|A\|_\infty$ is the maximum absolute row sum of A,

$$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}| \,. \tag{47}$$

HINT: First write $\|A\vec{x}\|_\infty = \max_i \left|\sum_{j=1}^{n} a_{ij}x_j\right|$*. Then apply triangle inequality and use the fact that* $|x_j| \le \max_i |x_i|, \ \forall j$*.*

Solution: We have,

$$\begin{aligned}
\|A\vec{x}\|_\infty &= \max_i \left|\sum_{j=1}^{n} a_{ij}x_j\right| && (48)\\
&\le \max_i \sum_{j=1}^{n} |a_{ij}x_j| && \text{(Triangle Inequality)} \quad (49)\\
&\le \max_i \left[\max_j |x_j| \Big(\sum_{j=1}^{n} |a_{ij}|\Big)\right] && (|x_j| \le \max_j |x_j|, \ \forall j) \quad (50)\\
&= \left(\max_i \sum_{j=1}^{n} |a_{ij}|\right) \|\vec{x}\|_\infty \,. && (51)
\end{aligned}$$

Thus, we have

$$\|A\|_\infty = \max_{\vec{x}\ne\vec{0}} \frac{\|A\vec{x}\|_\infty}{\|\vec{x}\|_\infty} \le \max_i \sum_{j=1}^{n} |a_{ij}|. \tag{52}$$

Assume that the index of the maximum absolute row sum is m. We can construct a vector $\vec{x}$ such that $x_j = 1$ if $a_{mj} \ge 0$, and $x_j = -1$ if $a_{mj} < 0$. This leads to

$$\|A\vec{x}\|_\infty = \max_i \left|\sum_{j=1}^{n} a_{ij}x_j\right| \ge \left|\sum_{j=1}^{n} a_{mj}x_j\right| = \sum_{j=1}^{n} |a_{mj}| \,. \tag{53}$$

Since $\|\vec{x}\|_\infty = 1$, the resulting $\frac{\|A\vec{x}\|_\infty}{\|\vec{x}\|_\infty} = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$ as desired.

This shows that $\|A\|_\infty =$ maximum absolute row sum.

4. Gradients, Jacobians, and Hessians

The *gradient* of a scalar-valued function $g\colon \mathbb{R}^n \to \mathbb{R}$ is the column vector of length n, denoted as ∇g, containing the derivatives of components of g with respect to the input variables:

$$(\nabla g(\vec{x}))_i = \frac{\partial g}{\partial x_i}(\vec{x}),\ i = 1, \dots n. \tag{54}$$

The *Hessian* of a scalar-valued function $g\colon \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ matrix, denoted as $\nabla^2 g$, containing the second derivatives of components of g with respect to the input variables:

$$(\nabla^2 g(\vec{x}))_{ij} = \frac{\partial^2 g}{\partial x_i \partial x_j}(\vec{x}),\ \ i = 1, \dots, n,\ \ j = 1, \dots, n. \tag{55}$$

The *Jacobian* of a vector-valued function $\vec{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ is the $m \times n$ matrix, denoted as $D\vec{g}$, containing the derivatives of components of $\vec{g}$ with respect to the input variables:

$$(D\vec{g}(\vec{x}))_{ij} = \frac{\partial g_i}{\partial x_j}(\vec{x}),\ \ i = 1, \dots, m,\ \ j = 1, \dots, n. \tag{56}$$

For the remainder of the class, we will repeatedly have to take gradients, Hessians and Jacobians of functions we are trying to optimize. This exercise serves as a warm up for future problems.

For the first two parts, suppose $A \in \mathbb{R}^{n\times n}$ is a square matrix whose entries are denoted a_{ij} and whose rows are denoted $\vec{a}_1^\top, \dots, \vec{a}_n^\top$, and $\vec{b} \in \mathbb{R}^n$ is a vector whose entries are denoted b_i.

(a) Compute the Jacobians for the following functions.

i. $\vec{g}(\vec{x}) = A\vec{x}$.

Solution: We compute each partial derivative and reconstitute $D\vec{g}$ at the end. That is,

$$\begin{aligned}
[D\vec{g}(\vec{x})]_{jk} &= \frac{\partial g_j}{\partial x_k}(\vec{x}) && (57)\\
&= \frac{\partial (A\vec{x})_j}{\partial x_k} && (58)\\
&= \frac{\partial (\vec{a}_j^\top \vec{x})}{\partial x_k} && (59)\\
&= (\vec{a}_j)_k && (60)\\
&= a_{jk} && (61)\\
&= [A]_{jk}, && (62)
\end{aligned}$$

where $[A]_{jk}$ is the (j, k)th entry of A. Thus $D\vec{g}(\vec{x}) = A$.

ii. $\vec{g}(\vec{x}) = f(\vec{x})\vec{x}$ where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable.

Solution: We again compute each partial derivative using the scalar product rule, obtaining

$$\begin{aligned}
[D\vec{g}(\vec{x})]_{jk} &= \frac{\partial g_j}{\partial x_k}(\vec{x}) && (63)\\
&= \frac{\partial (f(\vec{x})x_j)}{\partial x_k} && (64)\\
&= f(\vec{x})\frac{\partial x_j}{\partial x_k} + x_j \frac{\partial f}{\partial x_k}(\vec{x}) && (65)
\end{aligned}$$

$$= x_j[\nabla f(\vec{x})]_k + \begin{cases} f(\vec{x}), & \text{if } j = k \\ 0, & \text{if } j \neq k. \end{cases} \tag{66}$$

This gives a Jacobian whose entries are

$$[D\vec{g}(\vec{x})]_{jk} = x_j[\nabla f(\vec{x})]_k, \qquad \forall j \neq k \tag{67}$$

$$[D\vec{g}(\vec{x})]_{jj} = x_j[\nabla f(\vec{x})]_k + f(\vec{x}), \qquad \forall j. \tag{68}$$

For the purpose of self-grades it is fine to stop here. But one can also write the Jacobian as

$$D\vec{g}(\vec{x}) = \vec{x}[\nabla f(\vec{x})]^\top + f(\vec{x})I. \tag{69}$$

iii. $\vec{g}(\vec{x}) = f(A\vec{x} + \vec{b})\vec{x}$ where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is differentiable.

Solution: We compute each partial derivative using the scalar product rule and chain rule, obtaining

$$[D\vec{g}(\vec{x})]_{jk} = \frac{\partial g_j}{\partial x_k}(\vec{x}) \tag{70}$$

$$= \frac{\partial f(A\vec{x} + \vec{b})x_j}{\partial x_k} \tag{71}$$

$$= f(A\vec{x} + \vec{b})\frac{\partial x_j}{\partial x_k} + x_j\frac{\partial f(A\vec{x} + \vec{b})}{\partial x_k} \tag{72}$$

$$= f(A\vec{x} + \vec{b})\frac{\partial x_j}{\partial x_k} + x_j\sum_{\ell=1}^{n}\frac{\partial f(A\vec{x} + \vec{b})}{\partial (A\vec{x} + \vec{b})_\ell} \cdot \frac{\partial (A\vec{x} + \vec{b})_\ell}{\partial x_k} \tag{73}$$

$$= f(A\vec{x} + \vec{b})\frac{\partial x_j}{\partial x_k} + x_j\sum_{\ell=1}^{n}[\nabla f(A\vec{x} + \vec{b})]_\ell \cdot \frac{\partial (\vec{a}_\ell^\top \vec{x} + b_\ell)}{\partial x_k} \tag{74}$$

$$= f(A\vec{x} + \vec{b})\frac{\partial x_j}{\partial x_k} + x_j\sum_{\ell=1}^{n}[\nabla f(A\vec{x} + \vec{b})]_\ell \cdot (\vec{a}_\ell)_k \tag{75}$$

$$= f(A\vec{x} + \vec{b})\frac{\partial x_j}{\partial x_k} + x_j\sum_{\ell=1}^{n}[\nabla f(A\vec{x} + \vec{b})]_\ell \cdot a_{\ell k} \tag{76}$$

$$= f(A\vec{x} + \vec{b})\frac{\partial x_j}{\partial x_k} + x_j\sum_{\ell=1}^{n}[\nabla f(A\vec{x} + \vec{b})]_\ell \cdot [A^\top]_{k\ell} \tag{77}$$

$$= f(A\vec{x} + \vec{b})\frac{\partial x_j}{\partial x_k} + x_j[A^\top \nabla f(A\vec{x} + \vec{b})]_k \tag{78}$$

$$= x_j[A^\top \nabla f(A\vec{x} + \vec{b})]_k + \begin{cases} f(A\vec{x} + \vec{b}), & \text{if } j = k \\ 0 & \text{if } j \neq k. \end{cases} \tag{79}$$

This gives a Jacobian whose entries are

$$[D\vec{g}(\vec{x})]_{jk} = x_j[A^\top \nabla f(A\vec{x} + \vec{b})]_k, \qquad \forall j \neq k \tag{80}$$

$$[D\vec{g}(\vec{x})]_{jj} = x_j[A^\top \nabla f(A\vec{x} + \vec{b})]_k + f(A\vec{x} + \vec{b}), \qquad \forall j. \tag{81}$$

For the purpose of self-grades it is fine to stop here. But one can also write the Jacobian as

$$D\vec{g}(\vec{x}) = \vec{x}[A^\top \nabla f(A\vec{x} + \vec{b})]^\top + f(A\vec{x} + \vec{b})I \tag{82}$$

$$= \vec{x}[\nabla f(A\vec{x} + \vec{b})]^\top A + f(A\vec{x} + \vec{b})I. \tag{83}$$

(b) Compute the gradients and Hessians for the following functions.

i. $g_1(\vec{x}) = \vec{x}^\top A\vec{x}$.

Solution: We have

$$\begin{aligned} g_1(\vec{x}) &= \vec{x}^\top A\vec{x} && (84)\\ &= \vec{x}^\top \begin{bmatrix} \vec{a}_1^\top \\ \vdots \\ \vec{a}_n^\top \end{bmatrix} \vec{x} && (85)\\ &= \vec{x}^\top \begin{bmatrix} \vec{a}_1^\top \vec{x} \\ \vdots \\ \vec{a}_n^\top \vec{x} \end{bmatrix} && (86)\\ &= \sum_{i=1}^{n} x_i(\vec{a}_i^\top \vec{x}). && (87) \end{aligned}$$

Taking the derivative with respect to any x_j yields

$$\begin{aligned} \frac{\partial g_1}{\partial x_j}(\vec{x}) &= \frac{\partial}{\partial x_j} \sum_{i=1}^{n} x_i(\vec{a}_i^\top \vec{x}) && (88)\\ &= \frac{\partial}{\partial x_j}\left[x_j(\vec{a}_j^\top \vec{x}) + \sum_{\substack{i=1\\ i\neq j}}^{n} x_i(\vec{a}_i^\top \vec{x}) \right] && (89)\\ &= \frac{\partial}{\partial x_j} x_j(\vec{a}_j^\top \vec{x}) + \sum_{\substack{i=1\\ i\neq j}}^{n} \frac{\partial}{\partial x_j} x_i(\vec{a}_i^\top \vec{x}) && (90)\\ &= x_j \frac{\partial}{\partial x_j}(\vec{a}_j^\top \vec{x}) + \vec{a}_j^\top \vec{x} + \sum_{\substack{i=1\\ i\neq j}}^{n} x_i \frac{\partial}{\partial x_j}(\vec{a}_i^\top \vec{x}) && (91)\\ &= x_j(\vec{a}_j)_j + \vec{a}_j^\top \vec{x} + \sum_{\substack{i=1\\ i\neq j}}^{n} x_i(\vec{a}_i)_j && (92)\\ &= \vec{a}_j^\top \vec{x} + \sum_{i=1}^{n} x_i(\vec{a}_i)_j && (93)\\ &= \sum_{i=1}^{n} a_{ji}x_i + \sum_{i=1}^{n} a_{ij}x_i && (94)\\ &= (A\vec{x})_j + (A^\top \vec{x})_j && (95)\\ &= [(A + A^\top)\vec{x}]_j. && (96) \end{aligned}$$

Thus we have

$$\nabla g_1(\vec{x}) = (A + A^\top)\vec{x}. \quad (97)$$

For the Hessian, notice that

$$\begin{aligned} [\nabla^2 g_1(\vec{x})]_{jk} &= \frac{\partial^2 g_1}{\partial x_j \,\partial x_k}(\vec{x}) && (98)\\ &= \frac{\partial [\nabla g_1]_j}{\partial x_k}(\vec{x}) && (99) \end{aligned}$$

$$= \frac{\partial}{\partial x_k}\left[\sum_{i=1}^{n} a_{ji}x_i + \sum_{i=1}^{n} a_{ij}x_i\right] \tag{100}$$

$$= \left[\sum_{i=1}^{n} \frac{\partial}{\partial x_k} a_{ji}x_i + \sum_{i=1}^{n} \frac{\partial}{\partial x_k} a_{ij}x_i\right] \tag{101}$$

$$= a_{jk} + a_{kj} \tag{102}$$

$$= (A + A^\top)_{jk}. \tag{103}$$

Thus we have

$$\nabla^2 g_1(\vec{x}) = A + A^\top. \tag{104}$$

Notice the important special case that if A is symmetric then $\nabla g_1(\vec{x}) = 2A\vec{x}$ and $\nabla^2 g_1(\vec{x}) = 2A$.

ii. $g_2(\vec{x}) = \|\vec{x}\|_2^2$.

Solution: We take $A = I$ in g_1, obtaining

$$\nabla g_2(\vec{x}) = 2\vec{x} \qquad \text{and} \qquad \nabla^2 g_2(\vec{x}) = 2I. \tag{105}$$

iii. $g_3(\vec{x}) = g_2(A\vec{x} - \vec{b}) = \|A\vec{x} - \vec{b}\|_2^2$. (Use the chain rule and the Jacobians computed in part (a).)

Solution: As required, we use the chain rule: we compute

$$\nabla g_3(\vec{x}) = [Dg_3(\vec{x})]^\top \tag{106}$$

$$= ([Dg_2(A\vec{x} - \vec{b})][D(A\vec{x} - \vec{b})])^\top \tag{107}$$

$$= ([\nabla g_2(A\vec{x} - \vec{b})]^\top [D(A\vec{x} - \vec{b})])^\top \tag{108}$$

$$= [D(A\vec{x} - \vec{b})]^\top [\nabla g_2(A\vec{x} - \vec{b})] \tag{109}$$

$$= [A]^\top [2(A\vec{x} - \vec{b})] \tag{110}$$

$$= 2A^\top (A\vec{x} - \vec{b}). \tag{111}$$

For the Hessian, note that

$$\nabla^2 g_3(\vec{x}) = D[\nabla g_3(\vec{x})] \tag{112}$$

$$= D[2A^\top (A\vec{x} - \vec{b})] \tag{113}$$

$$= 2D[A^\top A\vec{x} - A^\top \vec{b}] \tag{114}$$

$$= 2A^\top A. \tag{115}$$

iv. $g_4(\vec{x}) = \log(\sum_{i=1}^{n} e^{x_i})$.

Solution: We use component-wise derivatives and the scalar-valued chain rule:

$$\frac{\partial g_4}{\partial x_j}(\vec{x}) = \frac{\partial}{\partial x_j} \log\left(\sum_{i=1}^{n} e^{x_i}\right) \tag{116}$$

$$= \frac{\frac{\partial}{\partial x_j} \sum_{i=1}^{n} e^{x_i}}{\sum_{i=1}^{n} e^{x_i}} \tag{117}$$

$$= \frac{\sum_{i=1}^{n} \frac{\partial}{\partial x_j} e^{x_i}}{\sum_{i=1}^{n} e^{x_i}} \tag{118}$$

$$= \frac{e^{x_j}}{\sum_{i=1}^{n} e^{x_i}}. \tag{119}$$

Thus the gradient is of the form

$$\nabla g_4(\vec{x}) = \begin{bmatrix} \partial g_4/\partial x_1\ (\vec{x}) \\ \vdots \\ \partial g_4/\partial x_n\ (\vec{x}) \end{bmatrix} = \frac{1}{\sum_{i=1}^{n} e^{x_i}} \begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_n} \end{bmatrix}. \tag{120}$$

The Hessian is computed using more component-wise derivatives and the scalar quotient rule:

$$\begin{aligned} \frac{\partial^2 g_4}{\partial x_j\, \partial x_k}(\vec{x}) &= \frac{\partial}{\partial x_k}\left(\frac{\partial g_4}{\partial x_j}\right)(\vec{x}) && (121)\\ &= \frac{\partial}{\partial x_k} \frac{e^{x_j}}{\sum_{i=1}^{n} e^{x_i}} && (122)\\ &= \frac{\left(\sum_{i=1}^{n} e^{x_i}\right)\frac{\partial}{\partial x_k} e^{x_j} - e^{x_j}\frac{\partial}{\partial x_k}\left(\sum_{i=1}^{n} e^{x_i}\right)}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2} && (123)\\ &= \frac{\left(\sum_{i=1}^{n} e^{x_i}\right)\frac{\partial}{\partial x_k} e^{x_j} - e^{x_j} e^{x_k}}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2} && (124)\\ &= -\frac{e^{x_j + x_k}}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2} + \begin{cases} \frac{e^{x_k}}{\sum_{i=1}^{n} e^{x_i}}, & \text{if } j = k \\ 0, & \text{if } j \neq k. \end{cases} && (125) \end{aligned}$$

This is the (i, j)th coordinate of the Hessian, whose entries can be defined as

$$[\nabla^2 g_4(\vec{x})]_{jk} = -\frac{e^{x_j + x_k}}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2} \qquad \forall j \neq k \tag{126}$$

$$\text{and} \qquad [\nabla^2 g_4(\vec{x})]_{jj} = \frac{e^{x_k}}{\sum_{i=1}^{n} e^{x_i}} - \frac{e^{x_j + x_k}}{\left(\sum_{i=1}^{n} e^{x_i}\right)^2} \qquad \forall j. \tag{127}$$

For the purpose of self-grades, the above solution is fine, but we we can also write the Hessian as

$$\nabla^2 g_4(\vec{x}) = \operatorname{diag}(\nabla g_4(\vec{x})) - [\nabla g_4(\vec{x})][\nabla g_4(\vec{x})]^\top \tag{128}$$

where $\operatorname{diag}(\cdot)\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$ forms a diagonal matrix whose diagonal entries are the entries of the input vector, i.e.,

$$\operatorname{diag}(\vec{v}) = \begin{bmatrix} v_1 & & \\ & \ddots & \\ & & v_n \end{bmatrix}. \tag{129}$$

v. $g_5(\vec{x}) = g_4(A\vec{x} - \vec{b}) = \log(\sum_{i=1}^{n} e^{\vec{a}_i^\top \vec{x} - b_i})$. (Use the chain rule and the Jacobians computed in part (a); you can use the gradient ∇g_4 and Hessian $\nabla^2 g_4$ in your answer without having to rewrite it.)

Solution: As prompted, we use the chain rule to compute the gradient:

$$\begin{aligned} \nabla g_5(\vec{x}) &= (Dg_5(\vec{x}))^\top && (130)\\ &= ([Dg_4(A\vec{x} - \vec{b})][D(A\vec{x} - \vec{b})])^\top && (131)\\ &= ([Dg_4(A\vec{x} - \vec{b})]A)^\top && (132)\\ &= A^\top [Dg_4(A\vec{x} - \vec{b})]^\top && (133) \end{aligned}$$

$$= A^\top [\nabla g_4(A\vec{x} - \vec{b})]. \tag{134}$$

To compute the Hessian, we write

$$\nabla^2 g_5(\vec{x}) = D(\nabla g_5)(\vec{x}) \tag{135}$$
$$= D(A^\top [\nabla g_4(A\vec{x} - \vec{b})]). \tag{136}$$

To evaluate this product, we look component-by-component and use a multivariate chain rule:

$$[D(A^\top [\nabla g_4(A\vec{x} - \vec{b})])]_{jk} = \frac{\partial}{\partial x_k}(A^\top [\nabla g_4(A\vec{x} - \vec{b})])_j \tag{137}$$
$$= \frac{\partial}{\partial x_k} \sum_{i=1}^n a_{ij} [\nabla g_4(A\vec{x} - \vec{b})]_i \tag{138}$$
$$= \sum_{i=1}^n a_{ij} \frac{\partial}{\partial x_k} [\nabla g_4(A\vec{x} - \vec{b})]_i \tag{139}$$
$$= \sum_{i=1}^n a_{ij} \sum_{\ell=1}^n \frac{\partial [\nabla g_4(A\vec{x} - \vec{b})]_i}{\partial (A\vec{x} - \vec{b})_\ell} \frac{\partial (A\vec{x} - \vec{b})_\ell}{\partial x_k} \tag{140}$$
$$= \sum_{i=1}^n a_{ij} \sum_{\ell=1}^n \frac{\partial^2 g_4(A\vec{x} - \vec{b})}{\partial (A\vec{x} - \vec{b})_i \, \partial (A\vec{x} - \vec{b})_\ell} \frac{\partial (A\vec{x} - \vec{b})_\ell}{\partial x_k} \tag{141}$$
$$= \sum_{i=1}^n a_{ij} \sum_{\ell=1}^n [\nabla^2 g_4(A\vec{x} - \vec{b})]_{i\ell} [D(A\vec{x} - \vec{b})]_{\ell k} \tag{142}$$
$$= \sum_{i=1}^n a_{ij} \sum_{\ell=1}^n [\nabla^2 g_4(A\vec{x} - \vec{b})]_{i\ell} [A]_{\ell k} \tag{143}$$
$$= \sum_{i=1}^n a_{ij} ([\nabla^2 g_4(A\vec{x} - \vec{b})]A)_{ik} \tag{144}$$
$$= \sum_{i=1}^n [A^\top]_{ji} ([\nabla^2 g_4(A\vec{x} - \vec{b})]A)_{ik} \tag{145}$$
$$= (A^\top [\nabla^2 g_4(A\vec{x} - \vec{b})]A)_{jk}. \tag{146}$$

Thus,

$$\nabla^2 g_5(\vec{x}) = A^\top [\nabla^2 g_4(A\vec{x} - \vec{b})]A. \tag{147}$$

vi. $g_6(\vec{x}) = e^{\|\vec{x}\|_2^2} = e^{g_2(\vec{x})}$. (Use the chain rule and the Jacobians computed in part (a).)

Solution: We use the multivariable chain rule, obtaining

$$\nabla g_6(\vec{x}) = (Dg_6(\vec{x}))^\top \tag{148}$$
$$= ([D\exp(g_2(\vec{x}))][Dg_2(\vec{x})])^\top \tag{149}$$
$$= [Dg_2(\vec{x})]^\top [D\exp(g_2(\vec{x}))]^\top \tag{150}$$
$$= [\nabla g_2(\vec{x})][e^{g_2(\vec{x})}]^\top \tag{151}$$
$$= [2\vec{x}][e^{g_2(\vec{x})}] \tag{152}$$
$$= 2e^{\|\vec{x}\|_2^2}\vec{x}, \tag{153}$$

since the transpose of a scalar is a scalar. For the Hessian, we obtain

$$\nabla^2 g_6(\vec{x}) = D(\nabla g_6)(\vec{x}) \tag{154}$$

$$= D(2e^{\|\vec{x}\|_2^2}\vec{x}). \quad (155)$$

Notice that the function being differentiated is of the form $f(\vec{x})\vec{x}$, for $f(\vec{x}) = 2e^{\|\vec{x}\|_2^2}$. Thus we obtain that

$$\nabla^2 g_6(\vec{x}) = D(f(\vec{x})\vec{x}) \quad (156)$$
$$= \vec{x}[\nabla f(\vec{x})]^\top + f(\vec{x})I. \quad (157)$$

It just remains to compute $\nabla f(\vec{x})$, which we compute componentwise using the scalar chain rule, and obtain

$$[\nabla f(\vec{x})]_j = \frac{\partial f}{\partial x_j}(\vec{x}) \quad (158)$$
$$= \frac{\partial}{\partial x_j}(2e^{\|\vec{x}\|_2^2}) \quad (159)$$
$$= 2e^{\|\vec{x}\|_2^2}\frac{\partial \|\vec{x}\|_2^2}{\partial x_j} \quad (160)$$
$$= 2e^{\|\vec{x}\|_2^2}[\nabla \|\vec{x}\|_2^2]_j \quad (161)$$
$$= 2e^{\|\vec{x}\|_2^2}[2\vec{x}]_j \quad (162)$$
$$= 4x_j e^{\|\vec{x}\|_2^2}. \quad (163)$$

Thus we have

$$\nabla f(\vec{x}) = 4e^{\|\vec{x}\|_2^2}\vec{x}. \quad (164)$$

This gives

$$\nabla^2 g_6(\vec{x}) = 4e^{\|\vec{x}\|_2^2}\vec{x}\vec{x}^\top + 2e^{\|\vec{x}\|_2^2}I \quad (165)$$
$$= 2e^{\|\vec{x}\|_2^2}(I + 2\vec{x}\vec{x}^\top). \quad (166)$$

vii. $g_7(\vec{x}) = e^{\|A\vec{x}-\vec{b}\|_2^2} = g_6(A\vec{x} - \vec{b})$. (Use the chain rule and the Jacobians computed in part (a); you can use the gradient ∇g_6 and Hessian $\nabla^2 g_6$ in your answer without having to rewrite it.)

Solution: Notice that in part **4.**((b))v, we did not use any specific functional properties of g_4, ∇g_4, or $\nabla^2 g_5$ to determine ∇g_5 and $\nabla^2 g_5$. Using the exact same analysis, we can conclude that

$$\nabla g_7 = A^\top[\nabla g_6(A\vec{x} - \vec{b})], \quad \text{and} \quad \nabla^2 g_7 = A^\top[\nabla^2 g_6(A\vec{x} - \vec{b})]A. \quad (167)$$

Consider the case now where all vectors and matrices above are scalar; do your answers above make sense? (No need to answer this in your submission.)

(c) Plot/hand-draw the level sets of the following functions:

i. $g(x_1, x_2) = \frac{x_1^2}{4} + \frac{x_2^2}{9}$

ii. $g(x_1, x_2) = x_1 x_2$

Also point out the gradient directions in the level-set diagram. Additionally, compute the first and second order Taylor series approximation around the point $(1, 1)$ for each function and comment on how accurately they approximate the true function.

Solution: Figures 1 and 2 contain the level sets and gradient directions for the given functions.

i. We first compute the first and second order partial derivatives of g as follows:

$$\frac{\partial g}{\partial x_1}(x_1, x_2) = \frac{x_1}{2}, \quad \frac{\partial g}{\partial x_2}(x_1, x_2) = \frac{2x_2}{9}, \tag{168}$$

$$\frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) = \frac{1}{2}, \quad \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) = 0, \tag{169}$$

$$\frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) = \frac{2}{9}, \quad \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) = 0. \tag{170}$$

The gradient of g is then given by,

$$\nabla g(x_1, x_2) = \begin{bmatrix} \frac{\partial g}{\partial x_1}(1,1) \\ \frac{\partial g}{\partial x_2}(1,1) \end{bmatrix}, \tag{171}$$

and the Hessian matrix is given by,

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) \\ \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) & \frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) \end{bmatrix}. \tag{172}$$

The first order Taylor series approximation around $(1, 1)$ can be computed as:

$$g(x_1, x_2) \approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{173}$$

$$= \frac{13}{36} + \frac{x_1}{2} - \frac{1}{2} + \frac{2x_2}{9} - \frac{2}{9} = \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36}. \tag{174}$$

The second order Taylor series approximation around $(1, 1)$ can be computed as:

$$g(x_1, x_2) \approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 - 1 & x_2 - 1 \end{bmatrix} H(1,1) \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{175}$$

$$= \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36} + \frac{1}{2}\left(\frac{1}{2}(x_1 - 1)^2 + \frac{2}{9}(x_2 - 1)^2\right) \tag{176}$$

$$= \frac{(x_1 - 1)^2}{4} + \frac{(x_2 - 1)^2}{9} + \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36}. \tag{177}$$

$$= \frac{x_1^2}{4} + \frac{x_2^2}{9} \tag{178}$$

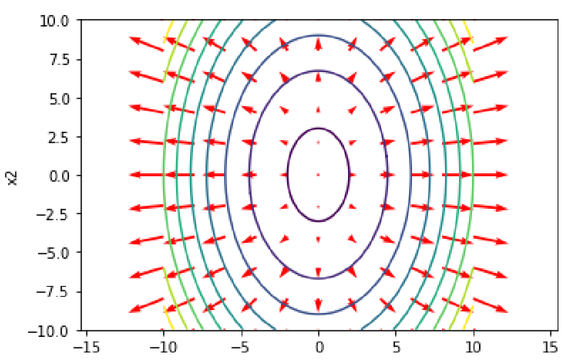

Figure 1: Level sets and gradient directions for the function $g(x_1, x_2) = \frac{x_1^2}{4} + \frac{x_2^2}{9}$.

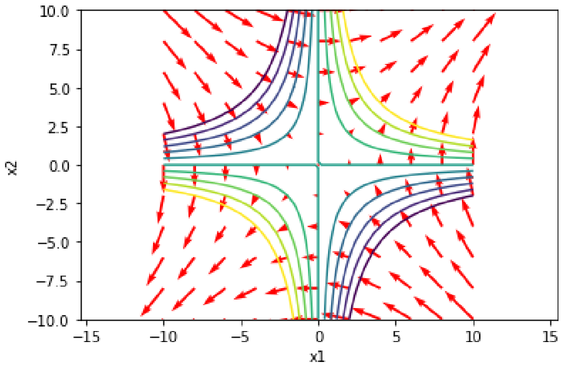

Figure 2: Level sets and gradient directions for the function $g(x_1, x_2) = x_1 x_2$.

The original function at (1.1,1.1) takes on the value 0.437. The first order approximation returns, evaluated at (1.1,1.1): $\frac{1.1}{2} + \frac{2.2}{9} - \frac{13}{36} = 0.433$. Additionally, observe that the second order approximation simplifies to return the original function!

ii. We follow the same steps as in the previous part of the problem. The partial derivatives for this g are given by:

$$\frac{\partial g}{\partial x_1}(x_1, x_2) = x_2, \quad \frac{\partial g}{\partial x_2}(x_1, x_2) = x_1, \tag{179}$$

$$\frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) = 0, \quad \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) = 1, \tag{180}$$

$$\frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) = 0, \quad \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) = 1. \tag{181}$$

The first order Taylor series approximation around $(1, 1)$ can be computed as:

$$g(x_1, x_2) \approx g(1, 1) + (\nabla g(1, 1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{182}$$

$$= 1 + x_1 - 1 + x_2 - 1 = x_1 + x_2 - 1. \tag{183}$$

The second order Taylor series approximation around $(1, 1)$ can be computed as:

$$g(x_1, x_2) \approx g(1, 1) + (\nabla g(1, 1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 - 1 & x_2 - 1 \end{bmatrix} H(1, 1) \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{184}$$

$$= x_1 + x_2 - 1 + \frac{1}{2}(2(x_1 - 1)(x_2 - 1)) \tag{185}$$

$$= (x_1 - 1)(x_2 - 1) + x_1 + x_2 - 1. \tag{186}$$

$$= x_1 x_2 \tag{187}$$

The original function evaluated at (1.1,1.1) is 1.21. The first order approximation around $(1.1, 1.1)$ is 1.2, but the second order approximation again exactly represents the function!

5. Homework Process

With whom did you work on this homework? List the names and SIDs of your group members.

NOTE: If you didn't work with anyone, you can put "none" as your answer.