Homework 4

Self grades are due at 11 PM one week after the homework.

1. **Mid-Semester Survey**

   Please complete this mid-semester survey at the following link: link. You will get a code at the end of the survey; write it in as the solution for this problem.

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:

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

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 distances from the points to their vector projections. Formally P_2 solves the minimization problem:

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

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

$$v^\star = \underset{v \in \mathbb{R}}{\operatorname{argmin}} \|\vec{x} - v\vec{u}\|_2^2 , \tag{3}$$

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:

$$\underset{Y : \operatorname{rank}(Y) \leq 1}{\operatorname{argmin}} \|X - Y\|_F . \tag{4}$$

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.

$$a^\star = \underset{a}{\operatorname{argmin}} \left\| \vec{x}_0 + a\vec{u} - \vec{x} \right\|_2. \tag{5}$$

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:

$$\left\| \vec{x} - \vec{z} \right\|_2^2 = \left\| \vec{x} - \vec{x}_0 \right\|_2^2 - ((\vec{x} - \vec{x}_0)^\top \vec{u})^2. \tag{6}$$

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

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

The squared objective writes

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

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

$$a^\star = (\vec{x} - \vec{x}_0)^\top \vec{u}. \tag{9}$$

The square of the distance between $\vec{x}$ and its vector projection on $\mathcal{L}$ ($\vec{z}$) is given by $\left\| \vec{z} - \vec{x} \right\|_2^2$. 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 $\left\| \vec{z} - \vec{x} \right\|_2^2$, takes the desired value:

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

(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:

$$\begin{aligned} \underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmin}} \sum_{i=1}^{n} \min_{v_i \in \mathbb{R}} \left\| \vec{x}_i - v_i u \right\|^2 &= \underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \vec{x}_i \right\|^2 - (\vec{x}_i^\top \vec{u})^2 && (11) \\ &= \underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmax}} \sum_{i=1}^{n} \vec{u}^\top \vec{x}_i \vec{x}_i^\top \vec{u} && (12) \\ &= \underset{\vec{u} \in \mathbb{R}^d : \vec{u}^\top \vec{u} = 1}{\operatorname{argmax}} \; \vec{u}^\top C \vec{u}. && (13) \end{aligned}$$

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 $\left\| \vec{u} \right\|_2 = 1$.

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:

$$\underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1,\vec{v}\in\mathbb{R}^n}{\operatorname{argmin}} \left\|X - \vec{v}\vec{u}^\top\right\|_F^2 \tag{14}$$

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:

$$\underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1,\vec{v}\in\mathbb{R}^n}{\operatorname{argmin}} \left\|X - \vec{v}\vec{u}^\top\right\|_F^2. \tag{15}$$

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

$$\left\|X - \vec{v}\vec{u}^\top\right\|_F^2 = \sum_{i=1}^n \|\vec{x}_i - v_i\vec{u}\|^2, \tag{16}$$

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

$$\vec{v}^\star = \underset{\vec{v}}{\operatorname{argmin}} \sum_{i=1}^n \|\vec{x}_i - v_i\vec{u}\|^2, \quad \vec{u}^\star = \underset{\vec{u}}{\operatorname{argmin}} \sum_{i=1}^n \|\vec{x}_i - v_i^\star\vec{u}\|^2 \tag{17}$$

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

$$\vec{u}^\star = \underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1}{\operatorname{argmin}} \min_{\vec{v}\in\mathbb{R}^n} \sum_{i=1}^n \|\vec{x}_i - v_i\vec{u}\|^2 \tag{18}$$

We now have a minimization of a sum of squares of vector norms $\|\vec{x}_i - v_i\vec{u}\|^2$, 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.,

$$\min_{x,y} f(x, y) = \min_{x,y}[g(x) + h(y)]. \tag{19}$$

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

$$(x^\star, y^\star) = \underset{x,y}{\operatorname{argmin}} f(x, y) = (\underset{x}{\operatorname{argmin}}\, g(x), \underset{y}{\operatorname{argmin}}\, h(y)). \tag{20}$$

We can split the minimization problem in 18 over each individual v_i. We have

$$\vec{u}^\star = \underset{\vec{u}\in\mathbb{R}^d:\vec{u}^\top\vec{u}=1,\vec{v}\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^n \min_{v_i\in\mathbb{R}} \|\vec{x}_i - v_i\vec{u}\|^2. \tag{21}$$

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

$$\|A\|_p \doteq \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p}. \tag{22}$$

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

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|. \tag{23}$$

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

$$\begin{aligned}
\|A\vec{x}\|_1 = \left\|\sum_{i=1}^{n} x_i \cdot \vec{a}_i\right\|_1 &\leq \sum_{i=1}^{n} \|x_i \cdot \vec{a}_i\|_1 && \text{(Triangle Inequality)} && (24)\\
&= \sum_{i=1}^{n} |x_i| \cdot \|\vec{a}_i\|_1 && \text{(}|x_i|\text{ can come out of norm} && (25)\\
&\leq \left(\sum_{i=1}^{n} |x_i|\right) \cdot \max_i \|\vec{a}_i\|_1 && (\|\vec{a}_j\|_1 \leq \max_i \|\vec{a}_i\|_1,\ \ \forall j) && (26)\\
&= \|\vec{x}\|_1 \cdot \max_i \|\vec{a}_i\|_1. && && (27)
\end{aligned}$$

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.

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

$$\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|. \tag{28}$$

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 $|x_j| \leq \max_i |x_i|,\ \ \forall j$.

Solution: We have,

$$\begin{aligned}
\|A\vec{x}\|_\infty &= \max_i \left|\sum_{j=1}^{n} a_{ij}x_j\right| && && (29)\\
&\leq \max_i \sum_{j=1}^{n} |a_{ij}x_j| && \text{(Triangle Inequality)} && (30)
\end{aligned}$$

$$\leq \max_i \left[\max_j |x_j| \Big(\sum_{j=1}^{n} |a_{ij}| \Big) \right] \qquad (|x_j| \leq \max_j |x_j|, \ \ \forall j) \tag{31}$$

$$= \left(\max_i \sum_{j=1}^{n} |a_{ij}| \right) \|\vec{x}\|_\infty . \tag{32}$$

Thus, we have

$$\|A\|_\infty = \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_\infty}{\|\vec{x}\|_\infty} \leq \max_i \sum_{j=1}^{n} |a_{ij}|. \tag{33}$$

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 $a_{mj} \geq 0$, and $x_j = -1$ if $a_{mj} < 0$. This leads to

$$\|A\vec{x}\|_\infty = \max_i \left| \sum_{j=1}^{n} a_{ij} x_j \right| \geq \left| \sum_{j=1}^{n} a_{mj} x_j \right| = \sum_{j=1}^{n} |a_{mj}| . \tag{34}$$

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

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

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

Solution: Approach 1: Rayleigh Coefficient

We have

$$\|A\|_2^2 = \left(\max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2} \right)^2 \tag{35}$$

$$= \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_2^2}{\|\vec{x}\|_2^2} \tag{36}$$

$$= \max_{\vec{x} \neq \vec{0}} \frac{(A\vec{x})^\top (A\vec{x})}{\vec{x}^\top \vec{x}} \tag{37}$$

$$= \max_{\vec{x} \neq \vec{0}} \frac{\vec{x}^\top A^\top A \vec{x}}{\vec{x}^\top \vec{x}} \tag{38}$$

$$= \lambda_{\max}\{A^\top A\} \tag{39}$$

$$= \sigma_{\max}\{A\}^2. \tag{40}$$

Taking square roots gives the solution.

Approach 2: SVD

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

$$\|A\|_2 = \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2} \tag{41}$$

$$= \max_{\vec{x} \neq \vec{0}} \frac{\left\|U \Sigma V^\top \vec{x}\right\|_2}{\|\vec{x}\|_2}. \tag{42}$$

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

$$\|A\|_2 = \max_{\vec{x} \neq \vec{0}} \frac{\left\|\Sigma V^\top \vec{x}\right\|_2}{\|\vec{x}\|_2}. \tag{43}$$

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

$$\|A\|_2 = \max_{\substack{\vec{x} \neq \vec{0} \\ \vec{z} = V^\top \vec{x}}} \frac{\|\Sigma \vec{z}\|_2}{\|V\vec{z}\|_2}. \tag{44}$$

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

$$\|A\|_2 = \max_{\substack{\vec{x} \neq \vec{0} \\ \vec{z} = V^\top \vec{x}}} \frac{\|\Sigma \vec{z}\|_2}{\|\vec{z}\|_2}. \tag{45}$$

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,

$$\|A\|_2 = \max_{\vec{z} \neq \vec{0}} \frac{\|\Sigma \vec{z}\|_2}{\|\vec{z}\|_2}. \tag{46}$$

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

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

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

$$\frac{\|\Sigma \vec{z}\|_2^2}{\|\vec{z}\|_2^2} = \sigma_1\{A\}^2 = \sigma_{\max}\{A\}^2. \tag{47}$$

Thus we have

$$\|A\|_2 = \max_{\vec{z} \neq \vec{0}} \frac{\|\Sigma \vec{z}\|_2}{\|\vec{z}\|_2} = \sigma_{\max}\{A\} \tag{48}$$

as claimed.

4. Gradients, Jacobian matrices and Hessians

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

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

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

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

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

$$(Dg)_{ij} = \frac{\partial g_i}{\partial x_j},\ \ i = 1, \ldots, m,\ \ j = 1, \ldots, n.$$

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.

(a) Compute the gradients and Hessians for the following functions. You are encouraged to use the chain rule as needed. You aren't required to open up the indices or compute element-wise for every problem, although for some problems it may be useful to do so.

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

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

iii. $g_3(\vec{x}) = \log\left(\sum_{i=1}^{n} e^{x_i}\right)$

iv. **(Practice)** $g_4(\vec{x}) = \log\left(\sum_{i=1}^{n} e^{\vec{a_i}^\top \vec{x} - b_i}\right)$

v. **(Practice)** $g_5(\vec{x}) = e^{\|A\vec{x} - b\|_2^2}$

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)

Solution:

i. Let $A = [a_1,\ a_2,\ \ldots\ , a_n]$ where a_i is the i-th column of A. Similarly, let $a_i^\top$ be the i-th row of $A^\top$. For notational convenience, let α_i^T denote the i-th row of A. Finally, let a_{ij} denote the (i, j)th entry of A. Then

$$\begin{aligned} g_1(\vec{x}) &= \vec{x}^\top A\vec{x} \\ &= \vec{x}^\top [a_1,\ a_2,\ \ldots\ , a_n]\vec{x} \\ &= \vec{x}^\top (a_1 x_1 + a_2 x_2 + \ \ldots\ + a_n x_n) \\ &= \sum_{i=1}^{n} (\vec{x}^\top a_i) x_i. \end{aligned}$$

Then,

$$\frac{\partial g_1}{\partial x_j}(\vec{x}) = \frac{\partial}{\partial x_j}\left[(\vec{x}^\top a_j)x_j + \sum_{i \neq j} (\vec{x}^\top a_i)x_i\right]$$

$$\begin{aligned} &= \vec{x}^\top a_j + a_{jj}x_j + \sum_{i\neq j} a_{ji}x_i \\ &= a_j^\top \vec{x} + \alpha_j^\top \vec{x}. \end{aligned}$$

It follows that $\nabla g_1(\vec{x}) = (A + A^\top)\vec{x}$. Note if A is symmetric this reduces to $2A\vec{x}$. Based on the definition of the Hessian, it follows that the ith column of the hessian is the ith column of $A + A^\top$. Thus $\nabla^2 g_1(\vec{x}) = A + A^\top$.

ii. Expanding the norm and using the fact that $c^\top \vec{x} = \vec{x}^\top c$ we have that $g_2(\vec{x}) = \vec{x}^\top A^\top A\vec{x} - 2b^\top A\vec{x} + b^\top b$. Using the previous results and the fact that $A^\top A$ is symmetric, it follows that $\nabla g_2(\vec{x}) = 2(A^\top A\vec{x} - A^\top b)$ and $\nabla^2 g_2(\vec{x}) = 2A^\top A$.

iii. $\frac{\partial g_3(\vec{x})}{\partial x_i} = \frac{e^{x_i}}{\sum\limits_{j=1}^{n} e^{x_j}}$ which gives us the gradient $\nabla g_3(\vec{x})$. For the Hessian, for $i \neq j$

$$\frac{\partial^2 g(\vec{x})}{\partial x_i \partial x_j} = -\frac{e^{x_i + x_j}}{(\sum\limits_{k=1}^{n} e^{x_k})^2}$$

and for $i = j$,

$$\frac{\partial^2 g(\vec{x})}{\partial x_i^2} = \frac{e^{x_i}}{\sum\limits_{k=1}^{n} e^{x_k}} - \frac{e^{x_i + x_j}}{(\sum\limits_{k=1}^{n} e^{x_k})^2}$$

which arranged into a matrix gives $\nabla^2 g_3(\vec{x})$

iv. Notice that $g_4(\vec{x}) = g_3(A\vec{x} - b)$ where the i^{th} row of A is a_i and the i^{th} entry of b is b_i. Using chain rule, we get $\nabla g_4(\vec{x}) = A^\top \nabla g_3(A\vec{x} - b)$. The Hessian is $\nabla^2 g_4(\vec{x}) = A^\top \nabla^2 g_3(A\vec{x} - b)A$.

v. Define $g_6(\vec{x}) = e^{\vec{x}^\top \vec{x}}$ so $g_5(\vec{x}) = g_6(A\vec{x} - b)$. We first compute the gradient of $g_6(\vec{x})$ using chain rule.

$$\begin{aligned} \nabla g_6(\vec{x}) &= \nabla_{\|\vec{x}\|^2}\left[e^{\|\vec{x}\|^2}\right]\nabla_{\vec{x}}\left[\vec{x}^\top I\vec{x}\right] \\ &= \left(e^{\|\vec{x}\|^2}\right)((I + I^\top)\vec{x}) \\ &= 2\vec{x}e^{\|\vec{x}\|^2} \end{aligned}$$

To find the hessian, we use the product rule $D(f(\vec{x})g(\vec{x})) = f(\vec{x})D(g(\vec{x})) + D(f(\vec{x}))g(\vec{x})$

$$\begin{aligned} \nabla^2 g_6(\vec{x}) &= D(\nabla g_6)(\vec{x}) \\ &= D(\vec{x})(2\vec{x})e^{\|\vec{x}\|^2} + 2\vec{x}D(\vec{x})\left(e^{\|\vec{x}\|^2}\right) \\ &= 2e^{\|\vec{x}\|^2}(I + 2\vec{x}\vec{x}^T) \end{aligned}$$

Now we can find the gradient and hessian of g_5 by applying the chain rule on g_6,

$$\begin{aligned} \nabla g_5(\vec{x}) &= \nabla_{\vec{x}} g_6(A\vec{x} - \vec{b}) \\ &= \nabla_{\vec{x}}\left[A\vec{x} - \vec{b}\right]\nabla_{A\vec{x}-\vec{b}}\left[g_6(A\vec{x} - \vec{b})\right] \\ &= 2A^\top(A\vec{x} - \vec{b})e^{\|A\vec{x}-\vec{b}\|^2} \end{aligned}$$

$$\nabla^2 g_5(\vec{x}) = D(\nabla g_5)(\vec{x})$$

$$
\begin{aligned}
&= D(\vec{x})(2A^\top(A\vec{x}-\vec{b}))e^{\|A\vec{x}-\vec{b}\|^2} + 2A^\top(A\vec{x}-\vec{b})D(\vec{x})\big(e^{\|A\vec{x}-\vec{b}\|^2}\big) \\
&= 2A^\top A e^{\|A\vec{x}-\vec{b}\|^2} + 4A^\top(A\vec{x}-\vec{b})(A\vec{x}-\vec{b})^\top A\big(e^{\|A\vec{x}-\vec{b}\|^2}\big) \\
&= e^{\|A\vec{x}-\vec{b}\|^2} 2A^\top(I + 2(A\vec{x}-\vec{b})(A\vec{x}-\vec{b}))A
\end{aligned}
$$

(b) Compute the Jacobians for the following maps

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

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

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

Solution:

i. Note $g_i(\vec{x}) = \alpha_i^\top \vec{x}$ where $\alpha_i^\top$ is the i-th row of A. Then $\dfrac{\partial g_i}{\partial x_j} = \alpha_{ij}$ which is simply the (i, j) entry of A. It follows that $Dg(\vec{x}) = A$.

ii. Again $g_i(\vec{x}) = f(\vec{x})x_i$. Then

$$
\begin{aligned}
\frac{\partial g_i}{\partial x_i} &= f(\vec{x}) + (\nabla f(\vec{x}))_i x_i \\
\frac{\partial g_i}{\partial x_j} &= 0 + (\nabla f(\vec{x}))_j x_i.
\end{aligned}
$$

It follows that $Dg(\vec{x}) = \vec{x}(\nabla f(\vec{x}))^\top + f(\vec{x})I$.

Alternate Solution: The Jacobian is just stacking up the transpose of the gradients of all the g_i's row by row. The gradient of g_i has been computed above and can be seen to be $\nabla f(\vec{x})x_i + f(\vec{x})e_i$ where e_i is the vector with the i^{th} entry as one and all other entries as zero. Hence stacking up the transpose of these vectors row by row, the first of the two terms in the gradient stacked up lead to the matrix $\vec{x}\nabla f(\vec{x})^\top$ and the second term stacked together leads to $f(\vec{x})I$. Hence we get the Jacobian to be $\vec{x}(\nabla f(\vec{x}))^\top + f(\vec{x})I$

iii. First, we derive $\nabla f(z)$. Let $z = A\vec{x} + b$. Let $\alpha_i^\top$ and a_i denote the i-th row and i-th column of A respectively. Finally let a_{ij} denote (i, j)th entry of A. Note $z_i = \alpha_i^\top \vec{x} + b_i$. Then by the chain rule we have

$$
\begin{aligned}
\frac{\partial f}{\partial x_j} &= \sum_{i=1}^n \frac{\partial f}{\partial z_i}\frac{\partial z_i}{\partial x_j} \\
&= \sum_{i=1}^n \frac{\partial f}{\partial z_i} a_{ij} \\
&= a_j^\top \nabla f(z).
\end{aligned}
$$

It follows that $\nabla f(A\vec{x} + b) = \nabla f(z) = A^\top \nabla f(A\vec{x} + b)$.

Returning to the original problem, we have $g_i(\vec{x}) = f(A\vec{x} + b)x_i$. Then using the derivation in the previous part, it follows that $Dg(\vec{x}) = \vec{x}(\nabla f(A\vec{x} + b))^\top A + f(A\vec{x} + b)I$.

For another way to derive the result, one can work out like in the previous part's alternate solution to arrive at the same answer.

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

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

ii. $g(x_1, x_2) = x_1x_2$

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:

$$\frac{\partial g}{\partial x_1}(x_1, x_2) = \frac{x_1}{2}, \quad \frac{\partial g}{\partial x_2}(x_1, x_2) = \frac{2x_2}{9}, \tag{49}$$

$$\frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) = \frac{1}{2}, \quad \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) = 0, \tag{50}$$

$$\frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) = \frac{2}{9}, \quad \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) = 0. \tag{51}$$

The gradient of g is then given by,

$$\nabla g(x_1, x_2) = \begin{bmatrix} \frac{\partial g}{\partial x_1}(1,1) \\ \frac{\partial g}{\partial x_2}(1,1) \end{bmatrix}, \tag{52}$$

and the Hessian matrix is given by,

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) \\ \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) & \frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) \end{bmatrix}. \tag{53}$$

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

$$g(x_1, x_2) \approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{54}$$

$$= \frac{13}{36} + \frac{x_1}{2} - \frac{1}{2} + \frac{2x_2}{9} - \frac{2}{9} = \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36}. \tag{55}$$

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

$$g(x_1, x_2) \approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x_1 - 1 & x_2 - 1 \end{bmatrix} H(1,1) \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{56}$$

$$= \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36} + \frac{1}{2}\left(\frac{1}{2}(x_1 - 1)^2 + \frac{2}{9}(x_2 - 1)^2\right) \tag{57}$$

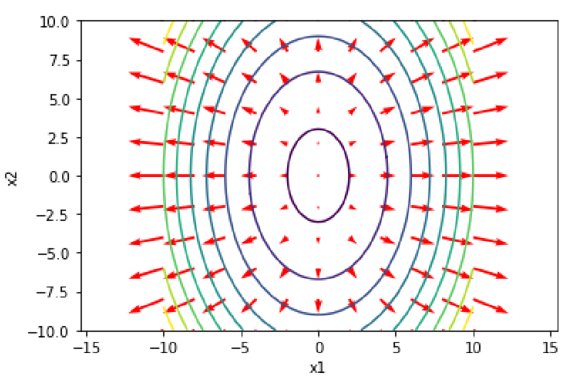

Figure 1: Level sets and gradient directions for the function $g(x_1, x_2) = \frac{x_1^2}{4} + \frac{x_2^2}{9}$.

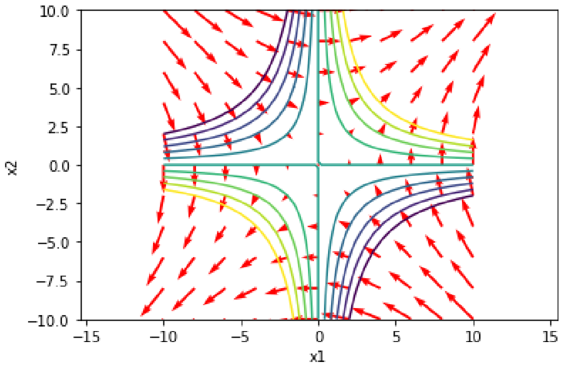

Figure 2: Level sets and gradient directions for the function $g(x_1, x_2) = x_1 x_2$.

$$= \frac{(x_1-1)^2}{4} + \frac{(x_2-1)^2}{9} + \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36}. \tag{58}$$

$$= \frac{x_1^2}{4} + \frac{x_2^2}{9} \tag{59}$$

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:

$$\frac{\partial g}{\partial x_1}(x_1, x_2) = x_2, \quad \frac{\partial g}{\partial x_2}(x_1, x_2) = x_1, \tag{60}$$

$$\frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) = 0, \quad \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) = 1, \tag{61}$$

$$\frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) = 0, \quad \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) = 1. \tag{62}$$

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

$$g(x_1, x_2) \approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{63}$$

$$= 1 + x_1 - 1 + x_2 - 1 = x_1 + x_2 - 1. \tag{64}$$

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

$$g(x_1, x_2) \approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 - 1 & x_2 - 1 \end{bmatrix} H(1,1) \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \tag{65}$$

$$= x_1 + x_2 - 1 + \frac{1}{2}\left(2(x_1 - 1)(x_2 - 1)\right) \tag{66}$$

$$= (x_1 - 1)(x_2 - 1) + x_1 + x_2 - 1. \tag{67}$$

$$= x_1 x_2 \tag{68}$$

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. Condition Number

In lecture and the course reader, we examined the sensitivity of solutions to linear system $A\vec{x} = \vec{y}$ (for nonsingular/invertible square matrix A) to perturbations in our measurements $\vec{y}$. Specifically, we showed that if we model measurement noise $\Delta\vec{y}$ as a linear perturbation on $\vec{y}$, resulting in a linear perturbation $\Delta\vec{x}$ on $\vec{x}$ — i.e., $A(\vec{x} + \Delta\vec{x}) = \vec{y} + \Delta\vec{y}$ — we can bound the magnitude of the solution perturbations $\Delta\vec{x}$ as

$$\frac{\|\Delta\vec{x}\|_2}{\|\vec{x}\|_2} \leq \kappa(A)\frac{\|\Delta\vec{y}\|_2}{\|\vec{y}\|_2}, \tag{69}$$

where $\kappa(A) = \frac{\sigma_{\max}\{A\}}{\sigma_{\min}\{A\}}$ is the condition number of A, or the ratio of A's maximum and minimum singular values. In this problem, we will establish a similar bound for perturbations on A.

(a) Consider the linear system $A\vec{x} = \vec{y}$ above, where $A \in \mathbb{R}^{n\times n}$ is invertible (i.e., square and nonsingular). Let $\Delta A \in \mathbb{R}^{n\times n}$ denote a linear perturbation on matrix A generating a corresponding linear perturbation $\Delta\vec{x}$ in solution $\vec{x}$, i.e.,

$$(A + \Delta A)(\vec{x} + \Delta\vec{x}) = \vec{y}. \tag{70}$$

Show that

$$\frac{\|\Delta\vec{x}\|_2}{\|\vec{x} + \Delta\vec{x}\|_2} \leq \kappa(A)\frac{\|\Delta A\|_2}{\|A\|_2}. \tag{71}$$

Solution: Rearranging the given linear system equation, we have

$$
\begin{aligned}
(A + \Delta A)(\vec{x} + \Delta\vec{x}) = \vec{y} &\implies A\vec{x} + A\Delta\vec{x} + \Delta A\vec{x} + \Delta A\Delta\vec{x} = \vec{y} && (72)\\
&\implies A\Delta\vec{x} + \Delta A\vec{x} + \Delta A\Delta\vec{x} = \vec{0} \qquad (A\vec{x} = \vec{y}) && (73)\\
&\implies A\Delta\vec{x} = -\Delta A(\vec{x} + \Delta\vec{x}) && (74)\\
&\implies \Delta\vec{x} = -A^{-1}\Delta A(\vec{x} + \Delta\vec{x}) && (75)\\
&\implies \|\Delta\vec{x}\|_2 = \left\|A^{-1}\Delta A(\vec{x} + \Delta\vec{x})\right\|_2 \leq \left\|A^{-1}\right\|_2 \|\Delta A\|_2 \|\vec{x} + \Delta\vec{x}\|_2 && (76)\\
&\implies \|\Delta\vec{x}\|_2 \leq \left\|A^{-1}\right\|_2 \|\Delta A\|_2 \|\vec{x} + \Delta\vec{x}\|_2 \frac{\|A\|_2}{\|A\|_2} && (77)\\
&\implies \frac{\|\Delta\vec{x}\|_2}{\|\vec{x} + \Delta\vec{x}\|_2} \leq \|A\|_2 \left\|A^{-1}\right\|_2 \frac{\|\Delta A\|_2}{\|A\|_2} && (78)\\
&\implies \frac{\|\Delta\vec{x}\|_2}{\|\vec{x} + \Delta\vec{x}\|_2} \leq \sigma_{\max}\{A\}\frac{1}{\sigma_{\min}\{A\}}\frac{\|\Delta A\|_2}{\|A\|_2} && (79)\\
&\implies \frac{\|\Delta\vec{x}\|_2}{\|\vec{x} + \Delta\vec{x}\|_2} \leq \kappa(A)\frac{\|\Delta A\|_2}{\|A\|_2} && (80)
\end{aligned}
$$

as desired.

(b) Note that Equations (69) and (71) above bound two slightly different quantities: $\dfrac{\|\Delta\vec{x}\|_2}{\|\vec{x}\|_2}$ and $\dfrac{\|\Delta\vec{x}\|_2}{\|\vec{x} + \Delta\vec{x}\|_2}$, respectively. In general, we wish to establish these bounds because we want to characterize the size of $\Delta\vec{x}$ under different sizes of perturbation. Which of these two bounds better serves this purpose? *HINT: Consider different relative values of $\vec{x}$ and $\Delta\vec{x}$. What happens to the bounds when $\Delta\vec{x} \gg \vec{x}$?*

Solution: When $\Delta\vec{x}$ is small relative to $\vec{x}$, both bounds are almost equivalent, since $\frac{\|\Delta\vec{x}\|_2}{\|\vec{x}+\Delta\vec{x}\|_2} \sim \frac{\|\Delta\vec{x}\|_2}{\|\vec{x}\|_2}$ for small $\Delta\vec{x}$. However, when $\Delta\vec{x}$ is very large relative to $\vec{x}$, $\frac{\|\Delta\vec{x}\|_2}{\|\vec{x}+\Delta\vec{x}\|_2} \sim \frac{\|\Delta\vec{x}\|_2}{\|\Delta\vec{x}\|_2} = 1$ regardless of the value of $\Delta\vec{x}$, so our bound in Equation (71) tells us nothing about $\Delta\vec{x}$'s size. Our bound on solution error for perturbations in $\vec{y}$ in Equation (69) is therefore much more useful for characterizing $\Delta\vec{x}$ over a wider range of perturbations than our bound on solution error for perturbations in A in Equation (71).

6. Direction of Steepest Ascent

For a differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ we want to show that the gradient $\nabla f(\vec{x})$ is the direction of steepest ascent at the point $\vec{x}$.

(a) Let us define the rate of change of the function $f(\vec{x})$ at the point $\vec{x}$ along an arbitrary unit vector $\vec{u}$ as:

$$D_{\vec{u}} f(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{u}) - f(\vec{x})}{h}. \tag{81}$$

We call this the directional derivative. Show that the directional derivative can be equivalently expressed as $D_{\vec{u}} f(\vec{x}) = \vec{u}^\top [\nabla f(\vec{x})]$.

HINT: Use Taylor approximation of the function around the point $\vec{x}$ and evaluate it at the point $\vec{x} + h\vec{u}$.

Solution: Using Taylor's theorem we can express the function $f(\vec{x})$ as

$$f(\vec{x} + h\vec{u}) = f(\vec{x}) + [\nabla f(\vec{x})]^\top [h\vec{u}] + o(h). \tag{82}$$

We rearrange the terms, and dividing both sides by h we get

$$\frac{f(\vec{x} + h\vec{u}) - f(\vec{x})}{h} = [\nabla f(\vec{x})]^\top [\vec{u}] + \frac{o(h)}{h}. \tag{83}$$

Now we take the limit of both sides as $h \to 0$; we get

$$\lim_{h \to 0} \frac{f(\vec{x} + h\vec{u}) - f(\vec{x})}{h} = [\nabla f(\vec{x})]^\top [\vec{u}] + \lim_{h \to 0} \left(\frac{o(h)}{h} \right) \tag{84}$$

$$= [\nabla f(\vec{x})]^\top [\vec{u}]. \tag{85}$$

Note that $\lim_{h \to 0} \frac{o(h)}{h} = 0$ because $o(h)$ decays faster than h as $h \to 0$.

(b) Show that

$$\frac{\nabla f(\vec{x})}{\|\nabla f(\vec{x})\|_2} = \underset{\|\vec{u}\|_2 = 1}{\operatorname{argmax}} \, \vec{u}^\top [\nabla f(\vec{x})]. \tag{86}$$

Solution: Using Cauchy-Schwarz inequality we can write:

$$\vec{u}^\top [\nabla f(\vec{x})] \leq \|\vec{u}\|_2 \|\nabla f(\vec{x})\|_2 \tag{87}$$

$$= \|\nabla f(\vec{x})\|_2 , \tag{88}$$

so the maximum value that the expression can take is $\|\nabla f(\vec{x})\|_2$. Now it remains to show that this value is attained for the choice $\vec{u} = \frac{\nabla f(\vec{x})}{\|\nabla f(\vec{x})\|_2}$.

$$\frac{[\nabla f(\vec{x})]^\top}{\|\nabla f(\vec{x})\|_2} \nabla f(\vec{x}) = \frac{\|\nabla f(\vec{x})\|_2^2}{\|\nabla f(\vec{x})\|_2} \tag{89}$$

$$= \|\nabla f(\vec{x})\|_2 . \tag{90}$$

7. 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.

