COMPSCI 527 Homework 0

Problem 0 (3 points)

Part 1: Eigenvalues and Eigenvectors

Problem 1.1

We define orthogonality by the dot product of the two real eigenvectors $\vec{v_1}, \vec{v_2}$ as the following:

$\vec{v_1} \cdot \vec{v_2} := {v_1}^\intercal v_2 = 0$

We also define eigenvectors by the following equation, where Q is some symmetric, real matrix and $\vec{v_1} \neq \vec{0}$:

$$Q\vec{v_1} = \lambda\vec{v_1}$$

We can take the dot product of each side of each equation with the transpose of the opposing eigenvector like so:

$$\begin{gathered} Q\vec{v_1} = \lambda_1\vec{v_1} \\ Q\vec{v_2} = \lambda_2\vec{v_2} \\ \downarrow \\ \vec{v_2}^\intercal Q\vec{v_1} = \lambda_1 {v_2}^\intercal v_1 \\ \vec{v_1}^\intercal Q\vec{v_2} = \lambda_2 {v_1}^\intercal v_2 \end{gathered}$$

We can take the transpose of either of the equations to match the other. For example:

$$\begin{gathered} \vec{v_2}^\intercal Q\vec{v_1} = \lambda_1 {v_2}^\intercal v_1 \\ \xrightarrow{\text{transpose}} \\ \vec{v_1}^\intercal Q\vec{v_2} = \lambda_1 {v_1}^\intercal v_2 \end{gathered}$$

Now, we can subtract the transposed result from its matching equation to set the left side to zero:

$$0 = \lambda_1 {v_1}^\intercal v_2 - \lambda_1 {v_1}^\intercal v_2$$

Which, simplified, leaves:

$$0 = (\lambda_1 - \lambda_2){v_1}^\intercal v_2$$

We are given that $\lambda_1 > \lambda_2$; therefore, the difference $\lambda_1 - \lambda_2$ is nonzero.

Therefore, we can conclude that ${v_1}^\intercal v_2 = 0$, which is only true if the vectors are orthogonal.

Therefore, two eigenvectors are orthogonal under the given conditions.

Problem 1.2

Using Lagrange multipliers, we can prove that the unit vector $\vec{x}$ represent the eigenvectors with the largest and smallest eigenvalues when $q(\vec{x}) = \vec{x}^{\mathsf{T}} Q \vec{x}$ is maximized and minimized.

Let the objective function be:

$$q(\vec{x}) = \vec{x}^{\mathsf{T}} Q \vec{x}$$

Let the constraint function be:

$$g(\vec{x}) = |\vec{x}|^2 - 1 = 0$$

Let q_x denote the partial derivative of $q(\vec{x})$, g_x denote the partial derivative of g(x), and λ as the multiplier, we solve for:

$$q_x = \lambda g_x$$

From the definition of our constraint, we have:

$$g_x(\vec{x}) = 2\vec{x}$$

And the partial derivative of $q(\vec{x})$ would be:

$$q_x(\vec{x}) = Q\vec{x} + (\vec{x}^{\mathsf{T}} Q)^{\mathsf{T}} = 2Q\vec{x}$$

Substituting these into the Lagrangian yields:

$$\lambda 2\vec{x} = 2Q\vec{x}$$

$$\lambda \vec{x} = Q\vec{x}$$

Here we can see that the Lagrangian multiplier can also represent the eigenvalue with $\vec{x}$ its associated eigenvector. Since $q(\vec{x}) = \vec{x}^{\mathsf{T}} Q \vec{x}$,

$$q(\vec{x}) = \lambda x^{\mathsf{T}} x = \lambda ||\vec{x}|| = \lambda$$

Since we are maximizing and minimizing $q(\vec{x})$ for all unit vectors $\vec{x}$, we are essentially maximizing and minimizing λ. Therefore, the associated eigenvectors of λ_1 and λ_2 are the eigenvectors with the largest and smallest eigenvalue.

Problem 1.3

Say the set of all real and symmetric matrices with equal eigenvalues is given by:

$$Q = \begin{bmatrix} r & s \\ s & t \end{bmatrix}$$

The objective is to determine under what conditions of r,s and t will the eigenvalues λ_1 and λ_2 of Q become equal. That is:

$$\lambda_1 = \lambda_2$$

Going by the definition given to evaluate the eigenvalues of a matrix, we have:

$$|Q - \lambda I| = 0$$

$$\left| \begin{bmatrix} r & s \\ s & t \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$$

$$\begin{vmatrix} r-\lambda & s \\ s & t-\lambda \end{vmatrix} = 0$$

$$(r-\lambda)(t-\lambda) - s^2 = 0 \tag{1}$$

On expanding the equation, we obtain a quadratic equation in the variable λ:

$$\lambda^2 - (r+t)\lambda + (rt - s^2) = 0$$

If $\lambda_1 = \lambda_2$, then the discriminant of the above equation has to be 0.

$$(r-t)^2 + 4s^2 = 0$$

By taking $4s^2$ on the left-hand-side and taking the square root, we will obtain complex solutions to the values of r,s and t. However, that is not possible since the given data claims that the given matrix is real-valued.

Hence, the only real solutions to the equation occurs when:

$$r = t, s = 0, r \in \mathbb{R}$$

That is,

the set of all real-valued symmetric matrices with equal eigenvalues are essentially scalar matrices.

$$Q = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$$

The eigenvalues of scalar matrices are obtained by equation (1):

$$(r-\lambda)(r-\lambda) = 0$$

$$\lambda = r$$

The eigenvalues of scalar matrices are equal to the diagonal elements of the given matrix.

Say the unit-norm eigenvector corresponding to λ_1 is $\vec{v_1}$. By defition of eigenvalues and eigenvectors:

$$Q\vec{v_1} = \lambda_1 \vec{v_1}$$

$$\begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$$

$$\vec{v_1} = r\vec{v_1}$$

It is clear that there is no unique eigenvector which satsifies the given condition.

There are infinitely many unit-norm eigenvectors for the given type of matrices and eigenvalues.

This is also justified by the fact that real-symmetric matrices perform a linear transformation on vectors that only perform scaling function. The elements of these matrices, also the eigenvalues, correspond to the change in the magnitude of these vectors alone as the diretion continues to remain unchanged. Hence, there exist infinite eigenvectors under such conditions.

Part 2: Singular Values and Singular Vectors

Problem 2.1

The newly obtained vectors post the linear tranformation are described as follows:

$$\vec{y_1} = \sigma_1 \vec{u_1} \tag{1}$$

$$\vec{y_2} = \sigma_2 \vec{u_2} \tag{2}$$

To Prove:

$\vec{u}_1$ and $\vec{u}_2$ are perpendicular to each other. That is, the dot product of these vectors is zero:

$$\vec{u}_1 \cdot \vec{u}_2 = {u_1}^\intercal u_2 = {u_2}^\intercal u_1 = 0$$

Proof:

Taking transpose of equation (2):

$$\vec{y_2}^\intercal = \sigma_2 \vec{u_2}^\intercal \tag{3}$$

Multiplying (1) with (3):

$$y_2{}^\intercal y_1 = \sigma_2 \vec{u_2}^\intercal \sigma_1 \vec{u_1} = \sigma_1 \sigma_2 u_2{}^\intercal u_1 \tag{4}$$

We also know that the longest vectors $\vec{y_1}$ and $\vec{y_2}$ are expressed as:

$$\vec{y_1} = A\vec{v_1}$$

$$\vec{y_2} = A\vec{v_2}$$

Substituting the above in (4):

$$\vec{v_2}^\intercal A^\intercal A \vec{v_1} = \sigma_1 \sigma_1 u_2{}^\intercal u_1$$

It is known that:

$$Q = A^\intercal A$$

The eigenvalues of Q are λ_1 and λ_2. The corresponding eigenvectors are $\vec{v}_1$ and $\vec{v}_2$ respectively.

This implies:

$$\vec{v_2}^\intercal Q \vec{v_1} = \sigma_1 \sigma_2 u_2{}^\intercal u_1$$

But,

$$Q\vec{v}_1 = \lambda_1\vec{v}_1$$

The final equation computes to:

$$\lambda_1 {v_2}^\intercal v_1 = \sigma_1\sigma_2 {u_2}^\intercal u_1$$

Remembering that $\vec{v}_1$ and $\vec{v}_2$ are mutually orthogonal vectors,

$${v_2}^\intercal v_1 = 0$$

Therefore,

$$\sigma_1\sigma_2 {u_2}^\intercal u_1 = 0$$

Since the eigenvalues are non-zero, λ_1 and λ_2, σ_1 and σ_2 are also non-zero.

Hence,

$${u_2}^\intercal u_1 = 0 = \vec{u}_2 \cdot \vec{u}_1$$

$\vec{u}_1$ and $\vec{u}_2$ are thus perpendicular to one another.

Problem 2.2

Given:

$$\vec{y} = A\vec{x}$$

To derive:

Equation of the ellipse E, that confirms to the below condition, in terms of A:

$$\vec{y}^\intercal E\vec{y} = 1$$

Method:

We are given that

$$\vec{y} = A\vec{x} \tag{1}$$

Multiplying (1) with $A^\intercal$:

$$A^\intercal\vec{y} = A^\intercal A\vec{x} \tag{2}$$

But it is known that

$$Q = A^\intercal A$$

Substituting in (2):

$$A^\intercal\vec{y} = Q\vec{x} \tag{3}$$

Multiplying (3) with Q^{-1}:

$$Q^{-1}A^\intercal\vec{y} = \vec{x} \tag{4}$$

Taking transpose of (4):

$$\vec{y}^\intercal A(Q^{-1})^\intercal = \vec{x}^\intercal \tag{5}$$

The area under consideration is a unit-circle with $\vec{x}$ as the radius. This gives:

$$x^\intercal x = 1$$

Multiplying (5) with (4):

$$x^\intercal x = \vec{y}^\intercal A(Q^{-1})^\intercal Q^{-1} A^\intercal \vec{y} = 1$$

Equating this to:

$$\vec{y}^\intercal E \vec{y} = 1$$

We get:

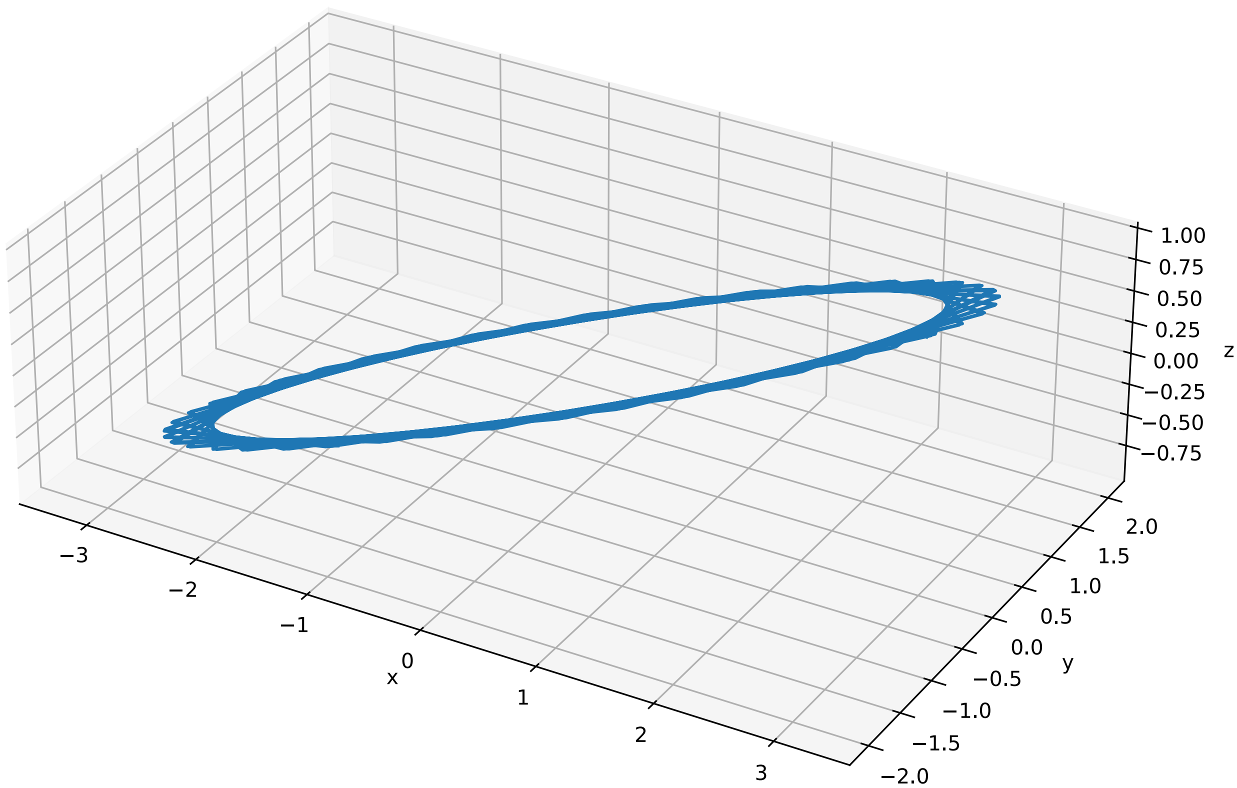
$$E = A(Q^{-1})^\intercal Q^{-1} A^\intercal$$

Problem 2.3

In [5]:

```
import numpy as np
from matplotlib import pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

%matplotlib inline

def ellipse(a):
    '''
    using input 3x2 matrix, the function multiplies the matrix
    with coordinates passed in as a nx2 vector describing an ellipse (i.e. a
    R^2 vector with n coordinates). The returned matrix is therefore 3xn in
    dimensions, and represents the vector transformed by the matrix.
    '''

    assert np.all(a.shape == (3, 2)), 'matrix must be 3 x 2'

    n = 100

    # Generates coordinates for the unit vector x = <cos(x), sin(x)>,
    # as x^Tx = 1 indicates that x is a vector anywhere on the unit circle
    # converts coordinates to numpy array and transposes the matrix

    x_coords = [[np.cos(t), np.sin(t)] for t in range(n)] # 2 x 100 matrix
    x_coords = np.array(x_coords).T
    y_coords = np.matmul(a, x_coords)              # y = matrix A * vector x
    return y_coords

def draw_3d_curve(y):
    fig = plt.figure(figsize=(12, 10))
    ax = fig.gca(projection='3d')
    ax.plot(y[0], y[1], y[2])
    axes = ['x','y','z']
    ranges = [np.ptp(yk) for yk in y]
    ax.set_box_aspect(ranges)
    ax.set_xlabel('x')
    ax.set_ylabel('y')
```

```
    ax.set_zlabel('z')
    plt.show()

a0 = np.array([[3., 1.], [0., 2.], [1., 0.]])

a0x = ellipse(a0)     # generate R^3 ellipse
draw_3d_curve(a0x)    # draw plot
```



Part 3: The Singular Value Decomposition

Problem 3.1

The cross product of two vectors is orthogonal to both. Take cross product of $\vec{u}$ and $\vec{v}$:

$$\vec{u} \times \vec{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$$

Problem 3.2

We modify Σ so that its dimensions are 3×2 by adding a row of zeros to the matrix. This collapses $\vec{u}_3$ to a point.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix}$$

Part 4: Calculus

Problem 4.1

We can constrain the given function $f(x, y) = y(y - 2x^2)$ to all values $y = mx$. The resulting function is therefore $\phi_m(x) = (mx)^2 - 2mx^3$.

From here, we can examine the extrema of the function located at the roots of the derivative $\phi_m(x') = 2m^2x - 6mx^2$, and the concavity of the function around those extrema with the second derivative $\phi_m(x'') = 2m^2 - 12mx$.

The zeros of $f'(x)$ are easily found by factoring the derivative to the following:

$2mx(m - 3x) = 0$

This equation has the solutions $x = 0, x = \frac{m}{3}$. Therefore, along the line $y = mx$, the point $(0, 0)$ is some extrema.

Likewise, there is a single zero on second derivative, indicating a single inflection point. This point is:

$2m(m - 6x) = 0$

This equation has the solution $x = \frac{m}{6}$.

Now, we can split up m into a set of critical intervals. To reason about what intervals we should establish, we have to consider the behavior of the function.

As a third degree polynomial, the function has at most three roots, two extrema, and one inflection point, regardless of the value of m.

We are investigating the point at which $y = 0$, which is only true if $x = 0$ or $m = 0$ by $y = mx$. Therefore, we want to look around the point $y = 0$ when $m = 0$, or when $x = 0$ and $m \in \mathbb{R}$. Therefore, we can set up the following table for the value of the function $\phi_m(x)$ and its derivatives:

m	$\phi_m(x)$	$\phi_m(x')$	$\phi_m(x'')$
$(-\infty, 0)$	$(mx)^2 + 2mx^3$	$2m^2x + 6mx^2$	$2m^2 + 12mx$
$= 0$	0	0	0
$(0, +\infty)$	$(mx)^2 - 2mx^3$	$2m^2x - 6mx^2$	$2m^2 - 12mx$

We can make a few conclusions from this table. Most obviously, when $m = 0$, $\phi_m(x) = 0$ for any $x \in \mathbb{R}$. This resolves our first case proposed.

When $m < 0$, the function's highest degree flips sign. Though this has implications on the location of the other extremum ($x = \frac{m}{6}$; it reflects it across the y axis), this doesn't affect the location of the extremum we are investigating at $x = 0$, compared to when $m > 0$. Therefore, for $m \in (-\infty, 0) \cup (0, +\infty)$, the concavity at the point $y = 0$ is consistent.

Specifically, the concavity at the point $y = 0$ is determined by the value of $\phi_m(x'')$ at the point (0,0). $\phi_m(0) = 2m^2$ is always positive for $m \in (-\infty, 0) \cup (0, +\infty)$; therefore, the point at y = 0 is always concave up, indicating a minimum.

Finally, we must determine the nature of the minimum at $y = 0$. In simple terms, a weak minimum is one where the value of x at the minimum is seen "around" that minimum along the function $f(x)$. We can disprove a weak minimum around $x = 0$ easily, by remembering that if the second order derivative continues to either positive or negative infinity on both ends, then the optimum is strict. Since $\phi_m(x'') = 2m^2$ is quadratic, its limits are the same sign and approach∞ at both ends of the function. We can therefore conclude that the point $(0, 0)$ is a strict local minimum for $m \neq 0, m \in \mathbb{R}$. At $m = 0$, the function resolves to $\phi_m(x) = 0$, a flat line (both a weak maximum and a weak minimum).

Problem 4.2

For the function $f(x, y) = y(y - 2x^2)$, the Hessian matrix ($H_f(x)$) and gradient ($\nabla f(x, y)$) are the following:

$$H_f(x) = \begin{bmatrix} \frac{\delta^2 f}{\delta x^2} & \frac{\delta^2 f}{\delta x \delta y} \\ \frac{\delta^2 f}{\delta x \delta y} & \frac{\delta^2 f}{\delta y^2} \end{bmatrix} \rightsquigarrow \begin{bmatrix} -4y & -4x \\ -4x & 2 \end{bmatrix}$$

$$\nabla f(x, y) = \langle f_x(x), f_y(y) \rangle \rightsquigarrow \langle -4xy, 2y - 2x^2 \rangle$$

At the point $(x, y) = (0, 0)$, these resolve to:

$$H_f(0) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\nabla f(0, 0) = \langle 0, 0 \rangle$$

Problem 4.3

To determine the type of point present at $(x, y) = (0, 0)$, we will need to use an alternative approach compared to 4.1, as the function is no longer limited to a special case where y is limited by x to a single dimension. To do this, we can find the eigenvalues of the Hessian, which, depending on their values, may provide additional information.

The eigenvalues of the Hessian are given by the equation

$$H_f(x)\vec{x} = \lambda\vec{x}$$

whose solutions can be obtained by taking the determinant of $H_f(x) - I(\lambda) = 0$.

This solution is therefore:

$$\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \rightsquigarrow \begin{bmatrix} -\lambda & 0 \\ 0 & 2 - \lambda \end{bmatrix}$$

$$det(\begin{bmatrix} -\lambda & 0 \\ 0 & 2-\lambda \end{bmatrix}) = 0 \rightsquigarrow -\lambda(2-\lambda) = 0$$

Thus, the eigenvalues of this function are $\lambda_1 = 0, \lambda_2 = 1$.

This case of the Hessian, where one or more of the values are zero, is an indeterminate case. Therefore, we cannot use the Hessian to determine the type of point $(0,0)$ is.

Since the gradient is zero, we can determine at least that the point is a stationary point, meaning we are left with three options:

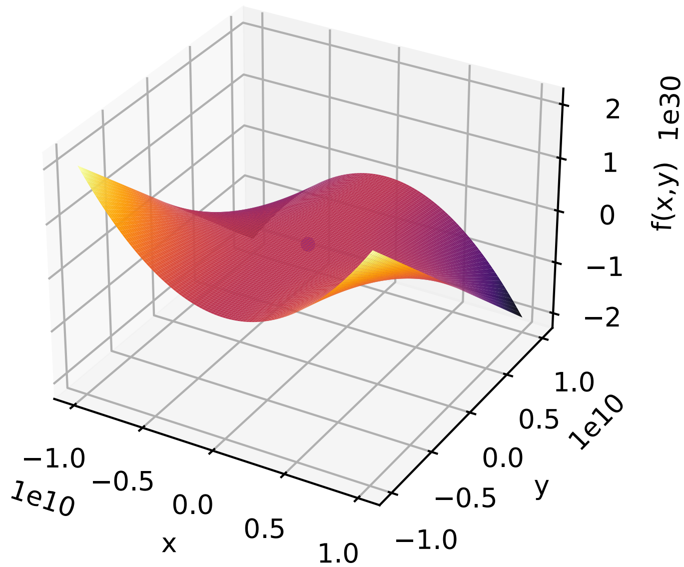
1. A saddle point,
2. A minimum, or
3. A maximum.

To make dealing with this question simpler, we can take a look at the function to examine it graphically and attempt to classify it by inspection.

In [6]:

```
from mpl_toolkits import mplot3d

%matplotlib inline

import jupyter
import numpy as np
import matplotlib.pyplot as plt


def f(x, y):                                    #the function
  return y*(y - 2*x**2)


fig = plt.figure()
ax = plt.axes(projection = "3d")

xy_bounds = np.linspace(-(10**10),10**10, 10**2)
x, y = xy_bounds, xy_bounds
x_ax, y_ax = np.meshgrid(x, y)                 #building grid

z_ax = f(x_ax, y_ax)

ax.plot_surface(
    x_ax, y_ax, z_ax,
    rstride = 1, cstride = 1,
    cmap = 'inferno')

ax.scatter(0,0,0, c = "blue")

ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('f(x,y)')
```

Out[6]: `Text(0.5, 0, 'f(x,y)')`



From our graph above, we can see that the point (seen here in the center of the space) is located in a position where it is neither a minimum (as there are points nearby less than the value of $f(x, y)$ at $(0, 0)$) nor a maximum (as there are points nearby greater than the value of $f(x, y)$ at $(0, 0)$). Therefore, we conclude by inspection that the point is a saddle point.