

Linear Algebra HW 10

gjd9961

November 2021

1 Problem 10.1

Compute critical points of f , g and h and determine if they are global/local maximizers/minimizers or saddle points. To determine the signs of eigenvalues it might be useful to remember that for $M \in \mathbb{R}^{n \times n}$ symmetric, $\text{tr}(M) = \sum_{i=1}^n M_{i,i} = \sum_{i=1}^n \lambda_i$.

a) $f : \mathbb{R} \rightarrow \mathbb{R}$ with $f(x) = (x^2 - 1)^2$

We compute the gradient and then the hessian to determine the behavior of the function. We can take the partial derivatives and calculate the gradient as such:

$$\nabla f(x) = f'(x) = (x^2 - 1)^2 \frac{df}{dx} = 4x(x^2 - 1)$$

We can then compute the hessian by taking the derivative of the gradient:

$$H_{f(x)} = \nabla f'(x) = 4x(x^2 - 1) \frac{d\nabla f(x)}{dx} = 12x^2 - 4$$

The polynomial resulting from the gradient has roots at $-1, 0, 1$, which are the critical points of this function. We can observe the behavior of the function around these points using the hessian to determine whether the critical points are local/global min/max's or saddle points.

$$\text{at } x = 1 \ H_f(x) = 8 \text{ at } x = -1 \ H_f(x) = 8 \text{ at } x = 0 \ H_f(x) = -4$$

Since the hessian is negative definite at $x = 0$ it is a local maximum. Since the hessian is positive definite at $x = 1$ and $x = -1$ they are each local minimums. As they are equal to each other, and the function is strictly increasing beyond those points, they are both in fact global minima as well.

b) $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ with $g(x, y, z) = (x^2 - z^2)y + 2$

We compute the gradient and then the hessian to determine the behavior of the function. We can take the partial derivatives and calculate the gradient as such:

$$\nabla g(x, y, z) = g'(x, y, z) \rightarrow \frac{\partial g}{\partial x} = 2xy \quad \frac{\partial g}{\partial y} = x^2 - z^2 \quad \frac{\partial g}{\partial z} = -2zy$$

Therefore:

$$\nabla g(x, y, z) = \begin{pmatrix} 2xy \\ x^2 - z^2 \\ -2zy \end{pmatrix}$$

Setting the gradient to 0, we can analyze where the function has critical points. Its clear to see from our system of equations that $x^2 = z^2$ and that any time the variables x, y, z are equal to 0, the gradient will be 0. Therefore, the following are critical points:

$$\begin{pmatrix} x \\ 0 \\ x \end{pmatrix} \text{ and } \begin{pmatrix} x \\ 0 \\ -x \end{pmatrix} \text{ and } \begin{pmatrix} -x \\ 0 \\ x \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix}$$

For vectors that take the form $(x, y, z)^T$. Now that we have the gradient and the location of the critical points, we can take the derivative of the gradient to compute the hessian to analyze the behavior of the function at the critical points:

$$H_{g(x,y,z)} = \begin{pmatrix} 2y & 2x & 0 \\ 2x & 0 & -2z \\ 0 & -2z & -2y \end{pmatrix}$$

One quick note: since the hessian is always square and symmetric, we know that the sum of it's eigenvalues are equal to the sum of its trace. In our case, the diagonal of the hessian has the entries of $2y, 0$ and $-2y$, which sums to 0 for any value of y . This will inform our decision making when understanding if a critical point is a global/local minimum or maximizer, undefined, or a saddle point.

Lets evaluate the critical points, starting with $(0, y, 0)$:

$$\text{If } \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} \rightarrow H_{g(x,y,z)} = \begin{pmatrix} 2y & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2y \end{pmatrix}$$

Since one of the columns in the hessian is the 0 vector, we know that $\dim(\text{Ker}(H_g(x, y, z))) = 1$, $\text{Rank}(H_g(x, y, z)) = 2$ and one eigenvalue will be 0. As the other two eigenvalues must be two non-zero numbers that sum to 0, we can conclude one eigenvalue is negative, one is 0, and one is positive. Therefore, the function at the point $(0, y, 0)$ is a **saddle point**.

$$\text{If } \begin{pmatrix} x \\ 0 \\ x \end{pmatrix} \rightarrow H_{g(x,y,z)} = \begin{pmatrix} 0 & 2x & 0 \\ 2x & 0 & 2x \\ 0 & 2x & 0 \end{pmatrix}$$

In this case, the first and third column are identical, meaning one is linearly dependent, therefore we know that $\dim(\text{Ker}(H_g(x, y, z))) = 1$, $\text{Rank}(H_g(x, y, z)) = 2$ and one eigenvalue will be 0. As the other two eigenvalues must be two non-zero numbers that sum to 0, we can conclude one eigenvalue is negative, one is 0, and one is positive. Therefore, the function at the point $(x, 0, x)$ is a **saddle point**.

$$\text{If } \begin{pmatrix} x \\ 0 \\ -x \end{pmatrix} \rightarrow H_{g(x,y,z)} = \begin{pmatrix} 0 & 2x & 0 \\ 2x & 0 & -2x \\ 0 & -2x & 0 \end{pmatrix}$$

As the first and third column are linearly dependent, we know that $\dim(\text{Ker}(H_g(x, y, z))) = 1$, $\text{Rank}(H_g(x, y, z)) = 2$ and one eigenvalue will be 0. As the other two eigenvalues must be two non-zero numbers that sum to 0, we can conclude one eigenvalue is negative, one is 0, and one is positive. Therefore, the function at the point $(x, 0, x)$ is a **saddle point**. The same logic follows for the critical point at $(-x, 0, x)$ as the hessian is nearly identical:

$$\text{If } \begin{pmatrix} x \\ 0 \\ -x \end{pmatrix} \rightarrow H_{g(x,y,z)} = \begin{pmatrix} 0 & -2x & 0 \\ -2x & 0 & 2x \\ 0 & 2x & 0 \end{pmatrix}$$

In short, the four critical points we found for the function were all saddle points.

c) $h : \mathbb{R}^3 \rightarrow \mathbb{R}$ with $h(x, y, z) = x^2 + y^2 + z^2 - 6x + 10y - 2z + 35$

We compute the gradient and then the hessian to determine the behavior of the function. We can take the partial derivatives and calculate the gradient as such:

$$\nabla h(x, y, z) = h'(x, y, z) \rightarrow \frac{\partial h}{\partial x} = 2x - 6 \quad \frac{\partial h}{\partial y} = 2y + 10 \quad \frac{\partial h}{\partial z} = 2z - 2$$

Therefore:

$$\nabla h(x, y, z) = \begin{pmatrix} 2x - 6 \\ 2y + 10 \\ 2z - 2 \end{pmatrix}$$

If we set the gradient to 0, we will need $x = 3, y = -5, z = 1$. We can calculate and then evaluate the hessian at this point to understand the behavior at this critical point.

$$H_{h(x)} = \nabla g'(x) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Since our Hessian is positive definite, which will not change as our inputs x, y, z vary, we can conclude that the point $(3, -5, 1)$ is a local minimum.

2 Problem 10.2

We consider the following constrained optimization problem in \mathbb{R}^2 :

$$\text{minimize } x^2 + y^2 \quad \text{subject to } 2x + y = 4. \quad (1)$$

We admit that this minimization problem has (at least) one solution (this comes from the fact that a continuous function on a compact set attains its minimum).

a) Using Lagrange multipliers, show that (1) has a unique solution and compute its coordinates.

We can use Lagrange multipliers to minimize this function and find the unique solution. Let $g(x, y) = x^2 + y^2$ and $h(x, y) = 2x + y - 4 = 0$. We can create our final function, $f(x)$ by adding the two functions together, with a λ parameter attached to our constraint, $h(x, y)$.

$$\mathcal{L}f(x, y, \lambda) = g(x, y) + \lambda h(x, y)$$

We then take the gradient and set it to 0 which will be useful to identify the critical points of the function $f(x)$. We will also use this to determine λ .

$$\begin{aligned} \nabla \mathcal{L}f(x, y, \lambda) &= \nabla g(x, y) + \nabla \lambda h(x, y) \\ 0 &= \nabla g(x, y) + \nabla \lambda h(x, y) \\ \frac{\partial \mathcal{L}}{\partial x} &= 2x + 2\lambda = 0 \rightarrow x = -\lambda \\ \frac{\partial \mathcal{L}}{\partial y} &= 2y + \lambda = 0 \rightarrow y = -\lambda \frac{1}{2} \\ \frac{\partial \mathcal{L}}{\partial z} &= 2x + y - 4 = 0 \end{aligned} \quad (2)$$

We can now evaluate the constraint function given the information we know about the gradient of our Lagrange for x,y.

$$2x + y - 4 = -2\lambda + -\lambda \frac{1}{2} - 4 \rightarrow \lambda = \frac{-8}{5}$$

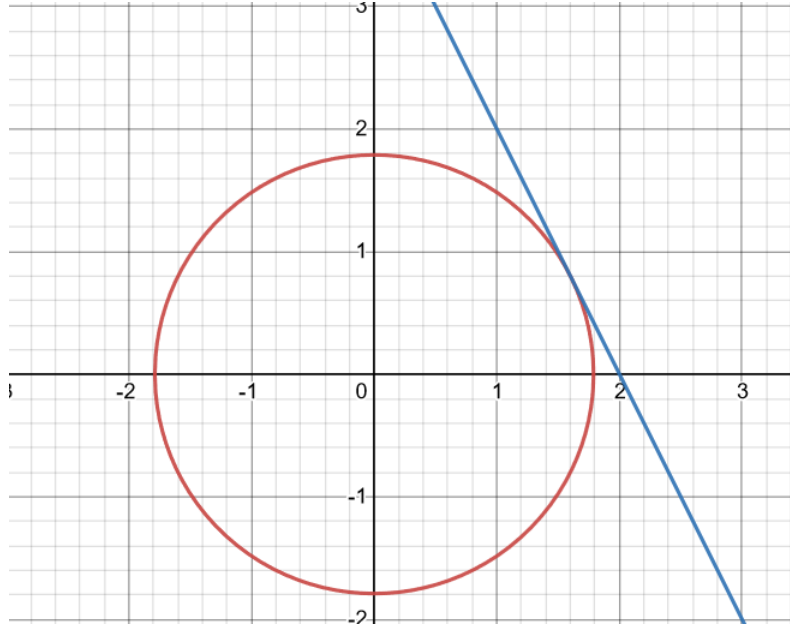
Plugging in λ for to determine our x,y values yields

$$x = \frac{8}{5} \quad y = \frac{4}{5}$$

And this is a unique solution as the coordinates are the only points that satisfy the above system of equations, and they are real numbers, not variables, so they are fixed.

b) Can you draw a picture in \mathbb{R}^2 representing the problem?

We'll represent the problem in \mathbb{R}^2 , which is made easy as our constraint is the function that defines a circle. We plug in the x,y coordinates that we solved for in part a) which yields the following image:



3 Problem 10.3

Let $u \in \mathbb{R}^n$ be a vector such that for all $i \neq j$, $|u_i| \neq |u_j|$. We consider the constrained optimization problem

$$\text{maximize } \langle u, x \rangle \quad \text{subject to } \|x\|_1 \leq 1.$$

a) Calling i_* the index at which $|u_i|$ is maximum, give a solution for the optimization problem (no Lagrange multiplier needed).

Firstly, we need to understand that when we try to maximize the dot product of u and x , we are really trying to maximize the sum of the multiplication of their respective entries, that is to say, we want to maximize: $u_1x_1 + \dots + u_nx_n$. Also, it's important to note that we know that the sum of the absolute values of x are less than or equal to 1, that is to say: $|x_1| \leq 1$ or $|x_1| + \dots + |x_n| \leq 1$. Let's pick the maximum entry of u , and call it u_* which is at entry u_{i_*} . Since we have the l_1 norm constraint, let's define each u_i as $|u_i| = |u_i| - \beta_i$ where $\beta_i \geq 0$. Then we have:

$$\langle u, x \rangle \leq |x_1|(|u_1| - \alpha_1) + \dots + |x_n|(|u_n| - \alpha_n)$$

We can then manipulate the expression in the following way:

$$\begin{aligned}
\langle u, x \rangle &\leq |x_1||u_1| + \dots |x_n||u_n| \\
\langle u, x \rangle &\leq \sum_{i \neq i_*}^n |x_i|(|u_i| - \alpha_i) + (|u_{i_*}| - \alpha_{i_*})|x_{i_*}| \\
\langle u, x \rangle - \sum_{i \neq i_*}^n |x_i|\alpha_i &\leq |u_{i_*}||x_{i_*}|
\end{aligned} \tag{3}$$

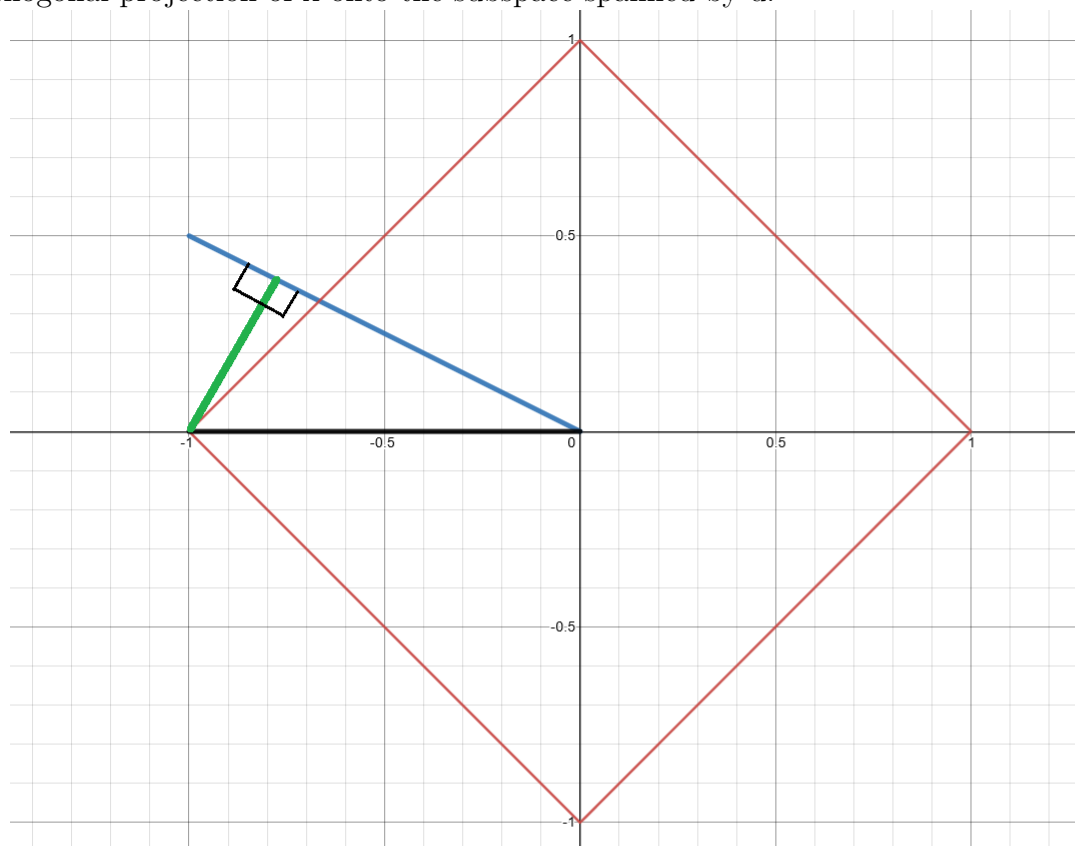
Therefore, the solution to this problem is a vector that contains all 0's except for the entry at the index that corresponds to the largest value of u (u_{i_*}), which will be either -1 or 1 . We will call this solution x_* . This makes sense, as the largest value that an entry can take in x vector is either -1 , or 1 , due to the l_1 norm constraint. Part c illustrates geometrically how this works, and provides great intuition for the solution.

b) By contradiction, show that this solution is unique.

We have found that the solution to our problem is when the vector x is all 0's except for the entry corresponding to the maximum absolute value entry u_{i_*} in the vector (when $x_* = x_{i_*}$). To prove that this is unique, let's assume that $x_* \neq x_{i_*}$ and instead $x_* = x_j$. If $x_* = x_j$ then there is a scenario in which some $i \neq i_*$ then some $x_i \neq 0$ (it would mean the other entries wouldn't have to be minimized). If this is the case, then we can look at our answer in part a and note that in the inequality each term before it is reduced by the corresponding term within our summation. Therefore, it cannot produce the maximum we found in part a, x_* . Therefore, we have shown that the solution in part a is unique, and that x_* is maximized only when $x_* = x_{i_*}$.

c) Give a graphical interpretation in the case $n = 2$. You should consider the orthogonal projector onto $\text{Span}(u)$.

In this image, the red counter line is the l_1 norm, the black line is the vector x such that it maximizes the dot product between x and u , and the blue vector is u . The green vector is the orthogonal projection of x onto the subspace spanned by u .



4 Problem 10.4

We will prove the spectral theorem in this problem: you are therefore not allowed to use the spectral theorem and its consequences to solve this exercise.

Let A be an $n \times n$ symmetric matrix. We consider the following optimization problem

$$\text{maximize } x^T A x \quad \text{subject to } \|x\| = 1. \quad (4)$$

This optimization problem admits a solution (this comes from the fact that a continuous function on a compact set achieved its maximum) that we denote by v_1 .

a) Using Lagrange multipliers, show that v_1 is an eigenvector of A .
Let's set up our Lagrangian function by defining

$$g(x) = x^T A x \text{ and } h(x) = \|x\| = 1 \rightarrow x^T x - 1 = 0$$

But wait! Since we want to maximize $x^T A x$, we can't use the normal procedure. We must first realize that maximizing $x^T A x$ is the same as minimizing $-x^T A x$, so $g(x) = -x^T A x$. Then we have:

$$\mathcal{L}(x, \lambda) = g(x) + \lambda h(x) = -x^T A x + x^T x - 1$$

We can now take the gradient of the function and analyze its critical points:

$$\begin{aligned} \nabla(x, \lambda) &= \nabla g(x) + \nabla \lambda h(x) \\ \frac{\partial g}{\partial x} &= -2Ax \\ \frac{\partial h}{\partial x} &= 2\lambda x \\ \frac{\partial \mathcal{L}}{\partial x} &= -2Ax + 2\lambda x \end{aligned} \quad (5)$$

Setting the gradient of the Lagrange to 0 we have:

$$\begin{aligned} -2Ax + 2\lambda x &= 0 \\ 2Ax &= 2\lambda x \\ Ax &= \lambda x \end{aligned} \quad (6)$$

Which is the exact definition of a matrix transformation (A) acts on an eigenvector (x) with an eigenvalue (λ). Therefore, x is an eigenvector with associated eigenvalue λ .

b) We now consider the optimization problem

$$\text{maximize } x^T A x \quad \text{subject to } \|x\| = 1 \quad \text{and} \quad \langle x, v_1 \rangle = 0. \quad (7)$$

For the same reason as above, this problem admits a solution that we denote by v_2 . Show that v_2 is an eigenvector of A that is orthogonal to v_1 .

We will repeat the same procedure as we did in part a, but this time with two constraints. Lets define the following:

$$g(x) = x^T A x \quad h(x) = x^T x \quad k(x) = x^T v_1$$

Then our Lagrangian function will be the following:

$$\mathcal{L}(x, \lambda_1, \lambda_2) = g(x) + \lambda_1 k(x) + \lambda_2 h(x)$$

We now take the gradient of the Lagrangian and set it to 0.

$$\begin{aligned} \nabla \mathcal{L}(x, \lambda_1, \lambda_2) &= \nabla g(x) + \nabla \lambda_1 k(x) + \nabla \lambda_2 h(x) \\ \frac{\partial g}{\partial x} &= -2Ax \\ \frac{\partial k}{\partial x} &= \lambda_1 v_1 \\ \frac{\partial h}{\partial x} &= 2\lambda_2 x \\ \frac{\partial \mathcal{L}}{\partial x} &= -2Ax + \lambda_1 v_1 + 2\lambda_2 x \end{aligned} \quad (8)$$

We will have to manipulate the expressions using the constraints to tease out the solution:

$$\begin{aligned} 0 &= -2Ax + \lambda_1 v_1 + 2\lambda_2 x \\ 0 &= -2v_1^T A x + \lambda_1 v_1^T v_1 + 2\lambda_2 v_1^T x \\ 0 &= -2v_1^T A x + \lambda_1 + 2 \times 0 \\ \lambda_1 &= 2\langle v_1, Ax \rangle \\ \lambda_1 &= 2\langle Ax, v_1 \rangle \\ \lambda_1 &= 2x^T A^T v_1 \\ \lambda_1 &= 2x^T A v_1 \\ \lambda_1 &= 2\lambda_1 x^T v_1 \\ \lambda_1 &= 2\lambda_1 \times 0 \\ \lambda_1 &= 0 \end{aligned} \quad (9)$$

We arrive at this solution by applying a few tricks. Firstly, we multiplied each side by v_1^T , then we used the fact that $\langle v_1, x \rangle = 0$ and $\|v_1\| = 1$ to simplify. Then we used the fact that the dot product is a linear operation closed under addition, then that A is a symmetric matrix, and then that v_1 is an eigenvector of A . We find that $\lambda_1 = 0$, which we can plug back into our equation to find λ_2 .

$$\begin{aligned}
0 &= -2Ax + \lambda_1 v_1 + 2\lambda_2 x \\
0 &= -2Ax + 2\lambda_2 x \\
Ax &= \lambda_2 x \quad \square
\end{aligned} \tag{10}$$

Which shows that x is another eigenvector of A , which is orthogonal to v_1 and has the associated eigenvalue λ_2 .

c) We now consider the optimization problem

$$\text{maximize } x^T A x \quad \text{subject to } \|x\| = 1 \quad \text{and} \quad \langle x, v_1 \rangle = 0 \quad \text{and} \quad \langle x, v_2 \rangle = 0. \tag{11}$$

Again, this problem admits a solution that we denote by v_3 . Show that v_3 is an eigenvector of A that is orthogonal to v_1 and v_2 .

Conclusion: by repeating this procedure, we obtain an orthonormal family v_1, \dots, v_n of eigenvectors of A . This proves the spectral theorem (without using any linear algebra result!).

We repeat the same procedure that we did in b, but now with 3 constraints. Lets call the following:

$$g(x) = x^T A x \quad h(x) = x^T x - 1 \quad k(x) = x^T v_1 \quad j(x) = x^T v_2$$

We now take the Lagrange:

$$\begin{aligned}
\nabla \mathcal{L}(x, v_1, v_2, \lambda_1, \lambda_2, \lambda_3) &= \nabla g(x) + \nabla \lambda_1 k(x) + \nabla \lambda_2 j(x) + \nabla \lambda_3 h(x) \\
\frac{\partial g}{\partial x} &= -2Ax \\
\frac{\partial k}{\partial x} &= \lambda_1 v_1 \\
\frac{\partial j}{\partial x} &= \lambda_2 v_2 \\
\frac{\partial h}{\partial x} &= 2\lambda_3 x \\
\frac{\partial \mathcal{L}}{\partial x} &= -2Ax + \lambda_1 v_1 + \lambda_2 v_2 + 2\lambda_3 x
\end{aligned} \tag{12}$$

Now we set the gradient to 0 and solve for the lambdas:

$$\begin{aligned}
0 &= -2Ax + \lambda_1 v_1 + \lambda_2 v_2 + 2\lambda_3 x \\
0 &= -2v_1^T A x + \lambda_1 v_1^T v_1 + \lambda_2 v_1^T v_2 + 2\lambda_3 v_1^T x \\
0 &= -2v_1^T A x + \lambda_1 \\
\lambda_1 &= 0
\end{aligned} \tag{13}$$

And we have found that $\lambda_1 = 0$. I didn't write some of the intermediate steps, as they are just repeats of the procedure we defined in b, in which we can manipulate the expression such that the orthogonal vectors equal 0 via dot product. Now we repeated the process for λ_2

$$\begin{aligned}
0 &= -2Ax + \lambda_1 v_1 + \lambda_2 v_2 + 2\lambda_3 x \\
0 &= -2v_2^T Ax + \lambda_1 v_2^T v_1 + \lambda_2 v_2^T v_2 + 2\lambda_3 v_2^T x \\
0 &= -2v_2^T Ax + \lambda_2 \\
\lambda_2 &= 0
\end{aligned} \tag{14}$$

And using the same procedure we have found that $\lambda_2 = 0$. We can now plug in the values for the first two lambdas to find the final one.

$$\begin{aligned}
0 &= -2Ax + \lambda_1 v_1 + \lambda_2 v_2 + 2\lambda_3 x \\
0 &= -2Ax + 2\lambda_3 x \\
Ax &= \lambda_3 x \quad \square
\end{aligned} \tag{15}$$

And here we have shown that x is an eigenvector of A with associated eigenvalue λ_3 and is orthogonal to v_1, v_2 .

5 Problem 10.5

We consider the problem with physics motivation of finding the maximal entropy distribution of a random variable (see last slides of Lecture 09) constraining values of some moments.

To keep things simple, we consider X that can take n different values x_1, \dots, x_n in \mathbb{R} . We wish to infer the probabilities p_1, \dots, p_n such that the entropy is maximal and the expected value of X is equal to a previously known scalar $\mu \in \mathbb{R}$. This corresponds to solving the constrained optimization problem

$$\text{maximize} \quad - \sum_i p_i \ln p_i \quad \text{subject to} \quad p_i \geq 0 \text{ for all } i \quad \text{and} \quad \sum_{i=1}^n p_i = 1 \quad \text{and} \quad \sum_{i=1}^n p_i x_i = \mu. \quad (16)$$

a) Rewrite the problem as a convex minimization problem (justify).

We can note that maximizing $-\sum_i p_i \ln p_i$ is the same as minimizing $\sum_i p_i \ln p_i$. Given that, we can rewrite the problem statement such that it is a convex minimization problem:

$$\text{Minimize} \quad \sum_i p_i \ln p_i \quad \text{subject to:} \quad p_i \geq 0 \text{ for all } i \quad \text{and} \quad \sum_{i=1}^n p_i = 1 \quad \text{and} \quad \sum_{i=1}^n p_i x_i = \mu$$

We can do so with Lagrange minimization technique that we have used in the above problems. Firstly, lets pack our p and x values into vectors in \mathbb{R}^n such that $p \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$ where $p_i = p_i$ and $x_i = x_i$. Also, let $\mathbf{1}$ refer to the a vector in \mathbb{R}^n where all the entries are equal to 1. Also, let $l \in \mathbb{R}^n$ where $l_i = \ln(p_i)$. Then lets note the following:

$$\begin{aligned} \sum_i p_i \ln p_i &\rightarrow \langle p_i, l_i \rangle \\ \text{where } l &\in \mathbb{R}^n \text{ and } l_i = \ln(p_i) \\ \sum_{i=1}^n p_i = 1 &\rightarrow \sum_{i=1}^n p_i - 1 = 0 \rightarrow \langle p_i, \mathbf{1} \rangle - 1 = 0 \\ \text{where } \mathbf{1} &\text{ refers a vector in } \mathbb{R}^n \text{ with all of its values equal to 1} \\ \sum_{i=1}^n p_i x_i = \mu &\rightarrow \sum_{i=1}^n p_i x_i - \mu = 0 \rightarrow \langle p_i, x_i \rangle - \mu = 0 \end{aligned} \quad (17)$$

We can then use the Lagrange function to find the solution to our constrained minimization problem (which will be the solution to the maximization problem in the problem statement as we flipped the sign of the original function). Our constrained optimization function now takes the form:

$$\mathcal{L}(p, x, \mu, \lambda_1, \lambda_2) = \langle p_i, l_i \rangle + \lambda_1(\langle p_i, \mathbf{1} \rangle - 1) + \lambda_2(\langle p_i, x_i \rangle - \mu)$$

b) Using KKT theorem, give the expression of the probability vector solution $p \in \mathbb{R}^n$ as a function of Lagrange multipliers and values x_i . Give also the relations between the Lagrange multipliers, μ and values x_i .

We can give the expression of the probability vector solutions as a function of the Lagrange multipliers by taking the gradient of the lagrange function and setting it equal to 0, then solving for the system of equations that results. As we have:

$$\mathcal{L}(p, x, \mu, \lambda_1, \lambda_2) = \langle p_i, l_i \rangle + \lambda_1(\langle p_i, 1 \rangle - 1) + \lambda_2(\langle p_i, x_i \rangle - \mu)$$

Then:

$$\begin{aligned} \nabla \mathcal{L}(p, l, x, \mu, \lambda_1, \lambda_2) &= \mathcal{L}'(p, l, x, \mu, \lambda_1, \lambda_2) \\ \frac{\partial \mathcal{L}}{\partial p} &= l_i + \lambda_1 + \lambda_2 x_i \\ \frac{\partial \mathcal{L}}{\partial x} &= \lambda_2 p_i \\ \frac{\partial \mathcal{L}}{\partial \mu} &= 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_1} &= \langle p_i, 1 \rangle - 1 \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= \langle p_i, x_i \rangle - \mu \end{aligned} \tag{18}$$

Setting the gradient equal to 0, we can see that:

$$-\frac{l_i + \lambda_1}{x_i} = \lambda_2 \text{ and that } \lambda_2 p_i = 0$$

. We can use this information to find the critical points of our function.

(19)

c) In the case where $n = 2$ and $x_1 = 0$ and $x_2 = 1$, solve for the values of the Lagrange multipliers and $p \in \mathbb{R}^2$. Could you have used an easier way to solve the problem in this simple case?