Linear Algebra HW 9

gjd9961

November 2021

1 Problem 9.1

a) Show that $\mathcal{M}$ is a convex set.

Since f(x) is convex we know that the inequality

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)y$$

must hold when $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ and $0 \leq \alpha \leq 1$. Lets set x, y to two distinct elements of M such that $x \neq y$ and $f(x) = f(y) = m$ where m is the minimum. Lets see if our inequality holds

$$
\begin{aligned}
f(\alpha x + (1 - \alpha)y) &\leq \alpha f(x) + (1 - \alpha)f(y) \\
f(\alpha x + (1 - \alpha)y) &\leq \alpha \times m + m - \alpha \times m \\
f(\alpha x + (1 - \alpha)y) &\leq m
\end{aligned}
\tag{1}
$$

Since m is the minimizer of f, we have $f(\alpha x + (1 - \alpha)y) = m$. Since a set is convex if $\alpha x + (1 - \alpha)y \in M$, and because $f(\alpha x + (1 - \alpha)y) = m$ then $f(\alpha x + (1 - \alpha)y) \in M$ where $f(x) = f(y)$. Therefore M is convex $\square$

b) Show that if f is strictly convex, then $\mathcal{M}$ has only one element.
With $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ and $0 \leq \alpha \leq 1$ we can use the definition of strict convexity:

$$
\begin{aligned}
f(\alpha x + (1 - \alpha)y) &< \alpha f(x) + (1 - \alpha)f(y) \\
f(\alpha x + (1 - \alpha)y) &< \alpha \times m + m - \alpha \times m \\
f(\alpha x + (1 - \alpha)y) &< m
\end{aligned}
\tag{2}
$$

We know that m is the minimizer of f but somehow, using two vectors, we must have a value smaller than m. This is a contradiction, and we have shown by definition of strict convexity, we cannot have two points $x, y \in M$ that both minimize the function f, as it should only have a single minimizer. So if f is strictly convex, M must have only one element.

2 Problem 9.2

Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. For $x \in \mathbb{R}^n$ we define

$$f(x) = x^T M x + \langle x, b \rangle + c.$$

f is called a quadratic function.

a) Compute the gradient $\nabla f(x)$ and the Hessian $H_f(x)$ at all $x \in \mathbb{R}^n$ (show the steps of your computations). Show that f is convex if and only if M is positive semi-definite.

For the gradient: we must differentiate f(x). The constant term c will become 0 when we differentiate, while the term $\langle x, b \rangle$ will be the following:

$$\langle x, b \rangle = \sum_{i=1}^{n} x_i \times b_i \longrightarrow \frac{df}{\partial x_i} = b_i$$

While the term $x^T M x$ can be expressed as $\langle x, Mx \rangle$

$$\langle x, Mx \rangle = \sum_{j=1}^{n} (\sum_{i=1}^{n} m_i \times x_i) x_i = \sum_{i=1}^{n} (m_{n,1} x_1 + \cdots + m_{n,n} x_n) x_i \tag{3}$$

So when we diffrerentiate with respect to an x_i we have

$$\frac{df}{\partial x_i} = (2m_{1,1} x_1 + \cdots + m_{n,n} x_n) + m_{2,1} x_2 + \cdots + m_{n,n} x_n = 2M_i x$$

Where M_i indicates the row of the matrix M.
Putting it all together:

$$\nabla f(x) = f'(x) = \begin{pmatrix} 2M_1 x + b_1 \\ \vdots \\ 2M_n x + b_n \end{pmatrix}$$

For the Hessian: The only term of $f(x)$ that we will care about is $x^T M x$ as $\langle x, b \rangle$ and c will both be equal to 0 once we double differentiate. We can express $x^T M x$ in the following way:

$$x^T M x = (x_1, x_2, \ldots, x_n) \begin{pmatrix} -- & m_1 & -- \\ -- & \ldots & -- \\ -- & m_n & -- \end{pmatrix} \begin{pmatrix} x_1 \\ | \\ x_n \end{pmatrix} = \sum_{i=1}^{n} (\sum_{j=1}^{n} m_{i,j} x_j) x_i$$

We can expand the summations and see that we get a sum like:

$$x^T M x = \sum_{i=1}^{n} (m_{1,1} x_1 + \cdots + m_{n,n} x_n) x_i$$

We can then begin to play around with differentiation and see what occurs to this summation. There are only two situations we must consider: when we differentiate with respect to

2

the same variable twice, and the situation in which we differentiate twice with respect to two different variables (one differentiation per variable).

If we differentiate with respect to the same variable twice (this will only happen on the diagonal of the hessian) then we get $H_{i,i} = 2m_{i,i}$ and when we differentiate twice, once for two variables, then we get $H_{i,j} = m_{i,j} + m_{j,i} = 2m_{i,j}$ where $m_{i,j} = mj, i$ as the matrix M is symmetric. Therefore what we have found is that the hessian matrix is twice the matrix M:

$$H_{f(x)} = 2M$$

Since M is a square, symmetric matrix, it is positive semi definite. $H_f(x)_{i,j} = 2Mi, j$ therefore M must be positive semi definite.

To show if M is positive semi definite then $f(x)$ is convex: $H_f(x)i, j = 2M_{i,j}$ M is positive semi definite then $H_f(x)$ is positive semi definite. Therefore, f is convex.

b) In this question, we assume M to be positive semi-definite. Show that f admits a minimizer if and only if $b \in Im(M)$.

We can show this using a bi-directional proof. Firstly, if $b \in Im(M)$ then $f(x)$ admits a minimizer. That means for some $y \in \mathbb{R}^n$, $My = b$.

$$
\begin{aligned}
f(x) &= x^T M x + \langle x, My \rangle + c \\
f(X) &= x^M x + x^T My + c \\
f(x) &= x^T M(x + y) + c \\
f(x) &= x^T 0 + c \\
f(x) &= 0 + c = c \\
f'(x) &= \nabla f(x) = 0
\end{aligned}
\tag{4}
$$

Now for the other direction: assuming that $f(x)$ admits a minimizer, then $b \in Im(M)$ Let $My = b$ and $x = -\frac{y}{2}$ then:

$$
\begin{aligned}
\nabla f(x) &= 2Mx + b = 2Mx + My \\
\nabla f(x) &= 2M(-y/2) + My = 0 \quad \square
\end{aligned}
\tag{5}
$$

Therefore f admits a minimizer.

3 Problem 9.3

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex if there exists $\alpha > 0$ such that the function $x \mapsto f(x) - \frac{\alpha}{2}\|x\|^2$ is convex. In other words, f is strongly convex if there exists $\alpha > 0$ and a convex function $g : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x) = g(x) + \frac{\alpha}{2}\|x\|^2.$$

a) Show that a strongly convex function is strictly convex. (Hint: start by showing that $x \mapsto \|x\|^2$ is strictly convex).
First lets start by showing that $x \mapsto \|x\|^2$ is strictly convex.

$$j(x) = \|x\|^2 = \langle x, x \rangle = \sum_{i=1}^{n} x_i^2$$

We can compute the hessian of $j(x)$, $H_j(x)$ and we find that for the diagonals we have value 2, and 0 elsewhere.

$$\frac{\partial j}{\partial x_i, \partial x_k} = \begin{cases} 0 & \text{when } i \neq k \\ 2 & \text{when } i = k \end{cases} \longrightarrow H_{j(x)} = 2 \times Id_n$$

As the hessian is just twice the identity matrix, and the identity matrix has strictly positive eigenvalues, so does the hessian matrix of $j(x) = \|x\|^2$ and therefore it is strictly convex. Since we're considering $\frac{\alpha}{2}\|x\|^2$, then the Hessian for that part of the equation is $H(x) = \alpha \times Id_n$ and therefore, the hessian is Positive Definite, as it has eigenvalues grater than 0. We know this because the identity has eigenvalues of 1, and since we scaled it by α the eigenvalues of the hessian are α where $\alpha > 0$. Now we must show that $f(x)$ is strictly convex. Lets call the right portion of the equation k where $k = \frac{\alpha}{2} \times \|x\|^2$

$$H_f(x) = H_g(x) + H_k(x)$$

Since we know that $H_k(x)$ is positive definite and strictly convex, and we know $g(x)$ is convex, it follows that $H_f(x)$ is positive definite with eigenvalues strictly greater than 0. We can show this further with the following. Consider $x, y \in \mathbb{R}^n$, $f(x) = g(x) + k(x)$ and $0 \leq \alpha \leq 1$

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$$
$$g(\alpha x + (1 - \alpha)y) + k(\alpha x + (1 - \alpha)y) < \alpha g(x) + (1 - \alpha)g(y) + \alpha k(x) + (1 - \alpha)k(y) \tag{6}$$

Lets call $a_1 = g(\alpha x + (1-\alpha)y)$ and $b_1 = k(\alpha x + (1-\alpha)y)$ and lets say $a_2 = \alpha g(x) + (1-\alpha)g(y)$ and $b_2 = \alpha k(x) + (1-\alpha)k(y)$. Becasue $g(x)$ is convex we know that $a_1 \leq a_2$, and since $k(x)$ is strictly convex then $b_1 < b_2$. It follows that $a_1 + b_1 < a_2 + b_2$ for any $x, y \in \mathbb{R}^n$ and therefore $f(x)$ is strictly convex.

b) Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Show that φ is strongly convex if and only if there exists $\alpha > 0$ such that for all $x \in \mathbb{R}^n$ the eigenvalues of $H_\varphi(x)$ are greater or equal than α.

Lets first prove this property in one direction, showing that a function is strongly convex if $h(x) - \frac{\alpha}{2}||x||^2$ is convex where $\alpha > 0$ and that the eigenvalues of $H_h(x)$ are greater to or equal to α. To show this let $z(x) = h(x) - \frac{\alpha}{2}||x||^2$. The hessian of z would be the following $H_z(x) = H_h(x) - \alpha \times Id_n$ (we computed this in part a). It follows that $\lambda_{z1}, \dots, \lambda_{zn} = \lambda_{h1} - \alpha, \dots, \lambda_{hn} - \alpha$, that is to say, the eigenvalues of the hessian of $h(x)$ are equal to the eigenvalues of the hessian of $z(x)$ minus α. Since $\lambda_{h1}, \dots, \lambda_{hn} \geq \alpha$ by assumption the lowest $\lambda_{z1}, \dots, \lambda_{zn}$ can be is 0, and therefore, $\lambda_{z1}, \dots, \lambda_{zn} \geq 0$ and $H_z(x)$ is positive semi definite, thus proving $z(x)$ is convex. If $z(x)$ is convex, $h(x)$ is strongly convex, by definition and virtue of the problem statement.

Now for the other direction: we assume $h(x)$ is strongly convex it follows that $\lambda_{h1}, \dots, \lambda_{hn} \geq \alpha$. We define $h(x) = g(x) + \frac{\alpha}{2}||x||^2$ where $g(x)$ is a concave function and $\alpha > 0$

$$H_h(x) = H_g(x) + \alpha Id_n \text{ by problem 3.a}$$

It follows that $\lambda_{h1}, \dots, \lambda_{hn} = \lambda_{g1} + \alpha, \dots, \lambda_{gn} + \alpha$ where $\lambda_{g1}, \dots, \lambda_{gn} \geq 0$. Therefore, at minimum, the values of $\lambda_{g1}, \dots, \lambda_{gn} \geq 0$ and it follows that $\lambda_{h1}, \dots, \lambda_{hn} \geq \alpha$ $\square$

4   Problem 9.4

Let $A \in \mathbb{R}^{n \times m}$ and $y \in \mathbb{R}^n$. For $x \in \mathbb{R}^m$ we define

$$f(x) = ||Ax - y||^2.$$

a) Compute the gradient $\nabla f(x)$ and the Hessian $H_f(x)$ at all $x \in \mathbb{R}^m$. Show that f is convex. To compute the gradient well need to take the first order derivatives to get our gradient vector. Since we have the expression $Ax - y$ nested inside the euclidian norm squared, well have to use the chain rule. Let $h(x) = Ax - y$ and $g(y) = ||y||^2$, lastly it follows that $f(x) = g(h(x))$

Calculating $g'(y)$ is fairly straight forward as $g(y) = \langle y, y \rangle = \sum_{i=1}^{n} y_i^2$ and therefore $\frac{dg}{\partial y_i} = 2y_i$

Similarly, when calculating $h'(x)$ we know that $h(x)_i = Ax - y = \langle x, a_i \rangle - y_i$ where a_i is the ith row of A and the corresponding entry of the y vector. Therefore:

$$h(x) = \begin{pmatrix} (\sum_{i=1}^{n} a_i \times x_i) - y_1 \\ \vdots \\ (\sum_{i=1}^{n} a_i \times x_i) - y_n \end{pmatrix} \text{ and } \nabla h(x) = h'(x) = \begin{pmatrix} a_i \\ \vdots \\ a_n \end{pmatrix} = A$$

And when we differentiate $h(x)$ with respect to any x_i we get $\frac{dh}{\partial x_i} = a_i$.
We have found that $g'(y) = 2y$ and that $h'(x) = A$, and using the chain rule $f'(x) = g'(h(x))h'(x) = 2(Ax - y)^T A$ and therefore $\nabla f(x) = f'(x) = 2A^T(Ax - y)$. If we differentiate once more to get the hessian we find that $H_f(x) = \frac{d\nabla f(x)}{dx} = 2A^T A$. Since $A^T A$ is square symmetric, $A^T A$ is positive semi definite. Therefore, $H_f(x)$ is positive semi definite and therefore $f(x)$ is convex.

b) Show that if $rank(A) < m$, then f is not strictly convex.

If $rank(A) < m$ we know by rank nullity theorem that $Ker(A) > 0$. For a function to be strictly convex, we need the following inequality to hold:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

where $0 \leq \alpha \leq 1$. Lets choose two elements from the kernel of A, as we are assuming $Rank(A) < m$ and therefore $dim(ker(A)) > 0$. Let $x, y \in Ker(A)$ and $x \neq y$ and lets take some $b \in \mathbb{R}^n$

$$
\begin{aligned}
||\alpha Ax + (1 - \alpha)Ay - b||^2 &< \alpha x||Ax - b||^2 + (1 - \alpha)||Ay - b||^2 \\
||\alpha \times 0 + (1 - \alpha) \times 0 - b||^2 &< \alpha||0 - b||^2 + (1 - \alpha)|| - b||^2 \\
|| - b||^2 &< \alpha||b^2|| - \alpha||b||^2 + ||b||^2 \\
|| - b||^2 &< || - b||^2 \text{ this is a contradiction} \quad \square
\end{aligned}
\tag{7}
$$

We have shown that if you allow the matrix A to have a kernel, it cannot possibly be strictly convex, as we will have elements in the kernel that contradict the definition of strict convexity.

c) Show that if $rank(A) = m$, then f is strongly convex.

If $Rank(A) = m$, this has interesting implications on what we calculated from part a, that the hessian of $h(x)$, $H_h(x) = 2A^T A$. If its the case that $Rank(A) = m$ we know from past home works and proofs that $A^T A$ will be positive definite, as $Ker(A^T A) = Ker(A)$ ($x \in Ker(A)$ then $A^T Ax = 0$ and $x^t A^T Ax = (Ax)^T(Ax) = 0$).

Since the dimension of the kernel of $A^T A$ is equal to 0, then all of the eigenvalues of the hessian of $H_h(x)$ must be greater than 0. If $\lambda_1, \ldots, \lambda_n > 0$ for all λ belonging to $H_h(x)$ then $f(x)$ is strongly convex.

5 Problem 9.5

Is the following function f convex? Compute its Hessian to conclude.

$$
\begin{aligned}
f: \quad \mathbb{R}^n &\to \mathbb{R} \\
x &\mapsto f(x) = \ln(e^{x_1} + \cdots + e^{x_n})
\end{aligned}
$$

We can re express our function f in the following way:

$$f(x) = \ln(e^{x_1} + \cdots + e^{x_n}) = ln(\sum_{i=1}^{n} z_i) = ln(I^T z)$$

where $z_i = e^{x_i}$ and $I \in R^n$ and $I = 1, \ldots, 1$ Using the chain rule we can differentiate the function, with the outside function being $ln(x)$ and the inside function being $I^T z$

$$\frac{df}{\partial x_i} = \frac{d}{\partial x_i} ln(1^T z) = \frac{1}{1^T z} z_i$$

Now that we've computed the gradient, we can compute the Hessian. We have two scenarious we must check: when we're on the diagonal of the hessian and when were not. That is when $i = j$ and when $i \neq j$

$$H_f(x)_{i,j} \text{ where } i = j \text{ then } = \frac{z_i}{I^T z} - \frac{z_i z_j}{(I^T z)^2}$$

and

$$H_f(x)_{i,j} \text{ where } i \neq j \text{ then } = -\frac{z_i z_j}{(I^T z)^2}$$

We can verity its convex in the following way:

$$V^T \nabla^2 f(x) v \geq 0$$
$$(v^T (I^T z) diag(z) v - v^T z z^T v) \geq 0$$
$$\sum_{i=1}^{n} z_i \times \sum_{i=1}^{n} v_i^2 z_i \geq (\sum_{i=1}^{n} z_i v_i)^2$$
$$\sum_{i=1}^{n} \sqrt{z_i} \sqrt{z_i} \sum_{i=1}^{n} v_i \sqrt{z_i} v_i \sqrt{z_i} \geq (\sum_{i=1}^{n} \sqrt{z_i} \sqrt{z_i} v_i)^2$$
$$\text{set } b = \sqrt{z_i} \text{ and set } a = v_i \sqrt{z_i} \text{ then}$$
$$(b^T b)(a^T a) \geq (b^T a)^2 \text{ True by Cauchy Shwarz} \quad \square$$

$$(8)$$

Since $H_f(x)$ is Positive Semi Definite $f(x)$ is convex.