164 Midterm 1 Solutions[1]

1. Question 1

True/False

(i) Let n be a positive integer. Every $n \times n$ real matrix has at least one real eigenvalue.

FALSE. The eigenvalues λ of the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ satisfy $\lambda^2 + 1 = 0$, so this matrix has no real eigenvalues.

(ii) Let A be an 5×5 real symmetric positive semidefinite matrix. Then all eigenvalues of A are positive.

FALSE. The zero matrix is positive semidefinite, but all of its eigenvalues are zero.

(iii) Let A be a 5×5 real symmetric matrix. Let $x \in \mathbb{R}^5$. As usual, define $\|x\| = (x^T x)^{1/2}$. Assume that, for any $x \in \mathbb{R}^5$, we have $\lim_{n\to\infty} \|A^n x\| = 0$. Then any eigenvalue λ of A must satisfy $|\lambda| < 1$.

TRUE. If A has an eigenvalue λ with $|\lambda| \geq 1$ with eigenvector $x \in \mathbb{R}^5$, $x \neq 0$, then $\lambda \in \mathbb{R}$ by the Spectral Theorem (Theorem 2.22 in the notes), and $A^n x = \lambda^n x$, so $\|A^n x\| = \|\lambda^n x\| = |\lambda|^n \|x\|$

(iv) The union of two convex sets is convex.

FALSE. $[-1, 0]$ and $[1, 2]$ are convex, but $[-1, 0] \cup [1, 2]$ is not convex, since $1/2 = (1/2)(0) + (1/2)(1)$, $0 \in [-1, 0]$ and $1 \in [1, 2]$, but $1/2 \notin [-1, 0] \cup [1, 2]$.

(v) Let $f, g \colon \mathbb{R} \to \mathbb{R}$ be two convex functions. Let $a, b \in \mathbb{R}$ be real numbers. Then $af + bg$ is a convex function.

FALSE. For any $x \in \mathbb{R}$, let $f(x) = x^2$, $g(x) = 0$, $a = -1$, $b = 0$. Then $af(x) = -x^2$ which is not convex, since $(1/2)(-(-1)^2) + (1/2)(-(1)^2) = -1 < 0 = -((1/2)(-1) + (1/2)(1))^2$.

2. Question 2

Let $f \colon \mathbb{R} \to \mathbb{R}$ be a function with three continuous derivatives. Assume that $f''(x) \geq 0$ for all $x \in \mathbb{R}$. Show that f is convex. (You can freely use Taylor's Theorem with integral remainder.)

Solution. From Taylor's Theorem, for any $x, y \in \mathbb{R}$, we have

$$f(x) = f(y) + (x - y)f'(y) + \int_y^x f''(t)(x - t)dt \geq f(y) + (x - y)f'(y). \qquad (*)$$

Let $a, b \in \mathbb{R}$ and let $t \in (0, 1)$. Set $y := ta + (1 - t)b$. From $(*)$, we deduce

$$tf(a) \geq tf(y) + t(a - y)f'(y), \qquad (1 - t)f(b) \geq (1 - t)f(y) + (1 - t)(b - y)f'(y)$$

Note that $t(a - y) + (1 - t)(b - y) = ta + (1 - t)b - y = 0$ by definition of y. So, adding the inequalities,

$$tf(a) + (1 - t)f(b) \geq f(y) + f'(y)[t(a - y) + (1 - t)(b - y)] = f(y) = f(ta + (1 - t)b).$$

[1]December 7, 2016,

3. Question 3

Maximize $f(x,y) = x^2 + y^2$ subject to the constraint $x^2 + 3y^2 = 1$.

Solution. Let $g(x,y) = x^2 + 3y^2 - 1$. We solve $\nabla f(x,y) = \lambda \nabla g(x,y)$. That is, we solve $(2x, 2y) = \lambda(2x, 6y)$. If $y \neq 0$, then $2 = 6\lambda$, so $\lambda = 1/3$, and $2x = 2\lambda x = (2/3)x$, so $x = 0$. Since $g(x,y) = 0$, we conclude that $y^2 = 1/3$, so that $y = \pm 1/\sqrt{3}$. So far, we have found candidate critical points $(0, 1/\sqrt{3})$ and $(0, -1/\sqrt{3})$. It remains to consider $y = 0$. In the case $y = 0$, since $g(x,y) = 0$, we have $x^2 = 1$. So, the only remaining candidate critical points are $(1,0)$ and $(-1,0)$.

In summary, by Lagrange Multipliers (Proposition 2.57 from the notes), the maximum of f must be among the four points: $(0, 1/\sqrt{3})$, $(0, -1/\sqrt{3})$, $(1,0)$ and $(-1,0)$. (Note that $\nabla g(x,y) = 0$ only when $x = y = 0$, so that $\nabla g(x,y) \neq 0$ for all (x,y) such that $g(x,y) = 0$, so Proposition 2.57 applies.) Plugging these four points into f, we find they have values $1/3, 1/3, 1, 1$, respectively. So, the maximum of f subject to the constraint $x^2 + 3y^2 = 1$ occurs at the points $(1,0)$ and $(-1,0)$, and the maximum value is 1.

4. Question 4

Let $f(x,y) = -x^2 - y^2$ for any $x, y \in \mathbb{R}$. In order to maximize f, find the first two iterations of the gradient ascent algorithm with parameter $\varepsilon = 1$, starting from the point $(x_0, y_0) = (2,2)$. Your answer should include two points (x_1, y_1) and (x_2, y_2) in the plane that are found next in the algorithm. (If you cannot remember where the parameter ε arises in the algorithm, just try to implement the gradient ascent algorithm as best you can, using two steps, and you will get partial credit.)

Solution. Note that $\nabla f(x,y) = -2(x,y)$. Gradient ascent says $x^{(n)} = x^{(n-1)} + \varepsilon \nabla f(x^{(n-1)})$. We therefore get

$$(x_1, y_1) = (x_0, y_0) + \nabla f(x_0, y_0) = (x_0, y_0) - 2(x_0, y_0) = -(x_0, y_0) = (-2, -2).$$
$$(x_2, y_2) = (x_1, y_1) + \nabla f(x_1, y_1) = (x_1, y_1) - 2(x_1, y_1) = -(x_1, y_1) = (2, 2).$$

5. Question 5

Suppose you enter the following expression into a computer program, such as Matlab, or a scientific calculator.

$$((1 + 2^{-53}) - 1)2^{53}.$$

Give a detailed explanation of what output this expression produces, and why this output is produced.

Solution. The output will be 0, even though we know it should be 1. The reason for this is that Matlab uses double-precision floating-point arithmetic. In double-precision floating-point arithmetic, which is the standard way to represent numbers in computers, a nonzero real number is stored as 1 followed by 52 binary digits after the decimal point, with an 11-digit binary exponent. That is, a nonzero real number on a computer is stored in the form

$$\pm 1.b_1 b_2 \ldots b_{52} \times 2^{c_1 c_2 \ldots c_{11} - 2^{10} + 1},$$

where $b_1, \ldots, b_{52}, c_1, \ldots, c_{11} \in \{0, 1\}$, and we interpret the decimal and the exponent as binary numbers. (Even though a number is stored using 64 bits, and we only described using 63 bits, there is one bit that is used to store the $\pm$ sign of the number.)

For example, when Matlab computes $1 + 2^{-53}$, it is performing the following addition

$$\Big(1.0\ldots0 \times 2^0\Big) + \Big(1.0\ldots0 \times 2^{-53}\Big).$$

Now, in order to add the numbers, the computer tries to represent the smaller number so that its exponent is 2^0, matching the larger number's exponent. But since only 52 binary digits of the number 2^{-53} are stored, the addition becomes

$$\Big(1.0\ldots0 \times 2^0\Big) + \Big(0.0\ldots0 \times 2^0\Big) = 1.$$

Finally, subtracting 1 from this expression gives the result of 0 for the expression $(1+2^{-53})-1$.