164 Final Solutions[1]

1. Question 1

True/False

(a) Let $f\colon \mathbb{R}\to\mathbb{R}$ be a C^3 function such that $f(x)\geq 0$ for all $x\in\mathbb{R}$. Then the gradient descent algorithm started at the point 0 will find the global minimum of f.

FALSE. The gradient descent algorithm may only terminate at a local minimum of a function. Let $x\in\mathbb{R}$ and consider $f(x)=(x-1)^2(x+1)^2$. Note that $f(x)\geq 0$ for all $x\in\mathbb{R}$. Then $f'(x)=(x-1)^2 2(x+1)+(x+1)^2 2(x-1)=2(x+1)(x-1)(x+1+x-1)=4x(x+1)(x-1)$. So, $f'(0)=0$, and the gradient descent algorithm cannot move anywhere, since $x_0=0$, and $x_1=x_0+\varepsilon f'(0)=x_0$, for any $\varepsilon>0$, and more generally $x_n=x_0$ for any $n\geq 0$. So, $f(x_n)=1$ for all $n\geq 0$, but $f(1)=0$. That is, the algorithm has not found the global minimum of f.

(b) Let A be a 5×5 real symmetric matrix. Let $x\in\mathbb{R}^5$. As usual, define $\|x\|=(x^Tx)^{1/2}$. Assume that, for any $x\in\mathbb{R}^5$, we have $\lim_{n\to\infty}\|A^nx\|=0$. 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 $A^nx=\lambda^nx$, so $\|A^nx\|=\|\lambda^nx\|=|\lambda|^n\|x\|$.

(c) Let $n\geq 2$ be an integer. Let $f\colon\mathbb{R}^n\to\mathbb{R}$ be a function. Assume that $f\in C^3$. Fix $x\in\mathbb{R}^n$. As usual, define the matrix of second derivatives $D^2f(x)$ so that $(D^2f(x))_{ij}=\frac{\partial^2 f(x)}{\partial x_i\partial x_j}$ for any $1\leq i,j\leq n$. Assume that $\nabla f(x)=0$ and all eigenvalues of $D^2f(x)$ are nonnegative. Then x is a local minimum of f.

FALSE. Let $f(x,y)=x^4-y^4$. Then $\nabla f(0,0)=0$, $D^2f(0)=0$, so all eigenvalues of $D^2f(0)$ are nonnegative, but $f(0,0)=0$ while $f(0,t)<0$ for all $t\neq 0$. Therefore, $(0,0)$ is not a local minimum of f.

(d) Let A be a 2×2 real matrix. For any $x,y\in\mathbb{R}^2$, define $\langle x,y\rangle_A:=x^TAy$. Then for any $x\in\mathbb{R}^2$, $\langle x,x\rangle_A\geq 0$.

FALSE. Let $A=\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}$. Let $x=\binom{1}{0}$. Then $x^TAx=-1<0$.

(e) Suppose we have a primal linear program which is infeasible. Then the dual linear program is unbounded.

FALSE; it is possible for both the primal and dual to be infeasible, from the first item in the Strong Duality Theorem for Linear Programming. For example, if $c=-1, b=-1, A=0$, the primal problem with $x\in\mathbb{R}$ has feasible set $0\cdot x=-1$, $x\geq 0$, and the dual problem has feasible set $0\cdot y\leq -1$. Both feasible regions are therefore empty.

(f) Suppose both the primal and dual linear program are feasible, where both the primal and dual are defined in the Reference Sheet. Then there exist x,y which are feasible for the primal and dual problems respectively such that $c^Tx=b^Ty$.

TRUE; it is the last part of the strong duality theorem.

(g) Let m,n be positive integers with $m\leq n$. Let A be a real $m\times n$ matrix with full row rank, and let $b\in\mathbb{R}^m$. Let $K=\{x\in\mathbb{R}^n\colon x\geq 0,\ Ax=b\}$. Then x is a vertex of K if and only if x is a basic feasible solution.

TRUE; Lemma 4.21 from the notes.

(h) Suppose we have a bounded, feasible linear program in standard form given by an $m \times n$ matrix A, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. Then the Simplex algorithm will find the minimum of this linear program in a number of steps which is a polynomial in n and in m.

FALSE; we mentioned this in class. If the feasible region is the (bounded, nonempty) hypercube $P = \{x \in \mathbb{R}^n \colon 0 \leq x_i \leq 1, \forall\, 1 \leq i \leq n\}$, then P has 2^n vertices, and the Simplex algorithm may need to visit all of these vertices. That is, for any $n \geq 1$, the algorithm might need 2^n steps to terminate, and 2^n exceeds any polynomial in n.

2. Question 2

If the statement below is true, prove it. If the statement below is false, do your best to explain why it is false (a counterexample would be best, but a counterexample would not be required.)

Let $f \colon \mathbb{R}^2 \to \mathbb{R}$ be a C^3 function with exactly one critical point $x \in \mathbb{R}^2$. Assume that x is a local minimum of f. Then x is a global minimum of f.

Solution. This statement is false. Let $f(x, y) = x^2 + y^2(1 + x)^3$ for any $x, y \in \mathbb{R}$. Then $\nabla f(x, y) = (2x+3(1+x)^2y^2, 2y(1+x)^3) = 0$ only when $2y(1+x)^3 = 0$ and $2x = -3(1+x)^2y^2$. From the first equation, either $y = 0$ or $x = -1$. If $y = 0$, then $2x = 0$, so $x = 0$. If $x = -1$, then $2x = 0$ so $x = 0$. So, the only critical point of f is $(0, 0)$. We have $\partial^2 f/\partial x^2 = 2 + 6(1 + x)y^2$, $\partial^2 f/\partial y^2 = 2(1 + x)^3$, and $\partial^2 f/\partial x\partial y = 6y(1 + x)^2$. So, at $(0, 0)$, we have $D^2 f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. So, $(0, 0)$ is a local minimum of f. However, $(0, 0)$ is not a global minimum of f since $f(0, 0) = 0$, and $f(-2, 3) < 0$.

3. Question 3

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

$$f(x) = f(y) + (x - y)f'(y) + \frac{1}{2}\int_y^x f''(t)(x - t)dt \geq f(y) + (x - y)f'(y). \qquad (*)$$

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,

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

4. Question 4

Let $m \geq n$ be positive integers. Let A be a real $m \times n$ matrix with rank n. Let $b \in \mathbb{R}^m$. Show that the global minimum of $\|Ax - b\|^2$ among all $x \in \mathbb{R}^n$ occurs when

$$x = (A^TA)^{-1}A^Tb.$$

(You may assume without proof that A^TA is invertible. As usual $\|x\| = (xx^T)^{1/2}$.)

Solution. Let $x \in \mathbb{R}^n$ and define

$$\begin{aligned} f(x) &:= \|Ax - b\|^2 = (Ax - b)^T(Ax - b) = x^T A^T Ax - b^T Ax - x^T Ab + b^T b \\ &= \sum_{i,j=1}^{n} x_i (A^T A)_{ij} x_j - \sum_{i=1}^{n} (b^T A)_i x_i - \sum_{i=1}^{n} (Ab)_i x_i + b^T b. \end{aligned}$$

For any $1 \leq i, j \leq n$,

$$\frac{\partial f}{\partial x_i} = 2 \sum_{j\colon j \neq i} (A^T A)_{ij} x_j + 2x_i (A^T A)_{ii} - (b^T A)_i - (Ab)_i.$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = (A^T A)_{ij}.$$

That is, $\nabla f(x) = 2A^T Ax - 2A^T b$ and $D^2 f(x) = 2A^T A$. In particular, if $x = (A^T A)^{-1} A^T b$, then $\nabla f(x) = 0$.

We now show that f is a convex function. Let $t \in (0, 1)$ and let $x, y \in \mathbb{R}^n$. Then

$$\begin{aligned} &tf(x) + (1 - t)f(y) - f(tx + (1 - t)y) \\ &= t \|Ax - b\|^2 + (1 - t) \|Ay - b\|^2 - \|A(tx + (1 - t)y) - b\|^2 \\ &= t \|Ax - b\|^2 + (1 - t) \|Ay - b\|^2 - \|t(Ax - b) + (1 - t)(Ay - b)\|^2 \\ &= t \|Ax - b\|^2 + (1 - t) \|Ay - b\|^2 - t^2 \|Ax - b\|^2 - (1 - t)^2 \|Ay - b\|^2 - 2t(1 - t)(Ax - b)^T(Ay - b) \\ &= t(1 - t) \|Ax - b\|^2 + t(1 - t) \|Ay - b\|^2 - 2t(1 - t)(Ax - b)^T(Ay - b) \\ &= t(1 - t)\Big(\|Ax - b\|^2 + \|Ay - b\|^2 - 2(Ax - b)^T(Ay - b)\Big) \\ &= t(1 - t) \|(Ax - b) - (Ay - b)\|^2 \geq 0, \qquad \text{since } 0 < t < 1. \end{aligned}$$

That is, f is convex.

So, f is convex, and the point $x = (A^T A)^{-1} A^T b$ satisfies $\nabla f(x) = 0$. We conclude that this point x is the global minimum of f, since if $y \in \mathbb{R}^n$, and if we define $g(t) := f(x + ty)$, $t \in \mathbb{R}$, then Taylor's Theorem with integral remainder implies that

$$g(t) = g(0) + (y - x)g'(0) + \frac{1}{2} \int_y^x g''(t)(x - t)dt.$$

Then $g'(0) = \langle y, \nabla f(x)\rangle = 0$ since $\nabla f(x) = 0$, and $g''(t) = y^T[D^2 f(x + ty)]y = 2y^A A^T Ay = 2(Ay)^T Ay \geq 0$. Therefore,

$$g(t) \geq g(0) = f(x).$$

That is, $f(z) \geq f(x)$ for all $z \in \mathbb{R}^n$.

5. Question 5

For a linear program in standard form, show that the dual problem of the dual problem is the primal problem.

The dual problem is maximize $b^T y$ subject to $A^T y \leq c$. Equivalently, the dual problem is minimize $-b^T y$ subject to $A^T y \leq c$. We rewrite the constraint as

$$\text{minimize} \quad \begin{pmatrix} -b \\ b \\ 0 \end{pmatrix}^T \begin{pmatrix} y^+ \\ y^- \\ z \end{pmatrix} \quad \text{subject to the constraints}$$

$$\begin{pmatrix} A^T & -A^T & I \end{pmatrix} \begin{pmatrix} y^+ \\ y^- \\ z \end{pmatrix} = c, \qquad \begin{pmatrix} y^+ \\ y^- \\ z \end{pmatrix} \geq 0.$$

By the definition of the dual, the dual of this problem is

$$\text{maximize} \quad c^T x \quad \text{subject to the constraints}$$

$$\begin{pmatrix} A^T & -A^T & I \end{pmatrix}^T x \leq \begin{pmatrix} -b \\ b \\ 0 \end{pmatrix}.$$

The constraint now says $Ax \leq -b$ and $-Ax \leq b$, i.e. $Ax \geq -b$. That is, $Ax = -b$ and $x \leq 0$. Equivalently, this linear program is to minimize $c^T(-x)$ subject to the constraints $A(-x) = b$ and $x \leq 0$. Relabeling $-x$ as x, this linear program can be written as: minimize $c^T x$ subject to the constraints $Ax = b$ and $x \geq 0$, as desired.

6. Question 6

Give an example of a linear program in **standard form** where the primal problem is unbounded, and the dual problem is infeasible. Or, prove that no such linear program exists.

Solution. We use the following linear program where $c = -1$, $A = 0$, $b = 0$:

$$\text{minimize} \quad -x_1 \quad \text{subject to the constraint} \quad 0 \cdot x_1 = 0,\ x_1 \geq 0.$$

This problem is unbounded since every positive integer n is feasible, and $-n \to -\infty$ as $n \to \infty$

The dual of this problem is:

$$\text{maximize} \quad 0 \quad \text{subject to the constraint} \quad 0 \cdot y_1 \leq -1.$$

Since no real number y_1 satisfies $0 \cdot y_1 \leq -1$, the dual problem is infeasible.

7. Question 7

Let K be the following set of positive semidefinite 2×2 matrices

$$K = \left\{ \begin{pmatrix} a & b \\ b & c \end{pmatrix} \geq 0 \colon a, b, c \in \mathbb{R},\ a + c = 1 \right\}.$$

Show that K has infinitely many extreme points. (Hint: which matrices in K have determinant zero?)

Solution. Show that K has infinitely many extreme points. Let $a, b \in \mathbb{R}$ and consider a matrix of the form $\begin{pmatrix} a & b \\ b & 1-a \end{pmatrix}$ For a matrix in K, note that

$$\det \begin{pmatrix} a & b \\ b & 1-a \end{pmatrix} = a(1-a) - b^2.$$

So, this matrix has determinant zero when $a(1-a) = b^2$. So, let $0 \leq a \leq 1$, and consider matrices of the form $\begin{pmatrix} a & \sqrt{a(1-a)} \\ \sqrt{a(1-a)} & 1-a \end{pmatrix}$. This matrix has determinant zero and trace 1, so its eigenvalues x, y satisfy $xy = 0$ and $x + y = 1$. That is, one of the eigenvalues is zero, and the other is one. We claim that every matrix of the form $\begin{pmatrix} a & \sqrt{a(1-a)} \\ \sqrt{a(1-a)} & 1-a \end{pmatrix}$, $0 \leq a \leq 1$ is an extreme point of K. This claim completes the problem. We have already verified that each of these matrices is in K.

So, consider any such matrix A. Let $x \in \mathbb{R}^2$ be the zero eigenvector of C, $x \neq 0$. We argue by contradiction. Suppose there exist $B, C \in K$, $0 < t < 1$ such that $A = tB + (1-t)C$ and $B \neq C$, then $0 = x^T A x = x^T(tB + (1-t)C)x = tx^T Bx + (1-t)x^T Cx$. Since $B \geq 0$, $C \geq 0$, we have $x^T Bx \geq 0$ and $x^T Cx \geq 0$. Since $t > 0$ and $(1-t) > 0$, we conclude that $x^T Bx = x^T Cx = 0$. That is, x is also a zero eigenvector for B and C. That is, both B, C have a zero eigenvalue. So, both B, C have zero determinant. In summary, there exist $0 \leq a, b, c \leq 1$ such that

$$A = \begin{pmatrix} a & \sqrt{a(1-a)} \\ \sqrt{a(1-a)} & 1-a \end{pmatrix}$$

$$B = \begin{pmatrix} b & \sqrt{b(1-b)} \\ \sqrt{b(1-b)} & 1-b \end{pmatrix}, \quad C = \begin{pmatrix} c & \sqrt{c(1-c)} \\ \sqrt{c(1-c)} & 1-c \end{pmatrix}.$$

Define a function $f(a) := \sqrt{a(1-a)} = \sqrt{a - a^2}$, $0 \leq a \leq 1$. Then $f'(a) = (1/2)[a(1-a)]^{-1/2}[-2a+1]$, and

$$\begin{aligned} f''(a) &= -[a(1-a)]^{-1/2} + (-2a+1)(1/2)(-1/2)[a(1-a)]^{-3/2}(-2a+1) \\ &= [a(1-a)]^{-3/2}[-a(1-a) - (1/4)(1-2a)(-2a+1)] = [a(1-a)]^{-3/2}[-1/4] \\ &= -[a(1-a)]^{-3/2}/4. \end{aligned}$$

So, $f''(a) < 0$ for all $0 < a < 1$, $f(0) = f(1) = 0$. So, f is strictly concave on $[0, 1]$. That is, if $0 < t < 1$, and if $b \neq c$, (which is true since $B \neq C$), then $tf(b) + (1-t)f(c) < f(tb + (1-t)c)$. By assumption $A = tB + (1-t)C$, so $a = tb + (1-t)b$, $\sqrt{a(1-a)} = t\sqrt{b(1-b)} + (1-t)\sqrt{c(1-c)}$. But $tf(b) + (1-t)f(c) < f(tb + (1-t)c)$ says $\sqrt{a(1-a)} > t\sqrt{b(1-b)} + (1-t)\sqrt{c(1-c)}$. We have found a contradiction. The proof is complete.

8. Question 8

Let $g\colon [0, 1] \to \mathbb{R}$ be a continuous function. Assume that, for any C^1 function $h\colon [0, 1] \to \mathbb{R}$ with $h(0) = h(1) = 0$, we have

$$\int_0^1 g(x)h(x)dx = 0.$$

Conclude that $g(x) = 0$ for all $x \in [0, 1]$.

Solution. We argue by contradiction. Assume $g(x) \neq 0$ for some $x \in [0,1]$. Without loss of generality, $g(x) > 0$. By the definition of continuity, $\forall\, \varepsilon > 0$, there exists $\delta > 0$ such that, for all $y \in [0,1]$ with $0 < |y-x| < \delta$, we have $|g(x) - g(y)| < \varepsilon$. So, choosing $\varepsilon := g(x)/2$, we get some $\delta > 0$ such that, for all $y \in [0,1] \cap [x-\delta, x+\delta]$, we have $g(y) > g(x) - \varepsilon > g(x)/2 > 0$. Choosing some smaller interval inside $[0,1] \cap [x-\delta, x+\delta]$ if necessary, there is some $z \in (0,1)$ such that $(z, 4z) \subseteq [0,1]$ and for all $y \in (z, 4z)$ we have $g(y) > g(x)/2 > 0$. Let $h\colon [0,1] \to \mathbb{R}$ so that

$$h(t) = \begin{cases} 0 & \text{, if } 0 \leq t \leq z \\ (t-z)^2(t-4z)^2 & \text{, if } z \leq t \leq 4z \\ 0 & \text{, if } t > 4z. \end{cases}$$

Then $h(0) = h(1) = 0$, $h \in C^1$, $h(y)g(y) \geq 0$ for all $y \in [0,1]$, and $h(y)g(y) > g(x)z^4/2$ for all $y \in [2z, 3z]$. Therefore,

$$\int_0^1 h(y)g(y)dy \geq \int_{2z}^{3z} h(y)g(y)dy > z^5 g(x)/2 > 0,$$

a contradiction. We conclude that $g(x) = 0$ for all $x \in [0,1]$.

9. Question 9

Let $f\colon [0,1] \to \mathbb{R}$ be a C^1 function. Assume $f(0) = 0$ and $f(1) = 2$. Show that

$$\sqrt{5} \leq \int_0^1 \sqrt{1 + \left(\frac{d}{dx} f(x)\right)^2}\, dx.$$

You may assume that there exists a function $g\colon [0,1] \to \mathbb{R}$ such that g is C^1, $g(0) = 0$, $g(1) = 2$, and such that

$$\int_0^1 \sqrt{1 + \left(\frac{d}{dx} g(x)\right)^2}\, dx = \min_{\substack{f\colon [0,1]\to\mathbb{R}, \\ f(0)=0,\, f(1)=2,\, f \in C^1}} \int_0^1 \sqrt{1 + \left(\frac{d}{dx} f(x)\right)^2}\, dx. \qquad (*)$$

Solution. Let $t \in \mathbb{R}$, and let $h\colon [0,1] \to \mathbb{R}$ such that $h(0) = h(1) = 0$ and such that h is a C^1 function. Let $g_t := g + th$. Then $g_t(0) = 0$, $g_t(1) = 1$. So, since g minimizes the length, the following derivative is zero:

$$\begin{aligned} 0 = \frac{d}{dt}|_{t=0} \int_0^1 \sqrt{1 + \left(\frac{d}{dx} g_t(x)\right)^2}\, dx &= \int_0^1 \left(1 + \left(\frac{d}{dx} g(x)\right)^2\right)^{-1/2} \frac{d}{dx} g(x) \frac{d}{dx} h(x) dx \\ &= -\int_0^1 \frac{d}{dx}\left[\left(1 + \left(\frac{d}{dx} g(x)\right)^2\right)^{-1/2} \frac{d}{dx} g(x)\right] h(x) dx. \end{aligned}$$

In the last line, we integrated by parts. Then, by Question 8,

$$\frac{d}{dx}\left[\left(1 + \left(\frac{d}{dx} g(x)\right)^2\right)^{-1/2} \frac{d}{dx} g(x)\right] = 0.$$

That is, there exists a constant $c \in \mathbb{R}$ such that

$$\left(1 + \left(\frac{d}{dx}g(x)\right)^2\right)^{-1/2} \frac{d}{dx}g(x) = c.$$

Rearranging a bit,

$$\left(\frac{d}{dx}g(x)\right)^2 (1 - c^2) = c^2.$$

If $c^2 = 1$ we get a contradiction $0 = 1$. So, we can divide both sides by $1 - c^2$ to conclude that $|(d/dx)g(x)|$ is constant. Since $(d/dx)g(x)$ is continuous, we conclude that $\frac{d}{dx}g(x)$ is constant. Since $g(0) = 0$ and $g(1) = 2$, we must therefore have $g(x) = 2x$ for all $x \in [0, 1]$. Equation $(*)$ therefore concludes the proposition.