IE510 Applied Nonlinear Programming Homework 1

Instructor name: Ruoyu Sun   Student name: ____________________

Warning: The work your are submitting for this homework assignment must be your own, and suspiciously similar homework submitted by multiple individuals may be reported to the University for investigation.

1 Reading

- Reading: Textbook Section 1.1
- Appendix A.

2 Problems

Note: You need to justify your answer for all questions (you cannot just answer yes or no for the question without a proof or analysis).

1. (15 + 5 points) Consider the sequence defined as

$$\mathbf{x}^{k+1} = A\mathbf{x}^k, k = 1, 2, \ldots,$$

where $A \in \mathbb{R}^{2\times 2}$, $\mathbf{x}^k \in \mathbb{R}^2$. In questions a) to c), The contraction/non-expansive mappings are defined for the Euclidean norm [1].

a) Let the matrix $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, where $a, b \in \mathbb{R}$. Find the sufficient and necessary condition C1 on a, b so that $f(x) = Ax$ is a contraction mapping. Find the sufficient and necessary condition C2 on a, b so that $f(x) = Ax$ is a nonexpansive mapping.

b) For the matrix defined in a), find the sufficient and necessary condition under which the sequence $\{\mathbf{x}^k\}$ converges for any initial point $\mathbf{x}^k$. Is it the same as C1 or C2 or none of them?

c) Now consider another matrix $A = \begin{bmatrix} 0.5 & 1 \\ 0 & 0.5 \end{bmatrix}$. Is $f(\mathbf{x}) = A\mathbf{x}$ a non-expansive mapping (under 2-norm)? Does the sequence $\{\mathbf{x}^k\}$ converge?

d) (bonus 5 points) For the matrix in c), can you find a norm of $\mathbb{R}^2$ so that $f(\mathbf{x}) = A\mathbf{x}$ is a contraction mapping with respect to this norm?

Hint 1: For any positive-definite matrix B, $\|x\|_B \triangleq x^T Bx$ is a norm.
Hint 2: You may use computer to help the search.

[1] Note: if the inequality $\|f(x) - f(y)\|_* \leq \gamma \|x - y\|_*$ where the constant $\gamma < 1$ holds for some norm $\|\cdot\|_*$ such as ℓ_1 norm or ℓ_2 norm, f is called a contraction mapping with respect to the norm $\|\cdot\|_*$.

2. (15 points) Consider the problem $\min_{x,y\in\mathbb{R}} (x+1)^2 + x^2y^2$.

   (1) Is the objective function convex, concave or neither? Justify your answer.

   (2) Identify all stationary points.

   (3) Does the problem have a global minimum? If yes, find it; if no, explain why.

   (4) Does the function $(x+1)^2 + x^2y^2$ have a global maximum? If yes, find it; if no, explain why.

3. (20 points) Exercise 1.1.1 in the textbook. Pay attention to the requirement "each value".

For each value of the scalar β, find the set of all stationary points of the following function of the two variables x and y

$$f(x, y) = x^2 + y^2 + \beta xy + x + 2y.$$

Which of these stationary points are global minima?

4. (25 points) Exercise 1.1.2 in the textbook. Note that you need to justify what you find. You cannot just present the points and claim they are local minima or global minma without justification.

In each of the following problems fully justify your answer using optimality conditions.

(a) Show that the 2-dimensional function $f(x, y) = (x^2 - 4)^2 + y^2$ has two global minima and one stationary point, which is neither a local maximum nor a local minimum.

(b) Find all local minima of the 2-dimensional function $f(x, y) = \frac{1}{2}x^2 + x\cos y$.

(c) Find all local minima and all local maxima of the 2-dimensional function $f(x, y) = \sin x + \sin y + \sin(x + y)$ within the set $\{(x, y) \mid 0 < x < 2\pi,\ 0 < x < 2\pi\}$.

(d) Show that the 2-dimensional function $f(x, y) = (y - x^2)^2 - x^2$ has only one stationary point, which is neither a local maximum nor a local minimum.

(e) Consider the minimization of the function f in part (d) subject to no constraint on x and the constraint $-1 \le y \le 1$ on y. Show that there exists at least one global minimum and find all global minima.

5. (15 points) Exercise 1.1.3 in the textbook.

Let $f : \Re^n \mapsto \Re$ be a differentiable function. Suppose that a point x^* is a local minimum of f along every line that passes through x^*; that is, the function

$$g(\alpha) = f(x^* + \alpha d)$$

is minimized at $\alpha = 0$ for all $d \in \Re^n$.

(a) Show that $\nabla f(x^*) = 0$.

(b) Show by example that x^* need not be a local minimum of f. *Hint*: Consider the function of two variables $f(y, z) = (z - py^2)(z - qy^2)$, where $0 < p < q$; see Fig. 1.1.5. Show that $(0, 0)$ is a local minimum of f along every line that passes through $(0, 0)$. Furthermore, if $p < m < q$, then $f(y, my^2) < 0$ if $y \neq 0$ while $f(0, 0) = 0$.

6. (10 points + 5 bonus points)

a) (10 points) Let $\mathbf{A} \neq 0$ be an $n \times n$ symmetric positive-semidefinite matrix. Visualize the function $f(\mathbf{x}) = \|\mathbf{A} - \mathbf{x}\mathbf{x}^T\|_F^2$ where $\mathbf{x} \in \mathbb{R}^n$, in the following settings, using any software:
(i) $n = 1, \mathbf{A} = 1$;
(ii) $n = 2, \mathbf{A} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$;
(iii) $n = 200$, $\mathbf{A} = BB^\top$ where B is a 200×200 random Gaussian matrix.

Hint: For the command of visualization at matlab, you can search "Creating 3-D plots, Mathworks" at google (if your are using matlab). For (i), using command "plot" is enough. For (ii) and (iii), you need to draw both 3D plots and the contour. You can use the command "surf" and "contour" if you are using matlab. For (iii), projection to 2-dim space is needed; you can try different ways of projection so that the plot makes sense to you. Needless to say, you can use similar tools in other software. Print the figures you draw.

(b) (5 bonus points) Consider a nonlinear neural-network problem

$$\min_{v \in \mathbb{R}^{m \times 1}, W \in \mathbb{R}^{m \times d}} \sum_{i=1}^{n} (y_i - v^T \sigma(W x_i))^2,$$

where $(x_k, y_k), k = 1, 2, \ldots, n$ are the training data, W, v are the optimization variables and $\sigma(\cdot)$ is the activation function given by (exponential linear units): $\sigma(z) = \begin{cases} e^z - 1, & z < 0, \\ z, & z \geq 0. \end{cases}$

Generate the data $(x_k, y_k), k = 1, 2, \ldots, n$ from a random distribution (e.g., the standard Gaussian distribution $N(0, 1)$); note that x_k is d-dim vector and y_k is a scalar. Visualize the function by projection onto a 2-dimensional space.

Note: you can try different ways of projection, and check which visualization makes the most sense to you. You can try different choices of n, d, m. For instance, try $(n, d, m) = (500, 30, 5)$

and $(n, d, m) = (500, 30, 200)$, and other choices of m between 5 and 200. The more interesting findings you give, the more points you get.

Note: All problems are referred using version 2 of the textbook. You do not need a textbook though, since all problems are included in this pdf.