Nonlinear Optimization	**Homework 3 (Solutions)**
MATH 408 Spring 2019	

Exercises: 2.2, 2.4 - 2.7, 2.13, 2.15, 2.17 in "Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB"

Problem 2.2 (3 pts) Let $a \in \mathbb{R}^n$ be a nonzero vectpr. Show that the maximum of $f(x) = a^T x$ over $B = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is attained at $x^* = \frac{a}{\|a\|}$ and that the maximal value is $\|a\|$.

By the Cauchy-Schwarz inequality, we know that for all x, $a^T x \leq \|a\|\|x\|$, so for $x \in B$, $a^T x \leq \|a\|$. On the other hand, by taking $x^* = \frac{a}{\|a\|} \in B$, we achieve equality $a^T x^* = \frac{a^T a}{\|a\|} = \frac{\|a\|^2}{\|a\|} = \|a\|$, proving the result.

Problem 2.4 (2 pts) Show that if A, B are $n \times n$ positive semidefinite (psd) matrices, then $A + B$ is also psd.

Suppose $A, B \succeq 0$. For all $x \in \mathbb{R}^n$, observe that

$$x^T (A + B) x = \underbrace{x^T A x}_{\geq 0} + \underbrace{x^T B x}_{\geq 0} \geq 0$$

hence $A + B \succeq 0$

Problem 2.5 (3pts) Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$ be two symmetric matrices. Prove the following are equivalent.

(1) A and B are psd

(2) $C = \begin{pmatrix} A & 0_{n \times m} \\ 0_{m \times n} & B \end{pmatrix}$ is psd.

I will give two solutions. The first is more straightforward while the second argues about eigenvalues. While there is nothing wrong with the eigenvalue argument, almost everyone who tried to make it did so imprecisely or was unclear about the fact that the eigenvalues of C are the same as those of A and B.

First suppose that $A, B \succeq 0$. Since any element of $\mathbb{R}^{n+m}$ is of the form (x, y) for some $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, we compute

$$\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} A & 0_{n \times m} \\ 0_{m \times n} & B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^T A x + y^T B y \geq 0$$

Conversely suppose $C \succeq 0$. For any $x \in \mathbb{R}^n$, note that $0 \leq (x, 0)^T C (x, 0) = x^T A x$, so $A \succeq 0$. One can show $B \succeq 0$ similarly by taking $y \in \mathbb{R}^m$ arbitrary and $x = 0$.

Now with eigenvalues. Observe that the characteristic polynomial of C is

$$p(\lambda) = \det(A - \lambda I)\det(B - \lambda I)$$

since C is block diagonal. The zeroes of the characteristic polynomial are the eigenvalues of C. Note that the eigenvalues of C are simply those of A and B because $\det(A - \lambda I)$ and $\det(B - \lambda I)$ are the characteristic polynomials of A and B respectively. Hence the eigenvalues of C are nonnegative iff the eigenvalues of both A and B are nonnegative which proves the claim.

Problem 2.6 (4pts) Let $B \in \mathbb{R}^{n \times k}$ and let $A = BB^T$.

(1) Prove that $A \succeq 0$

For any $x \in \mathbb{R}^n$, $x^T A x = x^T B B^T x = (B^T x)^T (B^T x) = \|B^T x\|^2 \geq 0$, so $A \succeq 0$.

(2) Prove that $A \succ 0$ if and only if B has full row rank.

Suppose $A \succ 0$. Then for any $x \neq 0$, $x^T A x = x^T B B^T x = \|B^T x\|^2 > 0$, so $B^T x \neq 0$. Thus $\text{Null}(B^T) = \{0\}$, meaning B has full row rank.

Conversely if B has full row rank, then for any $x \neq 0$, $B^T x \neq 0$, so $x^T A x = \|B^T x\|^2 > 0$, so $A \succ 0$.

Problem 2.7

(1) (3 pts) Let A be an $n \times n$ symmetric matrix. Show that $A \succeq 0$ iff there exists a $B \in \mathbb{R}^{n \times n}$ such that $A = BB^T$.
Suppose $A \succeq 0$ and let $A = U \Lambda U^T$ be an eigenvalue decomposition of A. Since $\Lambda_{ii} = \lambda_i(A) \geq 0$ for all i, $\sqrt{\Lambda} = \text{diag}(\sqrt{\lambda_1(A)}, \ldots, \sqrt{\lambda_n(A)})$ is well defined. Letting $B = U\sqrt{\Lambda}$, it follows that $A = BB^T$.
See Problem 2.6(1) for the reverse implication.

(2) (2 pts) Let $x \in \mathbb{R}^n$ and let A be defined as

$$A_{ij} = x_i x_j, \quad \text{for } i, j = 1, \ldots, n$$

Show that $A \succeq 0$ and that it is not a positive definite matrix when $n > 1$.
Observe that $A = xx^T$ where we think of x as an $n \times 1$ matrix so by Problem 2.6(1), $A \succeq 0$. Note that $A = xx^T$ is a rank 1 matrix with $\text{Range}(A) = \text{Span}\{x\}$, hence $\text{nullity}(A) = n - \text{rank}(A) = n - 1 > 0$ for $n > 1$, which means that A has an eigenvalue equal to 0, so A is not positive definite.

Problem 2.13 Determine if the following matrices are positive/negative semidefinite/definite or indefinite.

(1) (1 pt) $\begin{pmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 3 \end{pmatrix} \succeq 0$ (see problem 2.5)

(2) (1 pt) $\begin{pmatrix} 2 & 2 & 2 \\ 2 & 3 & 3 \\ 2 & 3 & 3 \end{pmatrix} \succeq 0$ (by principal minors test)

(3) (1 pt) $\begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \\ 3 & 1 & 2 \end{pmatrix}$ indefinite (see 1,2 and 1,3 minors)

(4) (2 pts) $\begin{pmatrix} -5 & 1 & 1 \\ 1 & -7 & 1 \\ 1 & 1 & -5 \end{pmatrix} \preceq 0$ (observe that the negative of this matrix is psd)

Problem 2.15 Determine whether the following functions are coercive or not. (1 pt each) Most did not approach this problem particularly rigorously but the grading was done lightly.

(1) $f(x_1, x_2) = x_1^4 + x_2^4$

Coercive. I was not expecting this level of detail, but here is a proof. By a lemma from class notes, recall that f is coercive if every sublevel set $L_\alpha := \{x : f(x) \leq \alpha\}$ is compact. By continuity of f, each sublevel set is closed, so we need only show it is bounded. Indeed, if $\|(x_1, x_2)\|_4^4 = x_1^4 + x_2^4 \leq \alpha$, then $\|(x_1, x_2)\|_4 \leq \alpha^{1/4}$ for any $(x_1, x_2) \in L_\alpha$. By equivalence of norms, there also exists some constant $M(\alpha)$ such that $\|(x_1, x_2)\|_2 \leq M(\alpha)$ for all $(x_1, x_2) \in L_\alpha$, hence L_α is bounded. This proves that L_α is compact for all $\alpha \in \mathbb{R}$.

(2) $f(x_1, x_2) = e^{x_1^2} + e^{x_2^2} - x_1^{200} - x_2^{200}$

Coercive. Doing this rigorously in similar fashion to above turns out to be a mess, but it boils down to the fact that $e^{x_i^2}$ grows faster than x_i^{200} for sufficiently large value of x_i, which most people realized.

(3) $f(x_1, x_2) = 2x_1^2 - 8x_1x_2 + x_2^2$

Not coercive. This is a quadratic defined by $H = \begin{pmatrix} 2 & -4 \\ -4 & 1 \end{pmatrix}$ which is indefinite.

(4) $f(x_1, x_2) = 4x_1^2 + 2x_1x_2 + 2x_2^2$

Coercive. This is a quadratic defined by $H = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$ which is positive definite.

(5) $f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3$

Not coercive. For $v = (1, -1, 0)$, $\|tv\| \to \infty$ as $t \to \infty$, but $f(tv) = 0$ for all t.

(6) $f(x_1, x_2) = x_1^2 - 2x_1x_2^2 + x_2^4$

Not coercive. Note $f = (x_1 - x_2^2)^2$ so for $v = (1,1)$, $\|tv\| \to \infty$ as $t \to \infty$, but $f(tv) = 0$ for all t.

(7) $f(x) = \frac{x^T Ax}{\|x\|+1}$ where $A \in \mathbb{R}^{n\times n}$ is positive definite.

By Rayleigh Ritz, $\lambda_{\min}(A)\|x\|^2 \leq x^T Ax$ for any x. Dividing both sides by $\|x\|+1$, we have

$$\frac{\lambda_{\min}(A)\|x\|^2}{\|x\|+1} \leq \frac{x^T Ax}{\|x\|+1}$$

Observe that the LHS $\to \infty$ as $\|x\| \to \infty$, hence the RHS is coercive.

Problem 2.17 (3pts each) Find all stationary points of the following functions and classify them according to whether they are saddle points, strict/nonstrict local/global minimum/maximum points:

For all of the following, the method is to set $\nabla f(x) = 0$ to find stationary points $\overline{x}$ and determine the eigenvalues of the Hessian $\nabla^2 f(\overline{x})$ at each stationary point $\overline{x}$. The eigenvalues of the Hessian evaluated at $\overline{x}$ give information about the curvature of the function, hence allow us to classify the stationary points.

(1) $f(x_1, x_2) = (4x_1^2 - x_2)^2$

$$\nabla f(x) = \begin{pmatrix} 16x_1(4x_1^2 - x_2) \\ -2(4x_1^2 - x_2) \end{pmatrix}$$

which equals 0 when $x_2 = 4x_1^2$, so the set of stationary points is the 1 dimensional family of points which satisfy this equation. Observe that for any such $\overline{x} = (x_1, 4x_1^2)$ the Hessian evaluates to

$$\nabla^2 f(\overline{x}) = \begin{pmatrix} 192x_1^2 - 16(4x_1^2) & -16x_1 \\ -16x_1 & 2 \end{pmatrix} = \begin{pmatrix} 128x_1^2 & -16x_1 \\ -16x_1 & 2 \end{pmatrix}$$

By the principal minors test, this is psd at every stationary point, hence each stationary point is at least a local minimizer. Observe that in fact $f(\overline{x}) = 0 \leq f(x)$ for all x for every stationary point $\overline{x}$, so they are nonstrict global minimizers.

(2) $f(x_1, x_2, x_3) = x_1^4 - 2x_1^2 + x_2^2 + 2x_2x_3 + 2x_3^2$

$$\nabla f(x) = \begin{pmatrix} 4x_1(x_1^2 - 1) \\ 2x_2 + 2x_3 \\ 2x_2 + 4x_3 \end{pmatrix}$$

which equals 0 when $x_1 = 0, \pm 1$ and $x_2 = -x_3$ and $x_2 = -2x_3$, i.e., $x_2 = x_3 = 0$. The Hessian is

$$\nabla^2 f(\overline{x}) = \begin{pmatrix} 12x_1^2 - 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix}$$

which is indefinite at $(0, 0, 0$ and positive definite at $(1, 0, 0)$ and $(-1, 0, 0)$, hence $(0, 0, 0)$ is a saddle point and $(1, 0, 0)$ and $(-1, 0, 0)$ are strict local minimizers. They are in fact nonstrict global because f is bounded below by $f(1, 0, 0) = f(-1, 0, 0) = -1$.

(3) $f(x_1, x_2) = 2x_2^3 - 6x_2^2 + 3x_1^2 x_2$

$$\nabla f(x) = \begin{pmatrix} 6x_1 x_2 \\ 6x_2^2 - 12x_2 + 3x_1^2 \end{pmatrix}$$

which equals 0 when $x_1 = x_2 = 0$ or when $x_1 = 0$ and $x_2 = 2$. The Hessian is

$$\nabla^2 f(\overline{x}) = \begin{pmatrix} 6x_2 & 6x_1 \\ 6x_1 & 12x_2 - 12 \end{pmatrix}$$

which is negative semidefinite at $(0,0)$ and positive definite at $(0,2)$. Hence $(0,2)$ is a strict local minimizer and $(0,0)$ is a candidate to be a local maximizer, however, notice that for any small neighborhood of $(0,0)$, f increases in the positive x_1 direction and decreases in the positive x_2 direction, so $(0,0)$ is a saddle point.

(4) $f(x_1, x_2) = x_1^4 + 2x_1^2 x_2 + x_2^2 - 4x_1^2 - 8x_1 - 8x_2$

$$\nabla f(x) = \begin{pmatrix} 4x_1^3 + 4x_1 x_2 - 8x_1 - 8 \\ 2x_1^2 + 2x_2 - 8 \end{pmatrix}$$

which equals 0 at $(1,3)$. The Hessian is

$$\nabla^2 f(\overline{x}) = \begin{pmatrix} 12x_1^2 + 4x_2 - 8 & 4x_1 \\ 4x_1 & 2 \end{pmatrix}$$

which is positive definite at $(1,3)$, hence $(1,3)$ is a strict local minimizer. It is in fact global because $f(x) = (x_1^2 + x_2 - 4)^2 + (x_1 - 1)^2 - 20 \geq -20$ while $f(1,3) = -20$.

(5) $f(x_1, x_2) = (x_1 - 2x_2)^4 + 64x_1 x_2$

$$\nabla f(x) = \begin{pmatrix} 4(x_1 - 2x_2)^3 + 64x_2 \\ -8(x_1 - 2x_2)^3 + 64x_1 \end{pmatrix}$$

which equals 0 at $(0,0), (-1, 1/2), (1, -1/2)$. The Hessian is

$$\nabla^2 f(\overline{x}) = \begin{pmatrix} 12(x_1 - 2x_2)^2 & 64 - 24(x_1 - 2x_2)^2 \\ 64 - 24(x_1 - 2x_2)^2 & 48(x_1 - 2x_2)^2 \end{pmatrix}$$

which is indefinite at $(0,0)$, and positive definite at $(-1, 1/2)$ and $(1, -1/2)$. Hence $(0,0)$ is a saddle point and $(-1, 1/2)$ and $(1, -1/2)$ are strict local minimizers.

(6) $f(x_1, x_2) = 2x_1^2 + 3x_2^2 - 2x_1 x_2 + 2x_1 - 3x_2$

$$\nabla f(x) = \begin{pmatrix} 4x_1 - 2x_2 + 2 \\ 6x_2 - 2x_1 - 3 \end{pmatrix}$$

which equals 0 at $(-3/10, 2/5)$. The Hessian is

$$\nabla^2 f(\overline{x}) = \begin{pmatrix} 4 & -2 \\ -2 & 6 \end{pmatrix}$$

which is constant and positive definite, hence $(-3/10, 2/5)$ is a strict global minimizer since f is coercive because f is quadratic and $\nabla^2 f \succ 0$ everywhere.

(7) $f(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^2 + x_1 - x_2$

$$\nabla f(x) = \begin{pmatrix} 2x_1 + 4x_2 + 1 \\ 4x_1 + 2x_2 - 1 \end{pmatrix}$$

which equals 0 at $(1/2, -1/2)$. The Hessian is

$$\nabla^2 f(\overline{x}) = \begin{pmatrix} 2 & 4 \\ 2 & 4 \end{pmatrix}$$

which is constant and indefinite, hence $(1/2, -1/2)$ is a saddle point.