March 15, 2017 Student Name: SOLUTIONS

MATH 4/5420 - Optimization
Midterm Exam - Spring 2017

1. (a) Given the linear function $f(x) = b^T x$, where $b \in \mathbb{R}^n$, compute the gradient ∇f, using the directional derivative calculation. [Avoid the use of coordinates in your proof]

Fix $x \in \mathbb{R}^n$.

We know (by Chain Rule) that for $p \in \mathbb{R}^n$, $\frac{d}{d\alpha} f(x+\alpha p)\Big|_{\alpha=0} = \nabla f(x)^T p$.

On the other hand

$$f(x+\alpha p) = b^T(x+\alpha p) = b^T x + \alpha b^T p \quad \text{so} \quad \frac{d}{d\alpha} f(x+\alpha p) = b^T p, \ \forall \alpha.$$

In particular, for $\alpha = 0$, $\frac{d}{d\alpha} f(x+\alpha p)\Big|_{\alpha=0} = b^T p$.

Since $(\nabla f(x))^T p = b^T p$ for all $p \in \mathbb{R}^n$ it follows that

$$\nabla f(x) = b$$

[e.g. take $p = \nabla f(x) - b$ to conclude $0 = [\nabla f(x) - b]^T p = [\nabla f(x) - b]^T [\nabla f(x) - b] = \|\nabla f(x) - b\|$ $\Rightarrow \nabla f(x) = b$.]

(b) Given the quadratic function $f(x) = x^T A x$, where A is a symmetric matrix, positive definite matrix, show that $x^* = 0$ is the unique global minimizer for f.

By definition, $x^T A x \geq 0$, for all $x \in \mathbb{R}^n$ and

$$x^T A x > 0 \quad \text{for } x \neq 0$$

It follows that

$$0 = f(0) < f(x) \quad \text{for all } x \neq 0 \Rightarrow x^* = 0 \text{ is the (unique) global minimizer}$$

Note: Could also involve $\nabla f(x) = 2Ax$, $\nabla^2 f(x) = 2A$

Necessary condition for min: $0 = \nabla f(x) = 2Ax \Rightarrow Ax = 0$

But A is pos def, hence invertible $\Rightarrow x = 0$ is the unique solution. In addition $\nabla^2 f(x) = 2A > 0 \Rightarrow x = 0$ is the unique (local) minimizer of f. In addition $f(x) = x^T A x \geq \alpha \|x\|^2$,

2. Using Taylor's Theorem [redacted]

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp)p, \text{ for some } t \in (0,1)$$

prove that if x^* is such that $\nabla f(x^*) = 0$, then: [Choose only one of the two]

(a) [Necessary condition] If x^* is a (local) minimizer, then $\nabla^2 f(x^*)$ is positive semidefinite.

At $x = x^*$: $f(x^*+p) = f(x^*) + \frac{1}{2} p^T \nabla^2 f(x^*+tp)p$ (since $\nabla f(x^*) = 0$).

Assume x^* local minimizer but $\nabla^2 f(x^*)$ not pos. semidef, that is $\exists\, p \in \mathbb{R}^n$ with $p^T \nabla^2 f(x^*) p < 0$. Then $p^T \nabla^2 f(x) p < 0$, for all x in a neighborhood of x^*. Thus, for small $\alpha > 0$

$$f(x^* + \alpha p) = f(x^*) + \frac{1}{2}\alpha^2 \underbrace{p^T \nabla^2 f(x^* + t\alpha p) p}_{<0} < f(x^*).$$

Contradic x^* is local min, QED

(b) [Sufficient condition] If $\nabla^2 f(x^*)$ is positive definite, then x^* is a (local) minimizer.

Since $\nabla^2 f(x^*)$ is pos def, so is $\nabla^2 f(x)$ for all x in a neighborhood of x^* $\Rightarrow$

$$f(x) = f(x^*+p) = f(x^*) + \frac{1}{2} \underbrace{p^T \nabla^2 f(x^* + tp) p}_{>0} > f(x^*)$$

for all x in that neighborhood.

(c) What can happen in (b) if the Hessian is not positive definite but only semi-positive definite? Give an example in 1-dim and 2-dim.

Assume x^* is such that $\nabla f(x^*) = 0$.

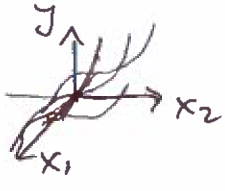
The necessary cond ($\nabla^2 f(x^*) \geq 0$) is not sufficient for x^* to be local min (if $\nabla^2 f(x^*)$ is not pos def), as one can see from the following two examples:

<u>Ex 1-d</u>: $f(x) = x^3$, $x^* = 0$ has $f''(x^*) = 0$ (so not pos def but pos semidef).

$x^* = 0$ is not a local min nor a local max

<u>Ex 2-d</u>: $f(x_1, x_2) = x_1^2 x_2^3$. Again $\nabla f(0,0) = 0$

$\nabla^2 f(0,0)$ is singular



but $(0,0)$ is not local min nor local max

3. Consider the optimization problem

$$\min \|Ax - b\|^2$$

where A is a (possibly nonsquare) $m \times n$ matrix with $m \geq n$, and $b \in \mathbb{R}^m$.

(a) Rewrite the objective function as a quadratic function of $x \in \mathbb{R}^n$ and write down (without computing) the gradient and Hessian of this quadratic function.

$$f(x) = \|Ax - b\|^2 = (Ax - b)^T (Ax - b) = (x^T A^T - b^T)(Ax - b)$$

$$= (x^T A^T A x) - 2b^T A x + b^T b$$

$$= x^T (A^T A) x - 2(A^T b)^T x + b^T b$$

So $\nabla f(x) = 2A^T A x - 2A^T b = 2A^T(Ax - b)$

$\nabla^2 f(x) = 2A^T A$.

(b) Show that if A has maximum rank (all n columns are linearly independent), then the quadratic function is strictly convex and hence the minimization problem has an unique minimizer.

The quadratic function $f(x) = x^T Q x -$ [illegible]$^T x + d$ is strictly convex when Q is positive definite

$(m \times n)$

$Q = A^T A$ is positive definite when A has max rank n

Indeed $x^T Q x = x^T A^T A x = (Ax)^T Ax = \|Ax\|^2 > 0$ for all $x \neq 0$

[The system $Ax = 0$ has the only solution $x = 0$, since A has rank $= n$]

(c) Write down the steepest descent method with exact line search. Then perform the first iteration, that is, given a starting point x^0, compute x^1.

Given $x^0 \in \mathbb{R}^n$

$$p_0 = -\nabla f(x^0) = -2A^T(Ax^{(0)} - b)$$

$$\alpha_0 = -\frac{\nabla f(x^0)^T p_0}{p_0^T (A^T A) p_0} = \frac{p_0^T p_0}{(Ap_0)^T A p_0} = \frac{\|p_0\|^2}{\|Ap_0\|^2}.$$

$$x^{(1)} = x^{(0)} + \alpha_0 p_0 = x^0 + \frac{\|p_0\|^2}{\|Ap_0\|^2}(-2A^T A x^0 + 2A^T b).$$

4. Consider the least-square approximation problem

$$\min_{x_1,x_2} \sum |y_j - (x_1 + x_2 t_j)|^2$$

of finding the line that best fits the data set $\{t_j, y_j\} = \{(-1,1), (0,-1), (1,0), (2,2)\}$, which amounts to finding the solution to the minimization problem

$$\min \|Ax - b\|^2$$

where

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } b = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 2 \end{bmatrix}$$

Following the steps in exercise 3, find the minimizer x^* and give a reason for why the steepest descent method converges.

A has (max) rank $= 2$.

$f(x) = \|Ax - b\|^2 \Rightarrow$ the minimizer x^* is the (unique) solution to the $0 = \nabla f(x) = 2A^T(Ax^* - b) \Leftrightarrow A^TAx^* = A^Tb$.

$\Leftrightarrow x^* = (A^TA)^{-1}A^Tb$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} = 2\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} = \text{non singular}$$

with

$$(A^TA)^{-1} = \frac{1}{2}\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}^{-1} = \frac{1}{2}\cdot\frac{1}{5}\begin{bmatrix} 3 & \\ -1 & \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

So $x^* = (A^TA)^{-1}A^Tb = \frac{1}{10}\begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3/10 \\ 4/10 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}$

The steepest descent converges because θ_k = angle between p_k & p_{k+1} $= \varepsilon$

so $\sum_{k\geq 1}^{\infty} \cos^2\theta_k \|\nabla f_k\|^2 = \sum_{k\geq 1}^{\infty} \|\nabla f_k\|^2 < +\infty$

$\Rightarrow \|\nabla f_k\| \to 0 \Rightarrow \|A^TA x^k - A^Tb\| \to 0 \Rightarrow A^TAx^k \to A^Tb$

But A^TA is invertible $\Rightarrow x^k \to (A^TA)^{-1}A^Tb = x^*$.

QED