COURSEWORK 2

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

477 - Computational Optimisation

Coursework 2

Author:
Jiahao Lin (CID: 00837321)

Date: November 21, 2016

1 Part 1

1.1 Q.1

a) To show that $\boldsymbol{\Delta} x_k$ is descent direction at x_k, we need to show that:

$$\nabla f(x_k)^\top \boldsymbol{\Delta} x_k < 0 \tag{1}$$
$$\tag{2}$$

Since:

$$\nabla^2 f(x_k) \boldsymbol{\Delta} x_k = -\nabla f(x_k) \tag{3}$$
$$\tag{4}$$

We can get:

$$\nabla f(x_k)^\top \boldsymbol{\Delta} x_k = -\nabla f(x_k)^\top \frac{\nabla f(x_k)}{\nabla^2 f(x_k)} \tag{5}$$
$$= -\frac{\nabla f(x_k)^\top \nabla f(x_k)}{\nabla^2 f(x_k)} \tag{6}$$

We can see that this is less than zero since the numerator is positive definite and the same for denominator:

$$\nabla^2 f(x_k) \geq m\boldsymbol{I} \tag{7}$$

b) We can set tolerane to e^{-08} and say that:

$$|f(x_{k+1}) - f(x_k)| < tor \tag{8}$$
$$||\nabla f(x_k)||_2 < tor \tag{9}$$
$$||x_{k+1} - x_k||_2 < tor \tag{10}$$

This is to check that the First Order Necessary Condition is satisfied.

c) Yes, since the function is strongly convex, but the condition is that the initial point x_0 has to be close enough to the optimal point.

d) First say that:

$$x_{k+1} = x_k + t_k \boldsymbol{\Delta} x_k \tag{11}$$

Then

$$f(x_{k+1}) = f(x_k + t_k \boldsymbol{\Delta} x_k) \tag{12}$$

Now use Taylor expansion to expand the above function into second order:

$$f(x_k + t_k \boldsymbol{\Delta} x_k) \approx f(x_k) + t_k \langle \nabla f(x_k), \boldsymbol{\Delta} x_k \rangle + \frac{1}{2} \nabla^2 f(x_k) ||t_k||_2^2 ||\boldsymbol{\Delta} x_k||_2^2 \tag{13}$$

$$\Delta x_k = -\frac{\nabla f(x_k)}{\nabla^2 f(x_k)} \tag{14}$$

$$\langle \nabla f(x_k), \Delta x_k \rangle = -\frac{\|\nabla f(x_k)\|_2^2}{\nabla^2 f(x_k)} \tag{15}$$

We need to show that:

$$f(x_k) + t_k \langle \nabla f(x_k), \Delta x_k \rangle + \frac{1}{2}\nabla^2 f(x_k) \|t_k\|_2^2 \|\Delta x_k\|_2^2 \leq f(x_k) + \alpha t_k \langle \nabla f(x_k), \Delta x_k)\rangle \tag{16}$$

$$t_k \langle \nabla f(x_k), \Delta x_k)\rangle + \frac{1}{2}\nabla^2 f(x_k) \|t_k\|_2^2 \|\Delta x_k\|_2^2 \leq \alpha t_k \langle \nabla f(x_k), \Delta x_k)\rangle \tag{17}$$

$$\frac{1}{2}\nabla^2 f(x_k) \|t_k\|_2^2 \|\Delta x_k\|_2^2 \leq (\alpha - 1) t_k \langle \nabla f(x_k), \Delta x_k)\rangle \tag{18}$$

$$\frac{1}{2}\nabla^2 f(x_k) t_k \|\Delta x_k\|_2^2 \leq (\alpha - 1) \langle \nabla f(x_k), \Delta x_k)\rangle \tag{19}$$

Since we have condition on α:

$$0 < \alpha < 0.5 \tag{20}$$

$$-1 < \alpha - 1 < -0.5 \tag{21}$$

Since $\langle \nabla f(x_k), \Delta x_k)\rangle < 0$:

$$-\langle \nabla f(x_k), \Delta x_k)\rangle \geq (\alpha - 1)\langle \nabla f(x_k), \Delta x_k)\rangle \geq -0.5 \langle \nabla f(x_k), \Delta x_k)\rangle \tag{22}$$

So we need to show that:

$$\frac{1}{2}\nabla^2 f(x_k) t_k \|\Delta x_k\|_2^2 \leq -0.5 \langle \nabla f(x_k), \Delta x_k)\rangle \tag{23}$$

$$\nabla^2 f(x_k) t_k \|\Delta x_k\|_2^2 \leq -\langle \nabla f(x_k), \Delta x_k)\rangle \tag{24}$$

$$t_k \leq -\frac{\langle \nabla f(x_k), \Delta x_k)\rangle}{\nabla^2 f(x_k) \|\Delta x_k\|_2^2} \tag{25}$$

Since we have condition on $\nabla^2 f(x_k)$:

$$\boldsymbol{mI} < \nabla^2 f(x_k) \leq \boldsymbol{MI} \tag{26}$$

$$\frac{\langle \nabla f(x_k), \Delta x_k)\rangle}{\nabla^2 f(x_k) \|\Delta x_k\|_2^2} \geq \frac{\langle \nabla f(x_k), \Delta x_k)\rangle}{\boldsymbol{M} \|\Delta x_k\|_2^2} \tag{27}$$

$$-\frac{\langle \nabla f(x_k), \Delta x_k)\rangle}{\nabla^2 f(x_k) \|\Delta x_k\|_2^2} \leq -\frac{\langle \nabla f(x_k), \Delta x_k)\rangle}{\boldsymbol{M} \|\Delta x_k\|_2^2} \tag{28}$$

$$\tag{29}$$

We got:

$$t_k \leq -\frac{\langle \nabla f(x_k), \Delta x_k)\rangle}{M \|\Delta x_k\|_2^2} \tag{30}$$

Which is the given condition on t_k
End —

$$t_k\langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle \tag{31}$$

Since we have the condition on t_k:

$$0 < t \leq -\frac{\langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle}{\boldsymbol{M}||\boldsymbol{\Delta} x_k||_2^2} \tag{32}$$

$$0 \geq \langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle t_k > -\langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle^\top \frac{\langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle}{\boldsymbol{M}||\boldsymbol{\Delta} x_k||_2^2} \tag{33}$$

$$0 \geq \langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle t_k > -\frac{||\nabla f(x_k)||^2||\boldsymbol{\Delta} x_k)||^2}{\boldsymbol{M}||\boldsymbol{\Delta} x_k||_2^2} \tag{34}$$

$$0 \geq \langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle t_k > -\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{35}$$

Now let's take a look at second term:

$$\frac{1}{2}\nabla^2 f(x_k)||t_k||_2^2||\boldsymbol{\Delta} x_k||_2^2 \tag{36}$$

Again substitute in condition for t:

$$0 < t \leq -\frac{\langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle}{\boldsymbol{M}||\boldsymbol{\Delta} x_k||_2^2} \tag{37}$$

$$0 < ||t||_2^2 \leq \frac{||\nabla f(x_k)||^2||\boldsymbol{\Delta} x_k)||^2}{\boldsymbol{M}^\top\boldsymbol{M}||\boldsymbol{\Delta} x_k||_2^4} \tag{38}$$

$$0 < ||t||_2^2 \leq \frac{||\nabla f(x_k)||^2}{\boldsymbol{M}^\top\boldsymbol{M}||\boldsymbol{\Delta} x_k||_2^2} \tag{39}$$

$$0 < ||t||_2^2||\boldsymbol{\Delta} x_k||_2^2 \leq \frac{||\nabla f(x_k)||^2}{\boldsymbol{M}^\top\boldsymbol{M}} \tag{40}$$

$$\tag{41}$$

Since we have the upper bound for $\nabla^2 f(x_k)$:

$$\boldsymbol{mI} < \nabla^2 f(x_k) \leq \boldsymbol{MI} \tag{42}$$

$$0 < ||t||_2^2||\boldsymbol{\Delta} x_k||_2^2\nabla^2 f(x_k) \leq \frac{\nabla^2 f(x_k)||\nabla f(x_k)||^2}{\boldsymbol{M}^\top\boldsymbol{M}} \tag{43}$$

$$0 < ||t||_2^2||\boldsymbol{\Delta} x_k||_2^2\nabla^2 f(x_k) \leq \frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{44}$$

$$0 < \frac{1}{2}\nabla^2 f(x_k)||t_k||_2^2||\boldsymbol{\Delta} x_k||_2^2 \leq \frac{1}{2}\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{45}$$

Now combine first and second term:

$$0 \geq \langle\nabla f(x_k), \boldsymbol{\Delta} x_k)\rangle t_k > -\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{46}$$

$$0 < \frac{1}{2}\nabla^2 f(x_k)||t_k||_2^2||\mathbf{\Delta}x_k||_2^2 \leq \frac{1}{2}\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{47}$$

$$-\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} < t_k\langle\nabla f(x_k), \mathbf{\Delta}x_k\rangle\rangle + \frac{1}{2}\nabla^2 f(x_k)||t_k||_2^2||\mathbf{\Delta}x_k||_2^2 \leq \frac{1}{2}\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{48}$$

From condition on α:

$$0 < \alpha < 0.5 \tag{49}$$

$$\boldsymbol{mI} < \nabla^2 f(x_k) \leq \boldsymbol{MI} \tag{50}$$

$$-\frac{||\nabla f(x_k)||^2}{\nabla^2 f(x_k)} \leq -\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{51}$$

$$\frac{1}{2}\frac{||\nabla f(x_k)||^2}{\nabla^2 f(x_k)} \geq \frac{1}{2}\frac{||\nabla f(x_k)||^2}{\boldsymbol{M}} \tag{52}$$

$$-\frac{||\nabla f(x_k)||^2}{\nabla^2 f(x_k)} < t_k\langle\nabla f(x_k), \mathbf{\Delta}x_k\rangle\rangle + \frac{1}{2}\nabla^2 f(x_k)||t_k||_2^2||\mathbf{\Delta}x_k||_2^2 \leq \frac{1}{2}\frac{||\nabla f(x_k)||^2}{\nabla^2 f(x_k)} \tag{53}$$

$$0 < \alpha\frac{||\nabla f(x_k)||^2}{\nabla^2 f(x_k)} < \frac{1}{2}\frac{||\nabla f(x_k)||^2}{\nabla^2 f(x_k)} \tag{54}$$

$$\tag{55}$$

e) Assume $t_0 = 1$, maximum number of backtracking step is max n that:

$$1 * \beta^n \leq -\frac{\langle\nabla f(x_n), \mathbf{\Delta}x_n\rangle\rangle}{M||\mathbf{\Delta}x_n||_2^2} \tag{56}$$

$$\beta^n \leq -\frac{\langle\nabla f(x_n), \mathbf{\Delta}x_n\rangle\rangle}{M||\mathbf{\Delta}x_n||_2^2} \tag{57}$$

$$\tag{58}$$

2 Part 2

2.1 Q.2

a) KKT conditions:

$$min \quad z_1^2 + (x_2+1)^2 \tag{59}$$

$$g(x^*) = exp(x_1^*) - x_2^* \leq 0 \tag{60}$$

$$\mu^* \geq 0 \tag{61}$$

$$2x_1^* + \mu^* exp(x_1^*) = 0 \tag{62}$$

$$2(x_2^*+1) + \mu^* * (-1) = 0 \tag{63}$$

$$\mu^*(exp(x_1^*) - x_2^*) = 0 \tag{64}$$

if $\mu^* = 0$:

$$2x_1^* = 0 \tag{65}$$

$$x_1^* = 0 \tag{66}$$
$$2(x_2^* + 1) = 0 \tag{67}$$
$$x_2^* = -1 \tag{68}$$
$$g(x^*) = exp(0) - (-1) = 2 > 0 \tag{69}$$
$$\tag{70}$$

This is contradicting our condition, so $\mu^* > 0$:

$$exp(x_1^*) - x_2^* = 0 \tag{71}$$
$$exp(x_1^*) = x_2^* \tag{72}$$
$$2(x_2^* + 1) + \mu^*(-1) = 0 \tag{73}$$
$$2(exp(x_1^*) + 1) = \mu^* \tag{74}$$
$$2x_1^* + \mu^* * exp(x_1^*) = 0 \tag{75}$$
$$2x_1^* + 2(exp(x_1^*) + 1) * exp(x_1^*) = 0 \tag{76}$$

Since $exp(x_1^*) > 0$:

$$2(exp(x_1^*) + 1) * exp(x_1^*) > 0 \tag{77}$$
$$\tag{78}$$

combine with

$$2x_1^* + 2(exp(x_1^*) + 1) * exp(x_1^*) = 0 \tag{79}$$

we got

$$2x_1^* < 0 \tag{80}$$
$$x_1^* < 0 \tag{81}$$
$$0 < exp(x_1^*) < 1 \tag{82}$$
$$1 < exp(x_1^*) + 1 < 2 \tag{83}$$
$$0 < (exp(x_1^*) + 1) * exp(x_1^*) < 2 \tag{84}$$
$$\tag{85}$$

combine with the above equation agian:

$$x_1^* + (exp(x_1^*) + 1) * exp(x_1^*) = 0 \tag{86}$$
$$-1 < x_1^* < 0 \tag{87}$$

b) KKT condition:

$$min \quad c^\top x + 8 \tag{88}$$
$$g(x^*) = \frac{1}{2}||x^*||^2 - 1 \leq 0 \tag{89}$$
$$\mu^* \geq 0 \tag{90}$$

$$c + \mu^* x^* = 0 \tag{91}$$

$$\mu^*(\frac{1}{2}\|x^*\|^2 - 1) = 0 \tag{92}$$

We can see that if $\mu^* = 0$, $c = 0$ as well, which contradicts our given condition $c \neq 0$, so we get $\mu^* > 0$

$$\frac{1}{2}\|x^*\|^2 - 1 = 0 \tag{93}$$

$$\|x^*\|^2 = 2 \tag{94}$$

$$\|a\mathbf{1}\|^2 = 2 \tag{95}$$

$$n|a|^2 = 2 \tag{96}$$

$$a = \pm\sqrt{\frac{2}{n}} \tag{97}$$

$$c^\top x + 8 = 4 \tag{98}$$

$$c^\top x = -4 \tag{99}$$

$$n(\boldsymbol{c}a) = -4 \tag{100}$$

$$\boldsymbol{c} = \frac{-4}{an} \tag{101}$$

$$c = \mp\frac{4}{\sqrt{2n}} \tag{102}$$

Please noted that here c and a have inverse signs.

c) KKT Condition For original problem:

$$min \quad f(x) \tag{103}$$

$$h(x^*) = 0 \tag{104}$$

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0 \tag{105}$$

KKT Condition For In-Eq problem:

$$g(x^*) = \frac{1}{2}\|h(x^*)\|^2 \leq 0 \tag{106}$$

$$\mu^* \geq 0 \tag{107}$$

$$\nabla f(x^*) + \mu^* h(x^*) = 0 \tag{108}$$

$$\frac{1}{2}\mu^*\|h(x^*)\|^2 = 0 \tag{109}$$

If $\mu^* = 0$:

$$\nabla f(x^*) = 0 h(x^*) maynotbe 0 \tag{110}$$

So $\mu^* > 0$:

$$\|h(x^*)\|^2 = 0 \nabla g(x^*) = \|h(x^*)\| = 0 h(x^*) = 0 \tag{111}$$

So $h(x^*)$ is not linearly independent, and x^* is not a regular point, so KKT theorem can not be applied on In-Eq problem.