Review of HW-G3-1, by Group 2

Problem 1.1

i) Correct. Minor errors:

- I think it is better to follow an order on mathematical development. Instead of showing up the final expression, write this one after the expression of Taylor expansion.

$$a \triangleq \frac{(x_2 - x_1)}{\|x_2 - x_1\|_2^2}$$

ii)Correct. A bit hard to read the second part where it goes from the equivalent to the equation of strong convexity.

iii) a) Correct. Mention in the second line respect to what the derivation is done.

iii) b) Correct.

iii) c) Correct. I would suggest to be more explicit for the reader and include additional steps.

iii) d) Incorrect.

F is convex. Then this term is zero

III.d)

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \boxed{\frac{\mu}{2} \|x_2 - x_1\|_2^2}$$

$$r(x_2) \geq r(x_1) + \nabla r(x_1)^T (x_2 - x_1) + \frac{L}{2} \|x_2 - x_1\|_2^2$$

$(+)$ $g(x) \triangleq f(x) + r(x)$

$\mu' = \mu + L$

$$g(x_2) \geq g(x_1) + \nabla g(x_1)^T (x_2 - x_1) + \frac{\mu'}{2} \|x_2 - x_1\|_2^2$$

so $g(x)$ μ'- strongly convex

Problem 1.2

a) Correct! Minor comment on readability:

$$\leq \int_0^1 \|\nabla f((1-t)x_1 + t x_2) - \nabla f(x_1)\| \, \|x_2 - x_1\| \, dt \quad \text{(Cauchy-Schwarz)}$$

$$\leq \int_0^1 Lt \, \|x_2 - x_1\|^2 \, dt \quad \text{(from (2))}$$

You could afford one more step here to show the reader how the “t” pops out of the norm.

b) Solution is correct but a bit hard to read.

$$\leq \nabla f(x_1)^T (x_1 - z) \quad \text{(Convexity)}$$

$$\nabla f(x_2)^T (z - x_2) + L/2 \, \|z - x_2\|^2 \quad \text{(from (A))}$$

There is a plus sign missing in the line break here.

$$= \nabla f(x_1)^T (x_1 - x_2) + \nabla f(x_1)^T (x_2 - z) - 1/L \, \nabla f(x_2)^T (\nabla f(x_2) - \nabla f(x_1)) + \tfrac{1}{2L} \|f(x_2) - \nabla f(x_1)\|^2$$

$$\left(\text{Note: } x_2 - z = 1/L (\nabla f(x_2) - \nabla f(x_1)) \right)$$

$$= \nabla f(x_1)^T (x_1 - x_2) - \tfrac{1}{L} \|\nabla f(x_2) - \nabla f(x_1)\|^2 + \tfrac{1}{2L} \|\nabla f(x_2) - \nabla f(x_1)\|^2$$

$$\therefore \; f(x_2) \geq \underbrace{f(x_1) + \nabla f(x_1)^T (x_2 - x_1)}_{B} + \tfrac{1}{2L} \|\nabla f(x_2) - \nabla f(x_1)\|^2. \quad \text{QED.}$$

Also here the steps are unclear, although they seem correct after investigation. Please spend more time on writing so that every step is easy to see for the reader.

c) Ok!

Problem 1.3

Correct.

I think it's better to explain how to define the equation for the convergence rate that you used in your solutions before this explanation:

If a sequence x_n converges to x_∞, we say that the convergence is <u>linear</u> if there exists an $r \in (0,1)$ such that

$$\frac{\| x_{n+1} - x_\infty \|}{\| x_n - x_\infty \|} \leq r \qquad \text{for all } n \text{ sufficiently large}$$

If the algorithm converges to x* as k→ ∞, the error e_k = ||x_k-x*|| denote how far the current value x_k is from the optimal value x*. To find the rate of convergence or convergence rate for an algorithm we investigate on this equation:

$$\lim_{k\to\infty} \frac{e_{k+1}}{{e_k}^p} = \lim_{k\to\infty} \frac{||x_{k+1} - x^*||}{||x_k - x^*||^p}$$

Problem 1.4

(a) Correct. I would suggest to check Newton-Method and realize that the Hessian matrix is diagonal. Then, the approach to decide between the Newton-Method and other convergence algorithms can changed due to the tradeoff between the number of iterations until reaching the optimal solution and the computing price.
(b) Correct.
(c) It seems correct.
(d) At exercise c, the Hessian goes from diagonal to non-diagonal due to the fact that we do not know how the Hessian matrix from r(x). Then, computing the Hessian matrix and its inverse can be relatively expensive. Therefore, I would suggest using another method such as quasi-newton methods.

Problem 1.5

Ok! Also needs more steps.

Peer-review of HW-G3-1

Here the assumption that $0 < \mu < L/2$ is needed to apply the inequality of L-smoothness of $f(x)$.