Optimization for Data Science, FS22 (Bernd Gärtner and Niao He)
Graded Assignment 4

Student Name, Legi Number: Deepak Narayanan, 21-959-622

Convergence Rate

Assignment 1

We are given $\delta_0 \geq 0$ and the sequence is given by $\delta_t - \delta_{t+1} \geq C\delta_t^{3/2}$. Firstly observe that $\delta_t \geq 0 \forall t$, since otherwise $\delta_t^{3/2}$ would not be defined properly. Secondly observe that, $\delta_t \geq \delta_{t+1} \because \delta_t \geq \delta_{t+1} + C\delta_t^{3/2}$, as $C\delta_t^{3/2} \geq 0$. Now we have,

$$\delta_t - \delta_{t+1} \geq C\delta_t^{3/2}$$

Dividing both sides by $\delta_t\delta_{t+1}$, we have,

$$\frac{1}{\delta_{t+1}} - \frac{1}{\delta_t} \geq C\frac{\sqrt{\delta_t}}{\delta_{t+1}}$$

Now using the fact that $x^2 - y^2 = (x-y)(x+y)$, we further have,

$$\left(\frac{1}{\sqrt{\delta_{t+1}}} - \frac{1}{\sqrt{\delta_t}}\right)\left(\frac{1}{\sqrt{\delta_{t+1}}} + \frac{1}{\sqrt{\delta_t}}\right) \geq C\frac{\sqrt{\delta_t}}{\delta_{t+1}}$$

This implies

$$\left(\frac{1}{\sqrt{\delta_{t+1}}} - \frac{1}{\sqrt{\delta_t}}\right) \geq C\frac{\sqrt{\delta_t}}{\delta_{t+1}\left(\frac{1}{\sqrt{\delta_{t+1}}} + \frac{1}{\sqrt{\delta_t}}\right)}$$

$$= C\frac{\delta_t\sqrt{\delta_{t+1}}}{\delta_{t+1}(\sqrt{\delta_t} + \sqrt{\delta_{t+1}})} = C\frac{\delta_t}{\sqrt{\delta_{t+1}}(\sqrt{\delta_t} + \sqrt{\delta_{t+1}})} = C\frac{1}{\sqrt{\frac{\delta_{t+1}}{\delta_t}} + \frac{\delta_{t+1}}{\delta_t}}$$

Now we know that

$$\delta_{t+1} \leq \delta_t \implies \sqrt{\delta_{t+1}} \leq \sqrt{\delta_t}$$

Using this we have,

$$\sqrt{\frac{\delta_{t+1}}{\delta_t}} + \frac{\delta_{t+1}}{\delta_t} \leq 2 \implies \frac{1}{\sqrt{\frac{\delta_{t+1}}{\delta_t}} + \frac{\delta_{t+1}}{\delta_t}} \geq \frac{1}{2}$$

Using this we have,

$$\left(\frac{1}{\sqrt{\delta_{t+1}}} - \frac{1}{\sqrt{\delta_t}}\right) \geq \frac{C}{2}$$

Now further, note that

$$1+\frac{3}{4}C\sqrt{\delta_0}\geq 1 \implies 1\geq\frac{1}{1+\frac{3}{4}C\sqrt{\delta_0}}$$

Now using this we further have,

$$\left(\frac{1}{\sqrt{\delta_{t+1}}}-\frac{1}{\sqrt{\delta_t}}\right)\geq\frac{C}{2}=\frac{C}{2}.1\geq\frac{C}{2(1+\frac{3}{4}C\sqrt{\delta_0})}=\frac{C}{2+\frac{3}{2}C\sqrt{\delta_0}}$$

Therefore we have,

$$\frac{1}{\sqrt{\delta_{t+1}}}-\frac{1}{\sqrt{\delta_t}}\geq\frac{C}{2+\frac{3}{2}C\sqrt{\delta_0}}$$

Now changing variable and taking sum from $k=0$ to $k=t-1$ on both sides we have,

$$\sum_{k=0}^{t-1}\frac{1}{\sqrt{\delta_{k+1}}}-\frac{1}{\sqrt{\delta_k}}\geq\sum_{k=0}^{t-1}\frac{C}{2+\frac{3}{2}C\sqrt{\delta_0}}$$

This implies

$$\frac{1}{\sqrt{\delta_t}}-\frac{1}{\sqrt{\delta_0}}\geq\frac{Ct}{2+\frac{3}{2}C\sqrt{\delta_0}} \implies \frac{1}{\sqrt{\delta_t}}\geq\frac{1}{\sqrt{\delta_0}}+\frac{Ct}{2+\frac{3}{2}C\sqrt{\delta_0}} \implies \sqrt{\delta_t}\leq\frac{1}{\frac{1}{\sqrt{\delta_0}}+\frac{Ct}{2+\frac{3}{2}C\sqrt{\delta_0}}}$$

Now squaring both sides, we have,

$$\delta_t\leq\left(\frac{1}{\frac{1}{\sqrt{\delta_0}}+\frac{Ct}{2+\frac{3}{2}C\sqrt{\delta_0}}}\right)^2$$

which is the required result. Now we can view the denominator on the RHS as

$$(a+bt)^2, a=\frac{1}{\sqrt{\delta_0}}, b=\frac{C}{2+\frac{3}{2}C\sqrt{\delta_0}}$$

Now

$$(a+bt)^2=a^2+b^2t^2+2abt$$

Now let L such that $0<L<b^2$ be a constant, then we have,

$$\frac{1}{(a+bt)^2}\leq\frac{1}{Lt^2}\forall t>=1$$

Now we have by definition of $\mathcal{O}-$notation that if

$$\exists M>0, f(x)\leq Mg(x), \forall x\geq x_0$$

then $f(x) = \mathcal{O}(g(x))$. In our case we have, $M = \frac{1}{L}, f(t) = \frac{1}{(a+bt)^2}, g(t) = \frac{1}{t^2}, t_0 = 1$. Therefore indeed,

$$\delta_t \leq \left(\frac{1}{\frac{1}{\sqrt{\delta_0}} + \frac{Ct}{2+\frac{3}{2}C\sqrt{\delta_0}}} \right)^2 = \mathcal{O}(1/t^2) \implies \delta_t = \mathcal{O}(1/t^2)$$

since if $f(x) \leq g(x) = \mathcal{O}(h(x))$, then $f(x) = \mathcal{O}(h(x))$. In this assignment we assume $\delta_t > 0 \forall t \geq 0$, since otherwise $1/\sqrt{\delta_t}$ will be undefined.

Cubic Regularization

Assignment 2

$$u(h(r)) = g^T h(r) + \frac{1}{2} h(r)^T H h(r) + \frac{M}{6}\|h(r)\|^3$$

$$v(r) = -\frac{1}{2} g^T \left(H + \frac{Mr}{2} I_d\right)^{-1} g - \frac{M}{12} r^3$$

We know $h(r) = -\left(H + \frac{Mr}{2} I_d\right)^{-1} g \implies g = -\left(H + \frac{Mr}{2} I_d\right) h(r)$. We also note that

$$H + \frac{Mr}{2} I_d = \left(H + \frac{Mr}{2} I_d\right)^T$$

Plugging in the value of g into $u(h(r))$, we have,

$$\begin{aligned} u(h(r)) = -h(r)^T \left(H + \frac{Mr}{2} I_d\right) h(r) + \frac{1}{2} h(r)^T H h(r) + \frac{M}{6}\|h(r)\|^3 = \\ -\frac{1}{2} h(r)^T H h(r) - \frac{Mr}{2} h(r)^T h(r) + \frac{M}{6}\|h(r)\|^3 = \\ -\frac{1}{2} h(r)^T H h(r) - \frac{Mr}{2}\|h(r)\|^2 + \frac{M}{6}\|h(r)\|^3 \end{aligned}$$

Similarly for $v(r)$), we have

$$\begin{aligned} v(r) = -\frac{1}{2} h(r)^T \left(H + \frac{Mr}{2} I_d\right) \left(H + \frac{Mr}{2} I_d\right)^{-1} \left(H + \frac{Mr}{2} I_d\right) h(r) - \frac{Mr^3}{12} = \\ -\frac{1}{2} h(r)^T \left(H + \frac{Mr}{2} I_d\right) h(r) - \frac{Mr^3}{12} = \\ -\frac{1}{2} h(r)^T H h(r) - \frac{Mr}{4}\|h(r)\|^2 - \frac{Mr^3}{12} \end{aligned}$$

Now taking difference, we have,

$$\begin{aligned} u(h(r)) - v(r) &= -\frac{1}{2} h(r)^T H h(r) - \frac{Mr}{2}\|h(r)\|^2 + \frac{M}{6}\|h(r)\|^3 - \left(-\frac{1}{2} h(r)^T H h(r) - \frac{Mr}{4}\|h(r)\|^2 - \frac{Mr^3}{12}\right) \\ &= -\frac{Mr}{4}\|h(r)\|^2 + \frac{Mr}{6}\|h(r)\|^3 + \frac{Mr^3}{12} = \frac{M}{12}\left(r^3 + 2\|h(r)\|^3 - 3r\|h(r)\|^2\right) \end{aligned}$$

Now we have the following identity.

$$\begin{aligned} x^3 + 2y^3 - 3xy^2 = x^3 + 2y^3 - 2xy^2 - xy^2 = x(x^2 - y^2) + 2y^2(y - x) \\ = x(x-y)(x+y) + 2y^2(y-x) = (y-x)(2y^2 - x(x+y)) = \\ (y-x)(2y^2 - xy - x^2) = \\ (x-y)(x^2 - y^2 + xy - y^2) = (x-y)((x-y)(x+2y)) = (x-y)^2(x+2y) \end{aligned}$$

Using this identity with $x = r, y = \|h(r)\|$, we have,

$$u(h(r)) - v(r) = \frac{M}{12} (r - \|h(r)\|)^2 (r + 2\|h(r)\|)$$

Now, taking derivative of $v(r)$, with respect to r, we have,

$$v'(r) = \frac{d(-\frac{1}{2}g^T \left(H + \frac{Mr}{2}I_d\right)^{-1} g - \frac{M}{12}r^3)}{dr}$$

$$= \frac{d(-\frac{1}{2}g^T \left(H + \frac{Mr}{2}I_d\right)^{-1} \left(-\left(H + \frac{Mr}{2}I_d\right)\right) h(r) - \frac{M}{12}r^3)}{dr}$$

$$= \frac{d(\frac{1}{2}g^T h(r) - \frac{Mr^3}{12})}{dr} = \frac{1}{2}g^T h'(r) - \frac{Mr^2}{4}$$

where $h'(r)$ is derivative of $h(r)$ with respect to r. Now we have,

$$g = -\left(H + \frac{Mr}{2}I_d\right) h(r)$$

Take derivative with respect to r on both sides we have,

$$0 = -Hh'(r) - \frac{M}{2}h(r) - \frac{Mr}{2}h'(r) \iff \left(H + \frac{Mr}{2}I_d\right) h'(r) = -\frac{M}{2}h(r)$$

This implies

$$h'(r) = -\frac{M}{2}\left(H + \frac{Mr}{2}I_d\right)^{-1} h(r)$$

Now plugging this back into $v'(r)$, along with the value of g, we have,

$$v'(r) = \frac{1}{2}g^T h'(r) - \frac{Mr^2}{4} = \frac{1}{2}\left(-h(r)^T \left(H + \frac{Mr}{2}I_d\right)\right)\left(-\frac{M}{2}\left(H + \frac{Mr}{2}I_d\right)^{-1}\right) h(r) - \frac{Mr^2}{4}$$

Therefore we have,

$$v'(r) = \frac{M}{4}\|h(r)\|^2 - \frac{Mr^2}{4} = \frac{M}{4}(\|h(r)\|^2 - r^2)$$

Now

$$u(h(r)) - v(r) = \frac{M}{12}(r - \|h(r)\|)^2 (r + 2\|h(r)\|) = \frac{M}{12}(r - \|h(r)\|)^2 (r + 2\|h(r)\|)\frac{(r + \|h(r)\|)^2}{(r + \|h(r)\|)^2}$$

$$= \frac{M}{12}\left(r^2 - \|h(r)\|^2\right)^2 (r + 2\|h(r)\|)\frac{1}{(r + \|h(r)\|)^2} =$$

$$\frac{M^2}{16.12\frac{M}{16}}\left(r^2 - \|h(r)\|^2\right)^2 (r + 2\|h(r)\|)\frac{1}{(r + \|h(r)\|)^2} = \frac{4}{3M}\frac{(r + 2\|h(r)\|)v'(r)^2}{(r + \|h(r)\|)^2}$$

which is the required proof needed. Note that all the terms are non-negative which implies,

$$u(h(r)) - v(r) \geq 0$$

We are given that the supremum is attained, that is the maximum and supremum are the same. We then have, $v'(r^*) = 0$, which implies, $u(h(r^*)) = v(r^*) = \sup_{r \in \mathcal{D}} v(r)$.

We have $v'(r^*) = 0$, for the following reasons. Firstly we note that,

$$\begin{aligned} v''(r) = \frac{M}{4}(2h(r)^T h'(r) - 2r) &= \frac{M}{2}\left(-\left(H + \frac{Mr}{2}I_d\right)^{-1} g\right)^T \left(-\frac{M}{2}\left(H + \frac{Mr}{2}I_d\right)^{-1} h(r)\right) - \frac{Mr}{2} \\ &= \frac{M^2}{4}\left(g^T\left(H + \frac{Mr}{2}I_d\right)^{-1}\right)\left(H + \frac{Mr}{2}I_d\right)^{-1}\left(-\left(H + \frac{Mr}{2}I_d\right)^{-1} g\right) - \frac{Mr}{2} \\ &= -\frac{M^2}{4}\left(g^T\left(H + \frac{Mr}{2}\right)^{-3} g\right) - \frac{Mr}{2}I_d \leq 0 \forall r \in \mathcal{D} \end{aligned}$$

since inverse of symmetric positive definite matrices is positive definite and powers of symmetric positive definite matrices is positive definite. This implies as long as the domain $\mathcal{D}$ is convex, $v(r)$ is a concave function over $\mathcal{D}$. Note that we are given the supremum is attained, indeed giving us the maximum, and it is attained for $r^* > 0$. This implies as long as the domain $\mathcal{D}$ is convex and since we know $v(r)$ is concave, the derivative at the maximizer is zero. Therefore what remains to be shown is that the domain $\mathcal{D}$, is convex. Indeed, suppose

$$r_1, r_2 \in \mathcal{D} \implies H + \frac{Mr_1}{2}I_d \succ 0, H + \frac{Mr_2}{2}I_d \succ 0$$

Now consider a convex combination, $\lambda \in [0, 1]$,

$$v = \lambda r_1 + (1 - \lambda) r_2$$

First observe $v \geq 0$. Secondly, we have

$$\begin{aligned} &H + \frac{Mv}{2}I_d = H + \frac{M(\lambda r_1 + (1-\lambda)r_2)}{2}I_d = \lambda H + (1-\lambda)H + \lambda\frac{Mr_1}{2}I_d + (1-\lambda)\frac{Mr_2}{2}I_d \\ &= \lambda\left(H + \frac{Mr_1}{2}I_d\right) + (1-\lambda)\left(H + \frac{Mr_2}{2}I_d\right) \succ 0 \left(\because H + \frac{Mr_1}{2}I_d \succ 0, H + \frac{Mr_2}{2}I_d \succ 0\right) \end{aligned}$$

This implies $v \in \mathcal{D}$. Therefore, $\mathcal{D}$ is indeed convex and hence $v'(r^*) = 0$.

Now what remains to show is that $\inf_{h \in \mathbb{R}^d} u(h) \geq \sup_{r \in \mathcal{D}} v(r)$, since if we show this, we would indeed have proven

$$\inf_{h \in \mathbb{R}^d} u(h) = \sup_{r \in \mathcal{D}} v(r)$$

as $h = h(r^*) = -\left(H + \frac{Mr^*}{2} I_d\right)^{-1} g$, would imply that the inequality becomes an equality.

Indeed we have,

$$\inf_{h \in \mathbb{R}^d} u(h) = \inf_{h \in \mathbb{R}^d} g^T h + \frac{1}{2} h^T H h + \frac{M}{6} \|h\|^3 = \inf_{u \in \mathbb{R}^d, \tau = \|h\|^2} g^T h + \frac{1}{2} h^T H h + \frac{M}{6} \tau^{3/2}$$

$$= \inf_{h \in \mathbb{R}^d, \tau \in \mathbb{R}} \sup_{r \in \mathbb{R}} \left\{ g^T h + \frac{1}{2} h^T H h + \frac{M}{6} \tau^{3/2} + \frac{M}{4} r(\|h\|^2 - \tau) \right\} \geq$$
$$\inf_{h \in \mathbb{R}^d, \tau \in \mathbb{R}} \sup_{r \in \mathcal{D}} \left\{ g^T h + \frac{1}{2} h^T H h + \frac{M}{6} \tau^{3/2} + \frac{M}{4} r(\|h\|^2 - \tau) \right\}$$

where the inequality comes from the fact that supremum over a superset of $\mathcal{D}$ is bigger than the supremum over $\mathcal{D}$.

Now consider a function $f(x, y)$ defined over $x \in X$, $y \in Y$. We then have,

$$g(y) = \inf_{x \in X} f(x, y) \leq f(x, y) \implies \sup_{y \in Y} g(y) \leq \sup_{y \in Y} f(x, y) \implies \tag{1}$$

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \sup_{y \in Y} f(x, y) \implies \sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y) \tag{2}$$

Using this inequality, and swapping the supremum and infimum, we have,

$$\inf_{h \in \mathbb{R}^d} u(h) \geq \inf_{h \in \mathbb{R}^d, \tau \in \mathbb{R}} \sup_{r \in \mathcal{D}} \left\{ g^T h + \frac{1}{2} h^T H h + \frac{M}{6} \tau^{3/2} + \frac{M}{4} r(\|h\|^2 - \tau) \right\}$$
$$\geq \sup_{r \in \mathcal{D}} \inf_{h \in \mathbb{R}^d, \tau \in \mathbb{R}} \left\{ g^T h + \frac{1}{2} h^T H h + \frac{M}{6} \tau^{3/2} + \frac{M}{4} r(\|h\|^2 - \tau) \right\}$$

Now consider

$$\inf_{h \in \mathbb{R}^d, \tau \in \mathbb{R}} \left\{ g^T h + \frac{1}{2} h^T H h + \frac{M}{6} \tau^{3/2} + \frac{M}{4} r(\|h\|^2 - \tau) \right\} = \inf_{h \in \mathbb{R}^d, \tau \in R} k(h, \tau)$$

$k(h, \tau)$ is a sum of function of two variables h and τ and each of the functions in their individual variables are convex (as we will show below). Therefore, indeed their sum over the joint domain is convex from the basic definition of convexity. We will further show that the minimum indeed exists for this above function and therefore the infimum is indeed the minimum and equals $v(r)$, which would then conclude the proof.

Now indeed, we have individual functions in h as

$$f(h) = g^T h + \frac{1}{2} h^T H h + \frac{Mr}{4} \|h\|^2$$

Its gradient and subsequently hessian is

$$f'(h) = g + Hh + \frac{Mr}{2} h$$

$$f''(h) = H + \frac{Mr}{2} I_d$$

Now for $r \in \mathcal{D}$, the second derivative is positive definitive, implying $f(h)$ is convex. Now for $f(\tau)$, we have

$$f(\tau) = \frac{M}{6}\tau^{3/2} - \frac{M}{4} r\tau$$

$$f'(\tau) = \frac{M}{4}\tau^{1/2} - \frac{M}{4} r$$

$$f''(\tau) = \frac{M}{8}\frac{1}{\sqrt{\tau}} > 0 \forall \tau > 0$$

Therefore both functions are convex and from basic definition of convexity we can check that the sum is convex. Therefore the minimum in attained when the derivatives are zero. This implies

$$f'(h) = 0 \implies h = -\left(H + \frac{Mr}{2}\right)^{-1} g$$
$$f'(\tau) = 0 \implies \tau = r^2$$

Plugging this back, we have

$$g^T h + \frac{1}{2} h^T H h + \frac{M}{6}\tau^{3/2} + \frac{M}{4} r(\|h\|^2 - \tau) = g^T h + \frac{1}{2} h^T H h + \frac{M}{6} r^3 + \frac{M}{4} r(\|h\|^2 - r^2)$$
$$= g^T h + \frac{1}{2} h^T H h + \frac{M}{6} r^3 - \frac{M}{4} r^3 + \frac{M}{4} r\|h\|^2 = g^T h + \frac{1}{2} h^T H h - \frac{M}{12} r^3 + \frac{M}{4} r h^T h$$
$$= g^T h + \frac{1}{2} h^T H h + \frac{M}{4} r h^T I_d h - \frac{M}{12} r^3 = g^T h + \frac{1}{2} h^T (H + \frac{Mr}{2} I_d) h - \frac{M}{12} r^3 = \frac{g^T h}{2} - \frac{M}{12} r^3$$

Now plugging in value of h,

$$\frac{g^T h}{2} - \frac{M}{12} r^3 = -\frac{1}{2} g^T \left(H + \frac{Mr}{2} I_d\right)^{-1} g - \frac{M}{12} r^3 = v(r)$$

Therefore we have,

$$\inf_{h \in \mathbb{R}^d, \tau \in \mathbb{R}} \left\{ g^T h + \frac{1}{2} h^T H h + \frac{M}{6}\tau^{3/2} + \frac{M}{4} r(\|h\|^2 - \tau) \right\} = v(r)$$

This implies,

$$\inf_{h \in \mathbb{R}^d} u(h) \geq \sup_{r \in \mathcal{D}} v(r)$$

But we showed that $u(h(r^*)) = \sup_{r \in \mathcal{D}} v(r) \implies \inf_{h \in \mathbb{R}^d} u(h) = \sup_{r \in \mathcal{D}} v(r)$

Moreau Envelope Approximation

Assignment 3

$$f_\mu(x) = \min_y \left\{ f(y) + \frac{1}{2\mu}\|x-y\|^2 \right\}$$

Firstly observe that,

$$\min_y \left\{ f(y) + \frac{1}{2\mu}\|x-y\|^2 \right\} \leq f(z) + \frac{1}{2\mu}\|x-z\|^2$$

where z belongs to the domain of the optimization problem. In particular, for $z = x$, we have,

$$\min_y \left\{ f(y) + \frac{1}{2\mu}\|x-y\|^2 \right\} \leq f(x) + \frac{1}{2\mu}\|x-x\|^2 = f(x) \implies f(x) - f_\mu(x) \geq 0$$

Suppose the optimal value of the above optimization problem is attained at y^*. We then have,

$$f_\mu(x) = f(y^*) + \frac{1}{2\mu}\|x-y^*\|^2$$

Now taking the difference we have,

$$f(x) - f_\mu(x) =$$

$$f(x) - f(y^*) - \frac{1}{2\mu}\|x-y^*\|^2 \leq |f(x) - f(y^*)| - \frac{1}{2\mu}\|x-y^*\|^2 \leq L\|x-y^*\| - \frac{1}{2\mu}\|x-y^*\|^2$$

where the last inequality follows from lipschitzness of f. Now suppose,

$$t = \|x-y^*\|$$

Then the RHS is a quadratic function in t.

$$f(t) = Lt - \frac{1}{2\mu}t^2$$

This is a concave quadratic with maximum value attained at

$$f'(t) = 0 \implies t = L\mu$$

Now the maximum value is

$$L.L\mu - \frac{1}{2\mu}L^2\mu^2 = \frac{L^2\mu}{2}$$

Therefore we have,

$$f(x) - f_\mu(x) \leq \max \left\{ L\|x-y^*\| - \frac{1}{2\mu}\|x-y^*\|^2 \right\} = \frac{L^2\mu}{2} < L^2\mu$$

Combining the inequalities, we have,

$$0 \leq f(x) - f_\mu(x) \leq L^2\mu$$

Nonexistence of Saddle Point

Assignment 4

$$f(x, y) = y/(x + y)$$

$X = [1, \infty), Y = [1, \infty)$ Now consider a new domain $X' = (0, \infty)$ for x. Then we have over X' for a fixed $y \in Y$,

$$\frac{\partial f}{\partial x} = -\frac{y}{(x+y)^2}$$
$$\frac{\partial^2 f}{\partial x^2} = \frac{2y}{(x+y)^3}$$

The above second derivative is positive for all $x \in X'$. This implies f is convex in x for a fixed y over X'. Now since it is convex over X', it is also convex over a subset of X', which in our required case is X. Therefore f is convex in x over X, for a fixed $y \in Y$.

Now consider $Y' = (0, \infty)$. Then over this new domain, for a fixed $x \in X$, we have

$$\frac{\partial f}{\partial y} = \frac{x}{(x+y)^2}$$
$$\frac{\partial^2 f}{\partial y^2} = -\frac{2x}{(x+y)^3}$$

Since the second derivative is negative for all $y \in Y'$ and for a fixed $x \in X$, f is concave in y, over Y' for a fixed $x \in X$. Since it is concave over all of Y', it is also concave over a subset of Y', which in our required case is Y itself. Therefore f is concave in y over Y for fixed $x \in X$.
Therefore f is convex-concave is $X \times Y$.

Assignment 5

Claim 1: (x^*, y^*) is a saddle point considering minimization in x and maximization in y iff the supremum in $\sup_{y\in Y}\inf_{x\in X} f(x,y)$ is attained at y^* and the infimum of $\inf_{x\in X}\sup_{y\in Y} f(x,y)$ is attained at x^* and these extrema are equal, that is, $\inf_{x\in X}\sup_{y\in Y} f(x,y) = \sup_{y\in Y}\inf_{x\in X} f(x,y)$.

Proof: Now suppose (x^*, y^*) is a saddle point. We then have over minimization in x and maximization in y, that

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*)\,\forall x \in X, y \in Y$$

Firstly note that,

$$f(x^*, y^*) = \inf_{x\in X} f(x, y^*) = \sup_{y\in Y} f(x^*, y)$$

This further implies

$$f(x^*, y^*) = \inf_{x\in X} f(x, y^*) \leq \sup_{y\in Y}\inf_{x\in X} f(x,y)$$

$$f(x^*, y^*) = \sup_{y\in Y} f(x^*, y) \geq \inf_{x\in X}\sup_{y\in Y} f(x,y)$$

Therefore if (x^*, y^*) is a saddle point, we have

$$\inf_{x\in X}\sup_{y\in Y} f(x,y) \leq \sup_{y\in Y}\inf_{x\in X} f(x,y)$$

But we know from Equations (1) and (2) in page 7 that

$$\sup_{y\in Y}\inf_{x\in X} f(x,y) \leq \inf_{x\in X}\sup_{y\in Y} f(x,y)$$

For both of these to hold, we therefore must have,

$$\inf_{x\in X}\sup_{y\in Y} f(x,y) = \sup_{y\in Y}\inf_{x\in X} f(x,y)$$

this implies that the supremum and infimum attain their extreme at exactly at the required points.

Now for the other direction, we have,

$$\sup_{y\in Y}\inf_{x\in X} f(x,y) = \inf_{x\in X} f(x, y^*) \leq f(x, y^*) \forall x \in X$$

$$\inf_{x\in X}\sup_{y\in Y} f(x,y) = \sup_{y\in Y} f(x^*, y) \geq f(x^*, y) \forall y \in Y$$

We also know extremal values are same. This implies,

$$f(x^*, y) \leq \sup_{y \in Y} f(x^*, y) = \inf_{x \in X} f(x, y^*) \leq f(x, y^*)$$

where the supremum is at least $f(x^*, y^*)$ and the infimum is at most $f(x^*, y^*)$ which implies

$$f(x^*, y^*) \leq \sup_{y \in Y} f(x^*, y) = \inf_{x \in X} f(x, y^*) \leq f(x^*, y^*)$$

which implies

$$\sup_{y \in Y} f(x^*, y) = \inf_{x \in X} f(x, y^*) = f(x^*, y^*)$$

following which we have,

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*) \, \forall x \in X, y \in Y$$

making (x^*, y^*) a saddle point.

Now for the question at hand, we have,

$$f(x, y) = y/(x + y)$$

Claim 2: f is a monotonically decreasing function in x for a given $y \in Y$, and f is a monotonically increasing function in y for a given $x \in X$.

Proof

Part 1: Suppose $x_1, x_2 \in X, x_1 \leq x_2$. We then have, $f(x_1) - f(x_2)$,

$$f(x_1) - f(x_2) = y/(x_1 + y) - y/(x_2 + y) = y(x_2 - x_1)/((x_1 + y)(x_2 + y)) \geq 0 (\because x_2 \geq x_1)$$

Similarly, Suppose $y_1, y_2 \in Y, y_1 \leq y_2$. We then have, $f(y_1) - f(y_2)$

$$f(y_1) - f(y_2) = y_1/(x + y_1) - y_2/(x + y_2) = (y_1 - y_2)x/((x + y_1)(x + y_2)) \leq 0 \because (y_1 \leq y_2)$$

Because of monotonicity, we have the infimum and supremum being the limits. Therefore we have,

$$\inf_{x \in X} \sup_{y \in Y} y/(x + y) = \inf_{x \in X} \lim_{y \to \infty} y/(x + y) = 1$$

Similarly,

$$\sup_{y \in Y} \inf_{x \in X} y/(x + y) = \sup_{y \in Y} \lim_{x \to \infty} y/(x + y) = 0$$

Therefore we have,

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \neq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

and by Claim 1, there is no saddle point.