CSE 592: Convex Optimization
HW 4

Bhushan Sonawane
SBU ID: 111511679

1 Mirror Descent

1.1 Calculate the Fenchel conjugate of $\Psi(\mathbf{x})$

Given function $\Psi(\mathrm{x}) = \frac{1}{2(p-1)} \|x\|_p^2$, where 1 < p ≤ 2

From Cauchy-Schwarz inequality for dual norm, we can write

$$y^T x \leq \|y\|_q \|x\|_p$$

subtracting $\frac{1}{2(p-1)} \|x\|_p^2$ from both sides, we get

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq \|y\|_q \|x\|_p - \frac{1}{2(p-1)} \|x\|_p^2 \tag{1}$$

for all x, The right hand side function is a quadratic function of $\|x\|$. We can find maximum value by taking derivative of R.H.S. and putting to zero.

$$\|y\|_q - \frac{1}{p-1} \|x\|_p = 0$$

$$\|x\|_p = (p-1) \|y\|_q \tag{2}$$

putting value of $\|x\|_p$ into equation (1)

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq \|y\|_q (p-1) \|y\|_q - \frac{(p-1)^2}{2(p-1)} \|y\|_q^2$$

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq \|y\|_q (p-1) \|y\|_q - \frac{(p-1)}{2} \|y\|_q^2$$

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq (p-1) \|y\|_q^2 - \frac{(p-1)}{2} \|y\|_q^2$$

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq \frac{(p-1)}{2} \|y\|_q^2 \tag{3}$$

but, we know that, $\frac{1}{p} + \frac{1}{q} = 1$

$$\frac{1}{p} = 1 - \frac{1}{q}$$

$$\frac{1}{p} = \frac{q-1}{q}$$

$$p = \frac{q}{q-1} \tag{4}$$

We can simplify $\dfrac{p-1}{2}$ from equation (3) using value of p obtained in equation (4) as follows:

$$\frac{p-1}{2} = \frac{\dfrac{q}{q-1} - 1}{2}$$

$$\frac{p-1}{2} = \frac{\dfrac{q-q+1}{q-1}}{2}$$

$$\frac{p-1}{2} = \frac{1}{2(q-1)} \tag{5}$$

putting value of $\dfrac{p-1}{2}$ from equation (5) into equation (3) we get,

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq \frac{1}{2(q-1)} \|y\|_q^2 \tag{6}$$

To show other inequality, let x be any vector $y^T x = \|y\|_q \|x\|_p$, scaled so that, $\|x\|_p = (p-1)\|y\|_q$ Hence,

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 = (p-1)\|y\|_q^2 - \frac{(p-1)^2}{2(p-1)} \|y\|_q^2$$

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 = (p-1)\|y\|_q^2 - \frac{(p-1)}{2} \|y\|_q^2$$

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 = \frac{(p-1)}{2} \|y\|_q^2$$

and from equation (5), we can write,

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 = \frac{1}{2(q-1)} \|y\|_q^2 \tag{7}$$

From equation (7),

$$\Psi^*(y) \geq \frac{1}{2(q-1)} \|y\|_q^2 \tag{8}$$

From equation (8),

$$\Psi^*(y) = \frac{1}{2(q-1)} \|y\|_q^2 \tag{9}$$

1.2 Calculate the gradient of Ψ(x)

Given function Ψ(x) $= \dfrac{1}{2(p-1)} \|x\|_p^2$, where 1 < p $\leq$ 2

$$\Psi(x) = \frac{1}{2(p-1)} \|x\|_p^2$$

can be written as,

$$\Psi(x) = \frac{1}{2(p-1)} \left(\sum_{i=1}^{d} |x_i|^p\right)^{\dfrac{2}{p}}$$

For i from 1 to d

$$\frac{\partial \Psi(x_i)}{\partial x_i} = \frac{1}{2*(p-1)} \frac{2}{p} \left(\sum_{i=1}^{d} |x_i|^p\right)^{\dfrac{2}{p}-1} p * |x_i|^{p-1} sign(x_i)$$

$$\frac{\partial\Psi(x_i)}{\partial x_i} = \frac{1}{(p-1)} * (\sum_{i=1}^{d} |x_i|^p)^{\frac{2-p}{p}} |x_i|^{p-1} sign(x_i)$$

$$\frac{\partial\Psi(x_i)}{\partial x_i} = \frac{1}{(p-1)} \|x\|_p^{(2-p)} |x_i|^{p-1} sign(x_i) \tag{10}$$

We can represent equation (10) in vectorized form as follows:

$$\nabla\Psi(x) = \frac{1}{(p-1)} \|x\|_p^{(2-p)} diag(|x|^{p-1}) sign(x) \tag{11}$$

Where,

$$sign(x) = \begin{bmatrix} sign(x_i) \\ \vdots \\ sign(x_d) \end{bmatrix}$$

and

$$diag(x) = \begin{bmatrix} |x_0|^{p-1} & 0 & 0 & \dots & 0 \\ 0 & |x_1|^{p-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & |x_d|^{p-1} \end{bmatrix}$$

1.3 Calculate the gradient of $\Psi^*(x)$

from equation (8), we know that,

$$\Psi^*(y) = \frac{1}{(q-1)} \|y\|_q^2$$

Also, in question 1.2, we have computed gradient of $\Psi(\mathrm{x}) = \frac{1}{2(p-1)} \|x\|_p^2$ which is similar to this problem where p is replaced by q.

Hence, from equation (9) and (11), we can write,

$$\nabla\Psi^*(x) = \frac{1}{(q-1)} \|x\|_q^{(2-q)} diag(|x|^{q-1}) sign(x) \tag{12}$$

Where,

$$sign(x) = \begin{bmatrix} sign(x_i) \\ \vdots \\ sign(x_d) \end{bmatrix}$$

and

$$diag(x) = \begin{bmatrix} |x_0|^{p-1} & 0 & 0 & \dots & 0 \\ 0 & |x_1|^{p-1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & |x_d|^{p-1} \end{bmatrix}$$

1.4 Put all togheter and write the update rule for Mirror Descent. Do simplify what you get, in order to have something "nice".

Update rule for mirror descent is

$$x_{t+1} = \nabla\Psi^*(\nabla\Psi(x_t) - \eta_t g_t) \tag{13}$$

From equation 10,

$$\nabla\Psi(x_t) - \eta_t g_t = \frac{1}{2(p-1)} \|x_t\|^{(2-p)} \, diag(|x_t|^{p-1}) - \eta_t g_t \tag{14}$$

From equation (12), (13) and (14),

$$x_{t+1} = \nabla\Psi^* \left(\frac{1}{2(p-1)} \|x_t\|^{(2-p)} \, diag(|x_t|^{p-1}) - \eta_t g_t \right) \tag{15}$$

$$\begin{aligned} & \frac{1}{(q-1)} \left\| \frac{1}{p-1} \|x_t\|_p^{2-p} \, diag(|x_t|^{(p-1)}) sign(x_t) - \eta_t g_t \right\|_q^{2-q} \\ x_{t+1} = \; & diag\left(\left| \frac{1}{(p-1)} \|x_t\|_p^{2-p} \, diag(|x_t|^{p-1}) sign(x_t) - \eta_t g_t \right|^{q-1} \right) \\ & sign\left(\frac{1}{(p-1)} \|x_t\|_p^{2-p} \, diag(|x_t|^{p-1}) sign(x_t) - n_t g_t \right) \end{aligned}$$

Where, diag and sign are similar functions as used in Q1.2 and Q1.3

2 Programming Exercise

2.1 Number of Iterations

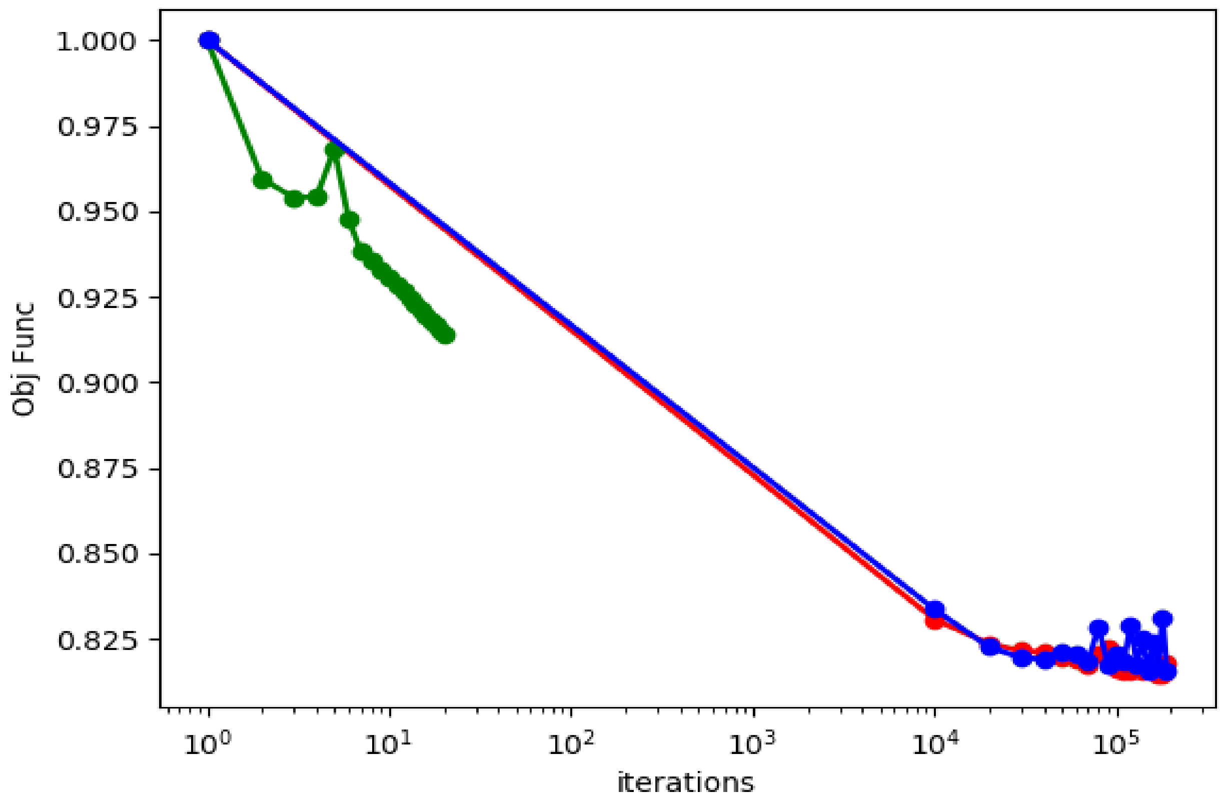

Figure 1: Number of Iterations

2.2 Objective function

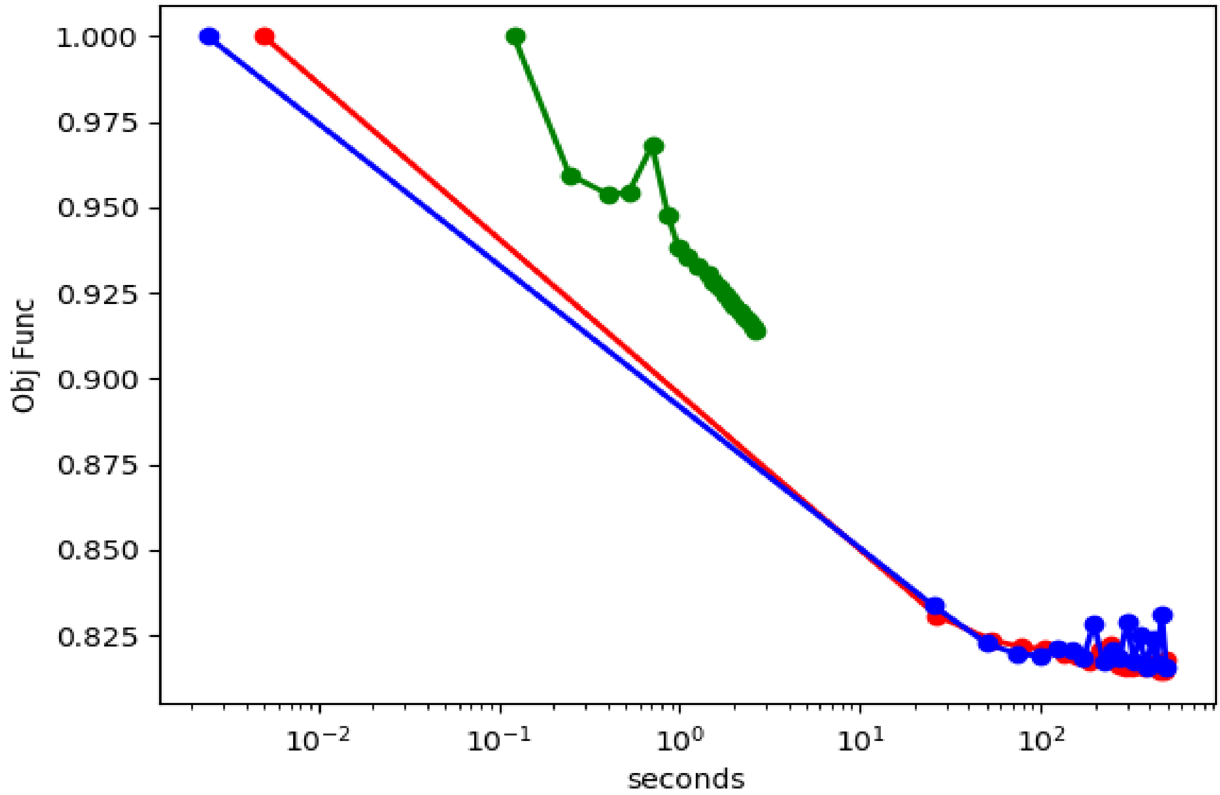

Figure 2: Objective function

2.3 Objective values

Solution found by stochastic subgradient descent:
matrix([-0.34586651, -0.0214911 , -0.01375237, -0.50988799, 0.03440775, 0.77621803, 0.03897345, 0.01697494, 0.12081295, 0.17626005, 0.01876806, 0.01353644, -0.01936219, 0.14952631, 0.03052357, -0.01309344, -0.04696571, 0.25944945, -0.0150916 , 0.00723274, -0.02510871, -0.22650021, 0.90452166, 1.18686794, 0.63974686, -1.48135024, 2.34533704, -3.87435276]))
Objective function: 0.8169245484868474

Solution found by stochastic adagrad:
matrix([-3.71737428e-01, -2.55940843e-02, -3.91759431e-02, -4.84993884e-01, -2.78116594e-03, 8.30884512e-01, -4.45414737e-02, 2.23335757e-02, 1.18274793e-01, 1.82187554e-01, -2.17690886e-02, 7.36946715e-03, -3.76505263e-02, 1.74162382e-01, 3.20339768e-02, 3.51520893e-02, 8.80603646e-03, 2.34145338e-01, 2.04078419e-02, 8.09923285e-03, -4.10348160e-02, -2.83439757e-01, 8.48389921e-01, 1.15675796e+00, 6.17861389e-01, -1.57458258e+00, 2.88194731e+00, -4.46994788e+00]))
Objective function: 0.8160413002545496

Solution found by subgradient descent: matrix([-0.02582542, -0.01474747, -0.00618211, -0.21100792, 0.00319675, 0.22824581, -0.00754905, 0.00146226, 0.14210423, 0.12311359, 0.01842417, -0.005359 , -0.05565465, 0.12481449, 0.00288305, 0.01901484, 0.00618816, 0.19496673, -0.01605185, 0.0104227 , 0.05892398, 0.05372047, 0.14753203, 0.20979349, 0.08846406, -0.39405411, -0.04870572, -0.1612745]))
Objective function: 0.9125785012725371