Convex Optimizations – Homework 4

Q1. Mirror Descent

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

We know that $\Psi(x) = \frac{1}{2(p-1)} \|x\|_p^2$ where $1 < p \leq 2$

Using Cauchy Shwarz inequality we get,

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

Here, $\|.\|_p$ is the p-norm and $\|.\|_*$ is its dual norm.

Eqn (1) is true for all x, the right-hand side of the equation is quadratic in terms of $\|x\|_p$

To maximize the RHS of the inequality, we can take the derivative of the right part of the equation w.r.t $\|x\|_p$ and equate it to 0.

$$\therefore \ \|y\|_* - \frac{2}{2(p-1)} \|x\|_p = 0$$

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

By substituting the value of (2) in (1), we get,

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq \|y\|_* \left((p-1)\|y\|_*\right) - \frac{1}{2(p-1)} \left((p-1)\|y\|_*\right)^2$$

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

We know that dual norm follows the property,

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

where $\|.\|_q$ is the dual norm of $\|.\|_p$.

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

Substituting (4) in (3),

$$y^T x - \frac{1}{2(p-1)} \|x\|_p^2 \leq \frac{1}{2(q-1)} \|y\|_q^2 \tag{5}$$

This shows that

$$\Psi^*(y) \leq \frac{1}{2(q-1)} \|y\|_q^2 \tag{6}$$

To show the other inequality, let x be any vector with $y^T x = \|y\|_q \|x\|$ scaled so that $\|x\|_p = (p-1)\|y\|_q$. Then we have for this x,

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

which shows that

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

From (7) and (8) we can conclude that,

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

1.2 Calculate the gradient of $\Psi(x)$

We know that $\Psi(x) = \frac{1}{2(p-1)}\ \|x\|_p^2$ where $1 < p \leq 2$

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

Differentiating $\Psi(x)$ w.r.t x_{i} ,

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

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

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

$$\therefore\ \frac{\delta\Psi}{\delta x_i} = \frac{|x_i|^{p-1}}{p-1} \left(\|x\|_p\right)^{2-p} sign(x_i) \qquad \forall i = 1 \ldots d$$

Converting the above equation in vector form

$$\therefore\ \boldsymbol{\nabla\Psi} = \frac{\|\boldsymbol{x}\|_{\boldsymbol{p}}^{\boldsymbol{2-p}}}{\boldsymbol{p-1}} * \boldsymbol{diag}(|\boldsymbol{x}|^{\boldsymbol{p-1}}) * \boldsymbol{sign}(\boldsymbol{x}) \tag{10}$$

Where $diag(|x|^{p-1})$ is a matrix that has terms of $|x|^{p-1}$ as its diagonal elements and the rest of the terms are zero.

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

and $sign(x)$ is a vector of the form,

$$sign(x) = \begin{bmatrix} sign(x_1) \\ \vdots \\ sign(x_d) \end{bmatrix}$$

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

From (9) know that $\Psi^*(x) = \frac{1}{2(q-1)} \|x\|_q^2$

$$\therefore \Psi^*(x) = \frac{1}{2\,(q-1)} \left(\sum_{i=1}^{d} |x_i|^q\right)^{\frac{2}{q}}$$

Differentiating $\Psi^*(x)$ w.r.t x_i ,

$$\frac{\delta \Psi^*}{\delta x_i} = \frac{1}{2\,(q-1)} * \frac{2}{q} * \left(\sum_{i=1}^{d} |x_i|^q\right)^{\frac{2}{q}-1} * q * |x_i|^{q-1} * \frac{\delta}{\delta x_i} |x_i|$$

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

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

$$\therefore \frac{\delta \Psi^*}{\delta x_i} = \frac{|x_i|^{q-1}}{q-1} \left(\|x\|_q\right)^{2-q} sign(x_i) \qquad \forall i = 1 \ldots d$$

Converting the above equation in vector form

$$\therefore \boldsymbol{\nabla \Psi^*} = \frac{\|\boldsymbol{x}\|_{\boldsymbol{q}}^{\boldsymbol{2-q}}}{\boldsymbol{q-1}} * \boldsymbol{diag}(|\boldsymbol{x}|^{\boldsymbol{q-1}}) * \boldsymbol{sign}(\boldsymbol{x}) \qquad (11)$$

Where $diag(|x|^{q-1})$ is a matrix that has terms of $|x|^{q-1}$ as its diagonal elements and the rest of the terms are zero.

$$diag(|x|^{q-1}) = \begin{bmatrix} |x_1|^{q-1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & |x_d|^{q-1} \end{bmatrix}$$

and $sign(x)$ is a vector of the form,

$$sign(x) = \begin{bmatrix} sign(x_1) \\ \vdots \\ sign(x_d) \end{bmatrix}$$

1.4 Write the update rule for mirror descent

The update rule for Mirror Descent can be written as,

$$X_{t+1} = \nabla\Psi^*(\nabla\Psi(X_t) - \eta_t g_t) \quad \ldots (given)$$

From (10) we get,

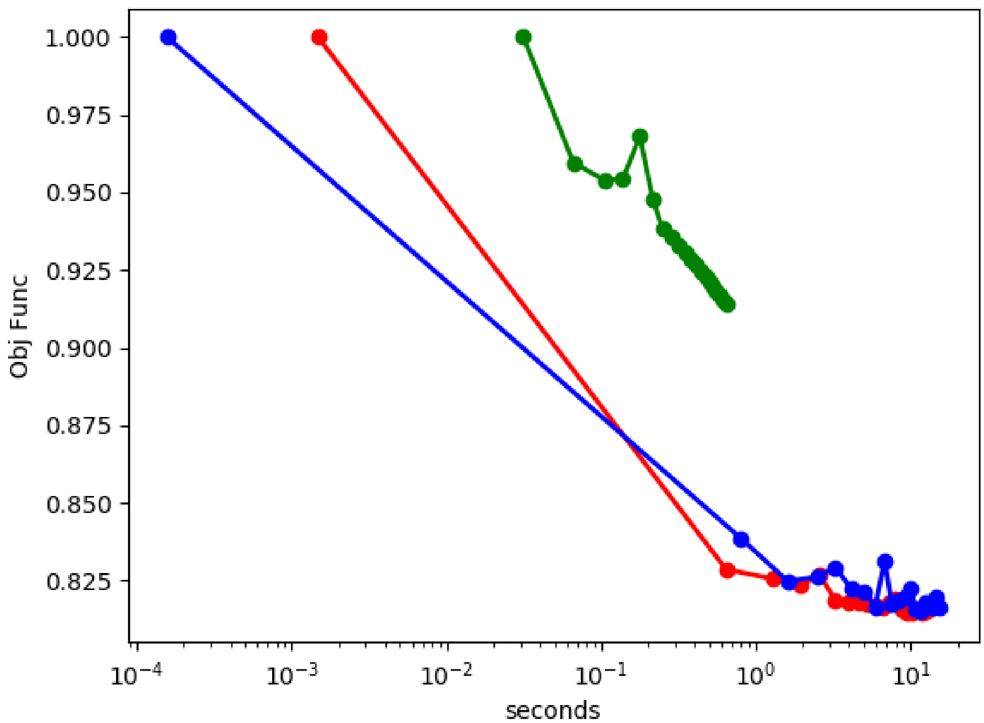
$$\therefore X_{t+1} = \nabla\Psi^*\left(\frac{\|X_t\|_p^{2-p}}{p-1} * diag(|X_t|^{p-1}) * sign(X_t) - \eta_t g_t\right)$$

From (11) we get,

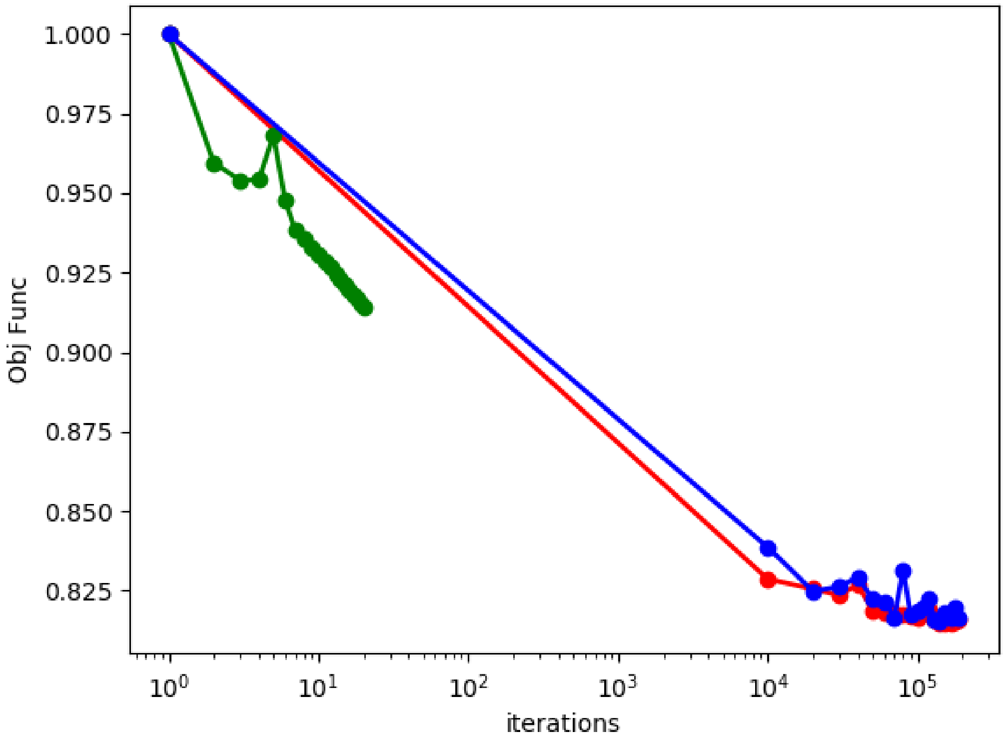
$$\therefore X_{t+1} = \frac{1}{q-1}\left\|\frac{\|X_t\|_p^{2-p}}{p-1} * diag(|X_t|^{p-1}) * sign(X_t) - \eta_t g_t\right\|_q^{2-q}$$
$$* diag\left(\left|\frac{\|X_t\|_p^{2-p}}{p-1} * diag(|X_t|^{p-1}) * sign(X_t) - \eta_t g_t\right|^{q-1}\right)$$
$$* sign\left(\frac{\|X_t\|_p^{2-p}}{p-1} * diag(|X_t|^{p-1}) * sign(X_t) - \eta_t g_t\right)$$

2 Programming Exercise

Plot 1:



Plot 2:



OUTPUT:

Solution found by stochastic subgradient descent

[[-3.82766340e-01], [-3.81199210e-02], [-1.33802728e-02], [-5.23892968e-01], [3.87125284e-04], [7.92598865e-01], [-1.63058725e-02], [9.88906872e-03], [9.18107436e-02], [1.54241552e-01], [3.11034142e-02], [2.14837698e-02], [-4.36862199e-02], [1.65145408e-01], [-2.44947948e-02], [2.29216069e-02], [-2.31771226e-02], [2.58326202e-01], [1.33468110e-02], [3.95091361e-03], [-3.98102673e-02], [-2.66643596e-01], [9.17257424e-01], [1.17738321e+00], [6.17161348e-01], [-1.51896124e+00], [2.42179921e+00], [-3.84489995e+00]]

Objective function 0.8148441242106866

Solution found by stochastic adagrad

[[-2.93993292e-01], [-8.47177834e-02], [-1.56053677e-03], [-4.88253260e-01],
[2.84568649e-02], [7.90104525e-01], [6.62801863e-03], [-2.14067903e-03],
[9.02412127e-02], [1.59807984e-01], [-3.86026941e-02], [-2.24010421e-02],
[-3.81838505e-02], [2.20424331e-01], [4.63755831e-02], [-2.89309034e-02],
[-5.72662586e-02], [2.29611015e-01], [-3.16711860e-02], [3.35261481e-03],
[-3.68488598e-02], [-2.38980461e-01], [8.62666343e-01], [1.18104634e+00],
[6.32419470e-01], [-1.52388920e+00], [2.91419386e+00], [-4.42720816e+00]]

Objective function 0.8176521573618533

Solution found by subgradient descent

[[-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