CS 331: Stochastic Gradient Descent Methods Assignment 1

Lukang Sun, ID: 182056

September 12, 2021

p1. Computing the Hesssian of the function, we know it is 2-smoothness, so step size $\gamma \le \frac{1}{L} = 0.5$. (After the first iteration, y-coordinate becomes 0, so y will not involve in the gradient descent process, only x-coordinate involves, however function is 0.02-smoothness with respect to the variable x, so after the first iteration, we can adjust the step size to 50.) Figure 1.shows the trajectory of gradient descent when $\gamma = 0.5$.

p2. Since $f(t) = e^t$ is convex function, so $D_f(x, y) \ge 0$, for any $x, y \in \mathbb{R}$, select $x = t, y = 0$, then $D_f(t, 0) = e^t - e^0 - \langle e^0, t - 0\rangle = e^t - 1 - t \ge 0$.

p3. $prox_R(x) = \arg\min_{a^T u=0} ||u-x||^2 = x - \frac{a^T x}{||a||^2} a$. Let $u = u_0 a + u_1 a^{\perp}, x = x_0 a + x_1 a^{\perp}$, where $u_1 a^{\perp} = u - \frac{a^T u}{||a||^2} a, x_1 a^{\perp} = x - \frac{a^T x}{||a||^2} a$, so $||u - x||^2 = ||u_0 a - x_0 a||^2 + ||u_1 a^{\perp} - x_1 a^{\perp}||^2$, so $\arg\min_{a^T u=0} ||u - x||^2 = \arg\min_{\{u:u_0=0\}} ||u_0 a - x_0 a||^2 + ||u_1 a^{\perp} - x_1 a^{\perp}||^2 = x_1 u^{\perp} = x - \frac{a^T x}{||a||^2} a$.

p4. $D_f(y, x) \ge \frac{\mu}{2} ||x - y||^2$ according to the statement of the problem, both side add $D_f(x, y)$, then we have

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle = D_f(x, y) + D_f(y, x) \ge D_f(x, y) + \frac{\mu}{2} ||x - y||^2,$$

which is the inequality we want to prove.

p5. let $g(t) = \langle \nabla f(y + t(x - y)), x - y\rangle$, then according to the mean value theorem, we have $g(1) - g(0) = g'(\xi)$, for some $\xi \in [0, 1]$, that is

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle = \langle \nabla^2 f(y + \xi(x - y))(x - y), x - y\rangle,$$

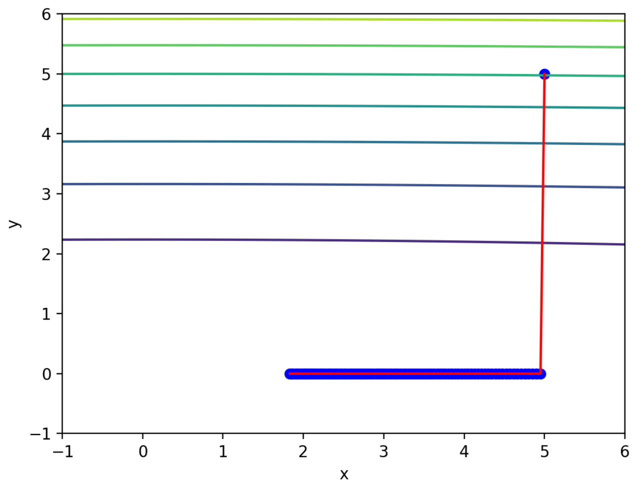

Figure 1: I set step size $\gamma = 0.5$, blue points are generated during the iteration, initial point is (5,5), the 100th iteration point is around(1.83,0), red line is the trajectory of the iteration.

since $\nabla^2 f$ is positive semi-definite, so R.H.S of the above formula is no less than 0, which means the symmetrized Bregman divergence is no less than 0, so f is convex.