Mathematics 2:
Homework 1

David Ocepek

March 31, 2022

1 Theoretical problems

1.1 Equivalence of 1. $\iff$ 4.

In order to prove 1. $\iff$ 4.$, we prove 1. \rightarrow$ 4. and 4 $\rightarrow$ 1. 1. is simply a special case of 4. when $k = 2$. We prove 4. $\rightarrow$ 1. by induction on the number of weighed points n. For $k = 1$ it is trivial to see the equation holds. Given $n = k - 1$ then

$$f(\sum_{i=1}^{k-1} \alpha_i x_i) \leq \sum_{i=1}^{k-1} \alpha_i f(x_i) \tag{1}$$

Adding x_k with weight α_k we need to devide all former α with $1 - \alpha_k$

$$f(\sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} x_i) \leq \sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} f(x_i) \tag{2}$$

Using the above equation to construct a weighed average

$$(1-\alpha_k) f(\sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} x_i) + \alpha_k f(x_k) \leq \sum_{i=1}^{k-1} \alpha_i f(x_i); \tag{3}$$

The following holds since we simply used inequality 2 to expand $f(\sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} x_i)$. Since $f(x)$ is continuous we know that $\sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} x_i \in D$; we can simply treat it as a point between x_1 and x_k. Therefore the following holds

$$\begin{aligned} f((1-\alpha_k) \sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} x_i + \alpha_k x_k) = f(\sum_{i=1}^{k-1} \alpha_i x_i) \leq \\ \leq (1-\alpha_k) f(\sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} x_i) + \alpha_k f(x_k) \end{aligned} \tag{4}$$

By transitivity of the operator $\leq$ it follows that $f(\sum_{i=1}^{k-1} \alpha_i x_i) \leq \sum_{i=1}^{k-1} \alpha_i f(x_i)$.

1.2 Bounds for f(x,y)

We calculate L by using the following inequality $||\nabla f|| \leq L$. The gradient equals

$$\nabla f = \begin{bmatrix} 2x + e^x - y \\ 2y - x \end{bmatrix} \tag{5}$$

Therefore the norm $||\nabla f||$ equals.

$$(2x + e^x - y)^2 + (2y - x)^2; \tag{6}$$

Here we removed the root since it does not change the ordinality of our norms. Since $f(x, y)$ is convex we know that the maximum will not be an interior point therefore we only need to check its borders. Formally, we'd construct the lagrangian L and give the KKT conditions, however we use some shortcuts to avoid most of the tedious calculations. For $(x, y) \in \{2, -2\} \times \{2, -2\}$ we get (2,-2) to be the greatest with $f(2, -2) \approx 215$ to be the largest, therefore the optimal L

$$\arg\max_{x,y} ||\nabla f|| = 14.67197 \leq L \tag{7}$$

We calculate β using the following inequality $||\nabla^2 f|| \leq \beta$. Where $||\nabla^2 f||$ equals

$$\nabla^2 f = \begin{bmatrix} 2 + e^x & -1 \\ -1 & 2 \end{bmatrix} \tag{8}$$

Finding the largest eigenvalue λ_{max} compute the $det(\nabla^2 f - \lambda I)$

$$\begin{vmatrix} 2 + e^x - \lambda & -1 \\ -1 & 2 - \lambda \end{vmatrix} \tag{9}$$

We get the following $(2 - \lambda)^2 + e^x(2 - \lambda) - 1 = \lambda^2 + (-e^x - 4)\lambda + (2e^x + 3)$. Solving the quadratic equation

$$\frac{(e^x + 4) + \sqrt{e^{2x} + 4}}{2} = 9.521999 \tag{10}$$

Here since all terms are positive we can clearly see that $x = 2$ maximizes λ.

In order for f to be strongly convex all eigenvalues of $\nabla^2 f$ need to be greater than α, therefore $\alpha = \arg\min_x min(\lambda)$. Calculating the derivative of $\frac{(e^x+4)-\sqrt{e^{2x}+4}}{2}$ we get

$$e^x - \frac{e^{2x}}{\sqrt{e^{2x} + 4}} \tag{11}$$

The following indicates there are no minimums in the interior of K. Hence we evaluate x at 2 and -2, the minimum being 1.065381 at -2. Since in the case 10 λ is at least four we conclude that the optimal value for α is 1.065381.

1.3 Projection formulas

For all free we conclude that if a point is in K then the projection should change nothing while for a projection of a point from outside K the point is projected to the nearest edge. Therefore are proofs are limited to points outside K

1. For $x^2 + y^2 \leq 1.5$

$$\begin{aligned} &\textit{minimize } (x_1 - x)^2 + (y_1 - y)^2 \\ &\textit{such that } x^2 + y^2 = 1.5, \end{aligned} \tag{12}$$

   we construct the lagrangian

$$\begin{aligned} &\mathcal{L}(x, y, \lambda) = (x_1 - x)^2 + (y_1 - y)^2 - \lambda(x^2 + y^2 - 1.5) \\ &\frac{\partial \mathcal{L}}{\partial x} = 2(x_1 - x) - 2\lambda x = 0 \\ &\frac{\partial \mathcal{L}}{\partial y} = 2(y_1 - y) - 2\lambda y = 0 \\ &\frac{\partial \mathcal{L}}{\partial \lambda} = -(x^2 + y^2 - 1.5) = 0 \end{aligned} \tag{13}$$

   Solving the following equation we get $y^{'} = \frac{y_1}{\sqrt{\frac{y_1^2 + x_1^2}{1.5}}}$ and $x^{'} = \frac{x_1}{\sqrt{\frac{y_1^2 + x_1^2}{1.5}}}$.
   Therefore our function is

$$f(x_1, y_1) = \begin{cases} x_1, y_1 & \text{if } x^2 + y^2 \leq 1.5 \\ x^{'}, y^{'} & \text{if } x^2 + y^2 > 1.5 \end{cases} \tag{14}$$

2. For $[-2, 2] \times [-2, 2]$

$$\begin{aligned} &\textit{minimize } (x_1 - x)^2 + (y_1 - y)^2 \\ &\textit{such that } x \in [-2, 2] \textit{ and } y \in [-2, 2], \end{aligned} \tag{15}$$

   We can formally write the KKT conditions and solve the equations, however we can avoid doing this by splitting our problem space into 9 spaces, split by lines going through points $(-2, -2)$, $(-2, 2)$, $(2, -2)$, $(2, 2)$. For each bounded by where at least y or x is in K we simply project the coordinate the is not to its nearest edge, in case neither are we project it to the nearest vertex.

$$f(x_1, y_1) = \begin{cases} x_1, y_1 & \text{if } x \in [-2, 2] & \wedge & y \in [-2, 2] \\ x_1, -2 & \text{if } x \in [-2, 2] & \wedge & y < -2 \\ x_1, 2 & \text{if } x \in [-2, 2] & \wedge & y > 2 \\ -2, y_1 & \text{if } x < -2 & \wedge & y \in [-2, 2] \\ 2, y_1 & \text{if } x > 2 & \wedge & y \in [-2, 2] \\ -2, -2 & \text{if } x < -2 & \wedge & y < -2 \\ -2, 2 & \text{if } x < -2 & \wedge & y > 2 \\ 2, -2 & \text{if } x > 2 & \wedge & y < -2 \\ 2, 2 & \text{if } x > 2 & \wedge & y > 2 \end{cases} \tag{16}$$

3. For triangle $\triangle ABC$, where A(-1,-1), B(1.5,-1), C(-1,1.5). Similairly we split the space into 7 subspaces, we project outliers to the nearest edge of $\triangle ABC$. We compute the following vector
$a = \vec{AB} = \begin{bmatrix} 2.5 \\ 0 \end{bmatrix}$, $b = \vec{AC} = \begin{bmatrix} 0 \\ -2.5 \end{bmatrix}$, $c = \vec{BC} = \begin{bmatrix} -2.5 \\ 2.5 \end{bmatrix}$.
We compute the necesary vectors for vector projection

$$f(x_1, y_1) = \begin{cases} C & \text{if } -1 < x < C + \alpha \begin{bmatrix} 1 & 1 \end{bmatrix}^T & \wedge & y > 1.5 \\ -1, 1 & \text{if } x < -1 & \wedge & y \in [-1, 1.5] \\ A & \text{if } x < -1 & \wedge & y < -1 \\ x, -1 & \text{if } x \in [-1, 1.5] & \wedge & y < -1 \\ x, -1 & \text{if } x \in [-1, 1.5] & \wedge & y < -1 \\ B & \text{if } x > 1.5 & \wedge & y \in (-\infty, \alpha B + \begin{bmatrix} 1 \\ 1 \end{bmatrix}] \\ B + BX & \text{if } x \in [-1, 1.5] & \wedge & y \in [B + \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}], C + \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}]] \end{cases} \tag{17}$$

The last inequality equals $BX = \frac{\vec{BC}x}{||\vec{BC}||||\vec{x}||}\vec{BC}$.

1.4 Gradient convergence

We change function f into $\frac{1}{2}x^T Hx$, where $H = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$.

1. We minimize $x_2^T Hx_2$ by γ and get $(I - 2\gamma)^T H(I - 2\gamma)$; we can ommit the x_1 from the equation since those just give us a regularizing constant. Evaluating the equation we get

$$2(1 - 2\gamma^2 + 4(1 - 4\gamma)^2 \tag{18}$$

   To minimize we set its derivative with respect to γ to zero

$$\begin{aligned} 4(1 - 2\gamma)(-2) + 8(1 - 4\gamma)(-4) = 0 \\ \gamma = \frac{5}{18} \end{aligned} \tag{19}$$

   Inserting gamma into gradient descent, we get that the lowest function value equals $\begin{bmatrix} 8/18 \\ -2/18 \end{bmatrix}$.

2. To find the minimum x we can reach in one step we minimize the norm $||x_1 - \gamma Hx_1||$

$$\begin{aligned} (x_1 - \gamma Hx_1)^T((x_1 - \gamma Hx_1)) = \\ x_1^T(I - 2\gamma H + \gamma^2 H^2)x_1 \end{aligned} \tag{20}$$

   In order to find the minimum of $||x_1 - \gamma Hx_1||$ with respect to γ we derivate

$$x_1^T(-2H + 2\gamma H^2)x_1 = 0 \tag{21}$$

Solving the following equation for γ we get $\gamma = 3/10$ and $x_2 = \begin{bmatrix} 2/5 \\ -1/5 \end{bmatrix}$.

2 Practical problems

2.1 PGD

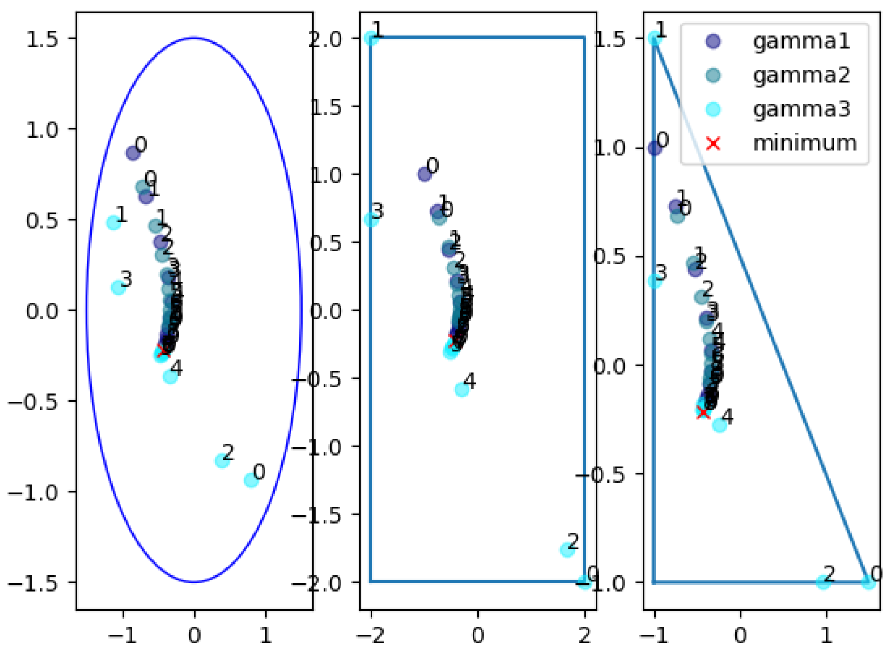

Figure 1: **Convergence using different gamma's.** The optimal Lipschitz convergence oscilates around the minimum, while beta and alpha GD slowly converges towards minimum.

2.2 GD

Using gradient descent with $\gamma = 6e - 4$ we reached -3.8627815600480115 in 9999 steps.