Optimization for Data Science: Special Assignment 1

Gabriel Hayat

April 28, 2019

Assignment 1. Given i $\in$ [n] and $\mathbf{x} \in \mathbb{R}^d$, let

$$g_i(\mathbf{x}) = \begin{cases} \mathbf{0} & \text{if } \mathbf{x} \in B_i \\ \frac{\mathbf{x}-\mathbf{c_i}}{||\mathbf{x}-\mathbf{c_i}||} & \text{otherwise} \end{cases}$$

Show that $g_i(\mathbf{x})$ is a subgradient of $h_i(\mathbf{x})$, *i.e*, $g_i(\mathbf{x}) \in \partial h_i(\mathbf{x})$. Is h_i differentiable?

Solution:

According to definition 4.1, $g_i(\mathbf{x}) \in \partial h_i(\mathbf{x})$ if

$$h_i(\mathbf{y}) \geq h_i(\mathbf{x}) + g_i(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) \quad \forall \mathbf{y} \in \mathbf{dom}(h_i) \tag{1}$$

We consider two cases:

Case 1 : $\mathbf{x} \in B_i$
Since both $h_i(\mathbf{x}) = \mathbf{0}$ and $g_i(\mathbf{x}) = \mathbf{0}$, the right hand side of inequality (1) is 0.
As

$$h_i(\mathbf{y}) \geq 0 \quad \forall \mathbf{y} \in \mathbb{R}^d \tag{2}$$

inequality (1) holds.

Case 2 : $\mathbf{x} \notin B_i$
As $h_i(\mathbf{x}) = ||\mathbf{x}-\mathbf{c_i}|| - r_i$ and $g_i(\mathbf{x}) = \frac{\mathbf{x}-\mathbf{c_i}}{||\mathbf{x}-\mathbf{c_i}||}$, we can write inequality (1) as:

$$h_i(\mathbf{y}) \overset{?}{\geq} ||\mathbf{x}-\mathbf{c_i}|| - r_i + \left[\frac{\mathbf{x}-\mathbf{c_i}}{||\mathbf{x}-\mathbf{c_i}||}\right]^T(\mathbf{y}-\mathbf{x}) \tag{3}$$

Consider the right handside of inequality (3):

$$||\mathbf{x}-\mathbf{c_i}|| - r_i + \left[\frac{\mathbf{x}-\mathbf{c_i}}{||\mathbf{x}-\mathbf{c_i}||}\right]^T(\mathbf{y}-\mathbf{x}) = \frac{1}{||\mathbf{x}-\mathbf{c_i}||}\left[||\mathbf{x}-\mathbf{c_i}||^2 - r_i||\mathbf{x}-\mathbf{c_i}|| + (\mathbf{x}-\mathbf{c_i})^T(\mathbf{y}-\mathbf{x})\right]$$

$$= \frac{1}{||\mathbf{x}-\mathbf{c_i}||}\left[(\mathbf{x}-\mathbf{c_i})^T(\mathbf{y}-\mathbf{c_i}) - r_i||\mathbf{x}-\mathbf{c_i}||\right] \tag{4}$$

$$\leq \frac{1}{||\mathbf{x}-\mathbf{c_i}||}\left[]||\mathbf{x}-\mathbf{c_i}||\ ||\mathbf{y}-\mathbf{c_i}|| - r_i||\mathbf{x}-\mathbf{c_i}||\right] \tag{5}$$

$$= ||\mathbf{y}-\mathbf{c_i}|| - r_i \tag{6}$$

where in (4), we used the result $||\mathbf{x}||^2 = \mathbf{x}^T\mathbf{x}$ and in (5), we used Cauchy-Schwarz inequality. Note that $||\mathbf{x}-\mathbf{c_i}|| \neq 0$ as $\mathbf{x} \notin B_i$.

We can now examine the inequality:

$$h_i(\mathbf{y}) \overset{?}{\geq} ||\mathbf{y}-\mathbf{c_i}|| - r_i \tag{7}$$

If $\mathbf{y} \in B_i$, then by definition the right hand side of inequality (7) will be negative. By inequality (2), (7) holds.
If $\mathbf{y} \notin B_i$, then by definition of h_i:

$$h_i(\mathbf{y}) = ||\mathbf{y}-\mathbf{c_i}|| - r_i$$

and inequality (7) holds. We conclude that equation (1) holds $\forall \mathbf{y} \in \mathbb{R}^d$, and $g_i(\mathbf{x})$ is a subgradient of $h_i(\mathbf{x})$.

Assume h_i differentiable. From Definition 1.7, we know there exists a unique $(1 \times d)$-matrix A and an error function $r : \mathbb{R}^d \to \mathbb{R}$ definied around $\mathbf{0} \in \mathbb{R}^d$ such that forall $\mathbf{y}$ in some neighborhood of $\mathbf{x}$,

$$h_i(\mathbf{y}) = h_i(\mathbf{x}) + A(\mathbf{y}-\mathbf{x}) + r(\mathbf{y}-\mathbf{x}) \tag{8}$$

where $\lim_{v\to 0} \frac{|r(\mathbf{v})|}{||\mathbf{v}||} = \mathbf{0}$

Consider the points $\mathbf{x} = \mathbf{c_i} + r_i\mathbf{e}_1$ and $\mathbf{y} = \mathbf{c_i} + (r_i+\beta)\mathbf{e}_1$ for some small $\beta \in \mathbb{R}$, where $\mathbf{e}_1$ is the first vector of the natural basis. We then consider two cases:
Case 1: $\beta < 0 \Rightarrow \mathbf{x}, \mathbf{y} \in B_i$

$$0 = 0 + \beta A(\mathbf{e}_1) + r(\beta\mathbf{e}_1)$$

$$|A(\beta\mathbf{e}_1)| = \frac{|r(\beta\mathbf{e}_1)|}{|\beta|} \xrightarrow{\beta\to 0} 0 \tag{9}$$

where we used the fact that $||\beta\mathbf{e}_1|| = |\beta|$
Case 2: $\beta > 0 \Rightarrow \mathbf{x} \in B_i, \mathbf{y} \notin B_i$

$$||(r_i+\beta)\mathbf{e}_1|| - r_i = 0 + \beta A(\mathbf{e}_1) + r(\beta\mathbf{e}_1)$$

$$\beta = \beta A(\mathbf{e}_1) + r(\beta\mathbf{e}_1) \tag{10}$$

$$\beta(A(\mathbf{e}_1) - 1) = -r(\beta\mathbf{e}_1) \tag{11}$$

$$|A(\mathbf{e}_1)| = \left|1 - \frac{r(\beta \mathbf{e}_1)}{\beta}\right| \geq 1 - \frac{|r(\beta \mathbf{e}_1)|}{\beta} \xrightarrow{\beta \to 0} 1 \tag{12}$$

where we used the reverse triangle inequality and the fact that $||\beta \mathbf{e}_1|| = \beta$
We see that the two cases contradicts. We conclude that the original assumption is wrong and h_i is not differentiable. □

Assignment 2. For $\mathbf{x} \in \mathbb{R}^d$, show that there is $j \in [n]$ such that $g_j(\mathbf{x})$ is a subgrandient of $h(\mathbf{x})$, *i.e*, $g_j(\mathbf{x}) \in \partial h(\mathbf{x})$. Describe how to compute such subgradient and provide the running time of this procedure.

Solution:

Take $j = \operatorname*{argmax}_{i \in [n]} h_i(x)$. Then, $h(\mathbf{x}) = h_j(\mathbf{x})$.
We showed in the previous question that for any $i \in [n]$, $g_i(\mathbf{x}) \in \partial h_i(\mathbf{x})$. In particular, take $i = j$ and $g_j(\mathbf{x}) \in \partial h_j(\mathbf{x}) = \partial h(\mathbf{x})$.

To compute such gradients:

1. For each $i \in [n]$, compute $h_i(\mathbf{x})$:
   (a) Compute $||\mathbf{x} - \mathbf{c_i}|| = D_i$
   (b) If $D_i \leq r_i$, then $h_i(\mathbf{x}) = 0$ else $h_i(\mathbf{x}) = D_i - r_i$
2. Compute $j = \operatorname*{argmax}_{i \in [n]} h_i(x)$
3. If $h_j(\mathbf{x}) = \mathbf{0}$, then $g_j(\mathbf{x}) = \mathbf{0}$ else $g_j(\mathbf{x}) = \frac{\mathbf{x} - \mathbf{c_j}}{||\mathbf{x} - \mathbf{c_j}||}$

Step 1 is done in $\mathcal{O}(nd)$, step 2 in $\mathcal{O}(n)$ and step 3 in $\mathcal{O}(1)$ (as all calculations have been previously done and can be reused). As d is assumed to be constant, the gradient is computed in $\mathcal{O}(n)$. □

Assignment 3. Show that each h_i is convex and 1-Lipschitz. Then use this to show that h is convex and 1-Lipschitz.

Solution:

By Lemma 4.3, h_i is convex if $\partial h_i(\mathbf{x}) \neq \emptyset \quad \forall \mathbf{x} \in \mathbf{dom}(h_i)$ and $\mathbf{dom}(h_i)$ is convex. In Assignment 1, we have shown that:

$$h_i(\mathbf{y}) \geq h_i(\mathbf{x}) + g_i(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}(h_i) \tag{13}$$

Since $\mathbf{dom}(h_i)$ is convex, we conclude that h_i is also convex.
To show that h_i is 1-Lipschitz, we must show:

$$|h_i(\mathbf{x}) - h_i(\mathbf{y})| \leq ||\mathbf{x} - \mathbf{y}|| \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}(h_i) \tag{14}$$

Case 1: $\mathbf{x}, \mathbf{y} \in B_i$
The left hand-side of inequality (11) is 0 and thus holds by the definition of a norm.

Case 2: $\mathbf{x} \in B_i, \mathbf{y} \notin B_i \quad (resp.\ \mathbf{x} \notin B_i,\ \mathbf{y} \in B_i)$

$$|h_i(\mathbf{y}) - h_i(\mathbf{x})| = ||\mathbf{y} - \mathbf{c_i}|| - r_i \tag{15}$$

$$\leq ||\mathbf{y} - \mathbf{c_i}|| - ||\mathbf{x} - \mathbf{c_i}|| \tag{16}$$

$$\leq ||(\mathbf{y} - \mathbf{c_i}) - (\mathbf{x} - \mathbf{c_i})|| = ||\mathbf{y} - \mathbf{x}|| \tag{17}$$

where in (13), we used the fact that $\mathbf{x} \in B_i$ and in (14), we used the reverse triangle inequality.

Case 3: $\mathbf{x} \notin B_i, \mathbf{y} \notin B_i$

$$|h_i(\mathbf{y}) - h_i(\mathbf{x})| = ||\mathbf{y} - \mathbf{c_i}|| - r_i - ||\mathbf{x} - \mathbf{c_i}|| + r_i \tag{18}$$

$$= ||\mathbf{y} - \mathbf{c_i}|| - ||\mathbf{x} - \mathbf{c_i}|| \tag{19}$$

$$\leq ||\mathbf{y} - \mathbf{x}|| \tag{20}$$

where in (17), the reverse triangle inequality was used again.
We conclude that h_i is 1-Lipschitz.

For $j = \underset{i \in [n]}{\operatorname{argmax}}\, h_i(x)$, we have $h(\mathbf{x}) = h_j(\mathbf{x})$. In particular, h is also convex and 1-Lipschitz.

□

Assignment 4. Assume that you are given $\mathbf{x_0} \in \mathbb{R}^d$ such that $||\mathbf{x_0} - \mathbf{x}^*|| \leq R$ for some constant R $\in \mathbb{R}$, where $\mathbf{x}^*$ is a global minimum of h. Show that there exists $\gamma \in \mathbb{R}$ such that after T iterations of subgradient descent on $h(\mathbf{x})$ we get:

$$\min_{t \in 0, \ldots, T-1} (h(\mathbf{x}_t) - h(\mathbf{x}^*)) \leq \frac{R}{\sqrt{T}} \tag{21}$$

Solution:

We use the Vanilla analysis (Eq. 2.6 from notes):

$$\sum_{t=0}^{T-1} h(\mathbf{x}_t) - h(\mathbf{x}^*) \leq \frac{\gamma}{2} \sum_{t=0}^{T-1} ||\mathbf{g}_t|| + \frac{1}{2\gamma} ||\mathbf{x_0} - \mathbf{x}^*||^2 \tag{22}$$

From assignment 4, we know that h is 1-Lipschitz. According to Lemma 4.4, this is equivalent to $||\mathbf{g}|| \leq 1$. In addition, $||\mathbf{x_0} - \mathbf{x}^*|| \leq R$ allows us to write:

$$\sum_{t=0}^{T-1} h(\mathbf{x}_t) - h(\mathbf{x}^*) \leq \frac{\gamma}{2} T + \frac{1}{2\gamma} R^2 \tag{23}$$

We want to minimize the bound, *i.e* minimize the function $g(\gamma) = \frac{\gamma}{2}T + \frac{1}{2\gamma}R^2$. By setting its gradient to 0:

$$g^{'}(\gamma) = \frac{T}{2} - \frac{R^2}{2\gamma^2} = 0$$

$$\gamma = \frac{R}{\sqrt{T}}$$

Inequality (20) now becomes:

$$\sum_{t=0}^{T-1} h(\mathbf{x}_t) - h(\mathbf{x}^*) \leq \frac{R\sqrt{T}}{2} + \frac{R\sqrt{T}}{2} = R\sqrt{T} \tag{24}$$

$$\frac{1}{T}\sum_{t=0}^{T-1} h(\mathbf{x}_t) - h(\mathbf{x}^*) \leq \frac{R}{\sqrt{T}} \tag{25}$$

where inequality (22) is obtained by dividing by T.
We have obtained the result that the average error is bounded by the factor $\frac{R}{\sqrt{T}}$. This is only possible when at least one error in the sequence $t \in \{0, ..., T-1\}$ is bounded by $\frac{R}{\sqrt{T}}$. In particular, the minimum error in the sequence will be bounded by this factor. Hence the result. $\square$

Assignment 5. Given $\epsilon > 0$, show that $\mathcal{O}(n/\epsilon^2)$ time suffices to solve the $\epsilon -$ *intersection* problem for the set of B_i's.

Solution:

Using the result from assignment 4, we know that the minimum error after running subgradient descent for T steps can be bounded.
Choose $\epsilon > 0$ such that $\frac{R}{\sqrt{T}} < \epsilon$ and let $\min_{t \in 0,...,T-1}(h(\mathbf{x}_t) - h(\mathbf{x}^*)) = h(\mathbf{x}_m) - h(\mathbf{x}^*)$. Then,

$$h(\mathbf{x}_m) - h(\mathbf{x}^*) < \epsilon$$

$$h(\mathbf{x}^*) > h(\mathbf{x}_m) - \epsilon$$

We now have two cases to consider:
Case 1: $h(\mathbf{x}_m) - \epsilon \geq 0$
Since $\mathbf{x}^*$ is the global minimum of h, we conclude that $h(\mathbf{x}) > 0$ for every $\mathbf{x} \in \mathbb{R}^d$.
Case 2: $h(\mathbf{x}_m) - \epsilon < 0$
Then choose $\bar{\mathbf{x}} = \mathbf{x}_m$ such that $h(\bar{\mathbf{x}}) < \epsilon$

Therefore, running subgradient descent for $T > \frac{R^2}{\epsilon^2}$ steps suffices to solve the $\epsilon -$ *intersection* problem.
However, each step of the algorithm requires taking a maximum over n elements to compute $h(\mathbf{x}_t)$, causing an additional factor of n. Thus, the $\epsilon -$ *intersection* problem can be solved in $\mathcal{O}(n/\epsilon^2)$ time.

$\square$

Assignment 6. Using elementary calculus (e.g. partial derivatives) show that:

$$\nabla f_i(\mathbf{x}) = \begin{cases} \mathbf{0} & \text{if } \mathbf{x} \in B_i \\ (1 - \frac{r_i}{||\mathbf{x}-\mathbf{c_i}||})(\mathbf{x} - \mathbf{c_i}) & \text{otherwise} \end{cases}$$

Show how to compute the gradient $\nabla f(\mathbf{x})$. How much time is needed to compute this gradient?

Solution:

We use the definition:

$$\frac{\partial f}{\partial x_i} = \lim_{h\to 0} \frac{f(x_1, ..., x_i + h, ..., x_d) - f(x_1, ..., x_n)}{h} \tag{26}$$

Case 1: $\mathbf{x} \in B_i$

$$\frac{\partial f}{\partial x_i} = \lim_{h\to 0} \frac{0-0}{h} = 0$$

Thus, we conclude that $\nabla f(\mathbf{x}) = \mathbf{0} \quad if \ \mathbf{x} \in B_i$.

Case 2: $\mathbf{x} \notin B_i$
Consider the function:

$$f_i(\mathbf{x}) = \frac{1}{2} d(\mathbf{x}, B_i)^2 = \frac{1}{2}||\mathbf{x} - \mathbf{c_i}||^2 - r_i||\mathbf{x} - \mathbf{c_i}|| + \frac{1}{2}r_i$$

Consider $p_1(\mathbf{x}) = \frac{1}{2}||\mathbf{x} - \mathbf{c_i}||^2$, $p_2(\mathbf{x}) = r_i||\mathbf{x} - \mathbf{c_i}||$, $p_3(\mathbf{x}) = \frac{1}{2}r_i$ such that $f_i(\mathbf{x}) = p_1(\mathbf{x}) - p_2(\mathbf{x}) + p_3(\mathbf{x})$

We then apply definition (23) to each part p_i:

$$\frac{\partial p_1}{\partial x_j} = \frac{1}{2} \lim_{h\to 0} \frac{(x_1 - c_{i1})^2 + .. + (x_j + h - c_{ij})^2 + .. + (x_d - c_{id})^2 - \Big((x_1 - c_{i1})^2 + .. + (x_d - c_{id})^2\Big)}{h}$$

$$= \frac{1}{2} \lim_{h\to 0} \frac{(x_j + h - c_{ij})^2 - (x_j - c_{ij})^2}{h}$$

$$= \frac{1}{2} \lim_{h\to 0} \frac{x_j^2 + 2hx_j + h^2 - 2(x_j + h)c_{ij} + c_{ij}^2 - (x_j^2 - 2x_jc_{ij} + c_{ij}^2)}{h}$$

$$= \frac{1}{2} \lim_{h\to 0} \frac{2hx_j + h^2 - 2hc_{ij}}{h}$$
$$= x_j - c_{ij}$$

Thus $\nabla p_1(\mathbf{x}) = \mathbf{x} - \mathbf{c_i}$

$$\frac{\partial p_2}{\partial x_j} = \lim_{h \to 0} r_i \frac{\sqrt{(x_1 - c_{i1})^2 + .. + (x_j + h - c_{ij})^2 + .. + (x_d - c_{id})^2} - \sqrt{(x_1 - c_{i1})^2 + .. + (x_d - c_{id})^2}}{h}$$

$$= \lim_{h \to 0} r_i \frac{(x_1 - c_{i1})^2 + .. + (x_j + h - c_{ij})^2 + .. + (x_d - c_{id})^2 - (x_1 - c_{i1})^2 + .. + (x_d - c_{id})^2}{h\sqrt{(x_1 - c_{i1})^2 + .. + (x_j + h - c_{ij})^2 + .. + (x_d - c_{id})^2} + \sqrt{(x_1 - c_{i1})^2 + .. + (x_d - c_{id})^2}}$$

$$= \lim_{h \to 0} r_i \frac{(x_j + h - c_{ij})^2 - (x_j - c_{ij})^2}{h\sqrt{(x_1 - c_{i1})^2 + .. + (x_j + h - c_{ij})^2 + .. + (x_d - c_{id})^2} + \sqrt{(x_1 - c_{i1})^2 + .. + (x_d - c_{id})^2}}$$

$$= \lim_{h \to 0} 2r_i \frac{x_j - c_{ij} + h}{\sqrt{(x_1 - c_{i1})^2 + .. + (x_j + h - c_{ij})^2 + .. + (x_d - c_{id})^2} + \sqrt{(x_1 - c_{i1})^2 + .. + (x_d - c_{id})^2}}$$

$$= 2r_i \frac{x_j - c_{ij}}{2||\mathbf{x} - \mathbf{c_i}||} \qquad = r_i \frac{x_j - c_{ij}}{||\mathbf{x} - \mathbf{c_i}||}$$

Thus $\nabla p_2(\mathbf{x}) = r_i \frac{\mathbf{x} - \mathbf{c_i}}{||\mathbf{x} - \mathbf{c_i}||}$
Since $p_3(\mathbf{x})$ is constant with respect to $\mathbf{x}$, $\nabla p_3(\mathbf{x}) = \mathbf{0}$
By linearity of operator ∇:

$$\nabla f_i(\mathbf{x}) = \nabla p_1(\mathbf{x}) - \nabla p_2(\mathbf{x}) + \nabla p_3(\mathbf{x}) = \mathbf{x} - \mathbf{c_i} - r_i \frac{\mathbf{x} - \mathbf{c_i}}{||\mathbf{x} - \mathbf{c_i}||} + \mathbf{0}$$

$$= \big(1 - \frac{r_i}{||\mathbf{x} - \mathbf{c_i}||}\big)\big(\mathbf{x} - \mathbf{c_i}\big)$$

Computing $\nabla f_i(\mathbf{x})$ requires $\mathcal{O}(d)$. Again, we use the linearity property of operator ∇ to write:

$$\nabla f(\mathbf{x}) = \sum_{t=1}^{n} \nabla f_i(\mathbf{x})$$

Since computing $\nabla f(\mathbf{x})$ requires summing over n elements, the total running time is $\mathcal{O}(nd)$ $= \mathcal{O}(n)$ (d is assumed constant). $\square$

Assignment 7. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ such that $||\mathbf{u}|| = ||\mathbf{v}||$ and for real numbers $\alpha, \beta \geq 1$

$$||\mathbf{u} - \mathbf{v}|| \leq ||\alpha\mathbf{u} - \beta\mathbf{v}||$$

Solution:

Assume $||\mathbf{u}|| = ||\mathbf{v}|| = a$. Notice that showing this inequality is equivalent to showing:

$$||\mathbf{u} - \mathbf{v}||^2 \leq ||\alpha\mathbf{u} - \beta\mathbf{v}||^2$$

Consider:
$||\mathbf{u} - \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2\mathbf{u}^T\mathbf{v} = 2a^2 - 2\mathbf{u}^T\mathbf{v}$

$||\alpha\mathbf{u} - \beta\mathbf{v}||^2 = \alpha^2||\mathbf{u}||^2 + \beta^2||\mathbf{v}||^2 - 2\alpha\beta\mathbf{u}^T\mathbf{v} = a^2(\alpha^2+\beta^2) - 2\alpha\beta\mathbf{u}^T\mathbf{v}$
We thus need to show:

$$2a^2 - 2\mathbf{u}^T\mathbf{v} \overset{?}{\leq} a^2(\alpha^2+\beta^2) - 2\alpha\beta\mathbf{u}^T\mathbf{v} \tag{27}$$

$$0 \overset{?}{\leq} a^2(\alpha^2+\beta^2-2) + 2(1-\alpha\beta)\mathbf{u}^T\mathbf{v} \tag{28}$$

We use Cauchy-Schwarz inequality $\mathbf{u}^T\mathbf{v} \geq -||\mathbf{u}||||\mathbf{v}|| = -a^2$ to show instead:

$$0 \overset{?}{\leq} a^2(\alpha^2+\beta^2-2) - 2a^2(1-\alpha\beta) \tag{29}$$

$$0 \overset{?}{\leq} a^2(\alpha^2+\beta^2+2\alpha\beta-4) \tag{30}$$

$$0 \overset{?}{\leq} (\alpha+\beta)^2 - 4 \tag{31}$$

$$(\alpha+\beta)^2 \overset{?}{\geq} 4 \tag{32}$$

As $\alpha, \beta \geq 1$ by hypothesis, inequality 29 holds. Note that step (28) used the fact that $a^2 \geq 0$
$\square$

Assignment 8. Using Assignment 7, show that for each f_i, it holds that

$$||\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{y})|| \leq 2||\mathbf{x} - \mathbf{y}|| \tag{33}$$

That is, f_i is smooth with parameter 2 (Lemma 2.4). Using this fact and Lemma 2.5 we also get that f is smooth with parameter 2n. Is f always strongly convex?

Solution:

Case 1: $\mathbf{x}, \mathbf{y} \in B_i$
We have that $\nabla f_i(\mathbf{x}) = \nabla f_i(\mathbf{y}) = \mathbf{0}$ and thus by definition of a norm, inequality (30) holds.

Case 2: $\mathbf{x} \in B_i, \mathbf{y} \notin B_i$ (*resp.* $\mathbf{x} \notin B_i$, $\mathbf{y} \in B_i$)

$$||\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{y})|| = ||\mathbf{y} - \mathbf{c_i}|| - r_i \tag{34}$$

$$\leq ||\mathbf{y} - \mathbf{c_i}|| - ||\mathbf{x} - \mathbf{c_i}|| \tag{35}$$

$$\leq ||\mathbf{y} - \mathbf{x}|| \leq 2||\mathbf{y} - \mathbf{x}|| \tag{36}$$

where inequality (32) used the fact that $\mathbf{x} \in B_i$ and inequality (33) used the reverse triangle inequality.
Case 3: $\mathbf{x} \notin B_i, \mathbf{y} \notin B_i$

$$||\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{y})|| = ||\mathbf{x} - \mathbf{c_i} - r_i\frac{\mathbf{x} - \mathbf{c_i}}{||\mathbf{x} - \mathbf{c_i}||} - \left(\mathbf{y} - \mathbf{c_i} - r_i\frac{\mathbf{y} - \mathbf{c_i}}{||\mathbf{y} - \mathbf{c_i}||}\right)|| \tag{37}$$

$$= ||\mathbf{x} - \mathbf{y} + r_i\left(\frac{\mathbf{y} - \mathbf{c_i}}{||\mathbf{y} - \mathbf{c_i}||} - \frac{\mathbf{x} - \mathbf{c_i}}{||\mathbf{x} - \mathbf{c_i}||}\right)|| \tag{38}$$

$$\leq ||\mathbf{x}-\mathbf{y}|| + r_i \left|\left| \frac{\mathbf{y}-\mathbf{c_i}}{||\mathbf{y}-\mathbf{c_i}||} - \frac{\mathbf{x}-\mathbf{c_i}}{||\mathbf{x}-\mathbf{c_i}||} \right|\right| \tag{39}$$

We now apply assignment 7 with $\mathbf{u} = \frac{\mathbf{y}-\mathbf{c_i}}{||\mathbf{y}-\mathbf{c_i}||}$, $\mathbf{v} = \frac{\mathbf{x}-\mathbf{c_i}}{||\mathbf{x}-\mathbf{c_i}||}$, $||\mathbf{u}|| = ||\mathbf{v}|| = 1$, $\alpha = \frac{||\mathbf{y}-\mathbf{c_i}||}{r_i}$, $\beta = \frac{||\mathbf{x}-\mathbf{c_i}||}{r_i}$, where $\alpha, \beta \geq 1$ as $\mathbf{x} \notin B_i, \mathbf{y} \notin B_i$:

$$\leq ||\mathbf{x}-\mathbf{y}|| + r_i \left|\left| \frac{||\mathbf{y}-\mathbf{c_i}||}{r_i} \frac{\mathbf{y}-\mathbf{c_i}}{||\mathbf{y}-\mathbf{c_i}||} - \frac{||\mathbf{x}-\mathbf{c_i}||}{r_i} \frac{\mathbf{x}-\mathbf{c_i}}{||\mathbf{x}-\mathbf{c_i}||} \right|\right| \tag{40}$$

$$= ||\mathbf{x}-\mathbf{y}|| + ||\mathbf{y}-\mathbf{c_i} - \big(\mathbf{x}-\mathbf{c_i}\big)|| \tag{41}$$

$$= ||\mathbf{x}-\mathbf{y}|| + ||\mathbf{y}-\mathbf{x}|| = 2||\mathbf{x}-\mathbf{y}|| \tag{42}$$

where the triangle inequality was used in equation (36).
In order for f to be strongly convex, we need to find $\mu \in \mathbb{R}_+$, $\mu > 0$ such that

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{\mu}{2}||\mathbf{x}-\mathbf{y}||^2 \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}(f) \tag{43}$$

Consider the case where $\exists\ \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that $f(\mathbf{x}) = f(\mathbf{y}) = 0$ ($f_i(\mathbf{x}) = f_i(\mathbf{y}) = 0 \quad \forall i \in [n]$). Then:

$$0 \overset{?}{\geq} \frac{\mu}{2}||\mathbf{x}-\mathbf{y}||^2 \tag{44}$$

which is false $\forall \mu \in \mathbb{R}_+, \mu > 0$.
Thus f is not always strongly convex. □

Assignment 9. Given $\epsilon > 0$, show that $\mathcal{O}(n^{3/2}/\epsilon)$ time suffices to solve the $\epsilon-$*intersection* problem for the set of B_i's. Use accelerated gradient descent to show this bound.

Solution:

Using accelerated gradient descent, we have that:

$$f(\mathbf{y}_t) - f(\mathbf{x}^*) \leq \frac{2L||\mathbf{z_0}-\mathbf{x}^*||}{T(T+1)} = \frac{4nR^2}{T(T+1)} \leq \frac{2nR^2}{T^2} \tag{45}$$

Choose $\epsilon > 0$ such that $\frac{2nR^2}{T^2} < \frac{\epsilon^2}{2}$. Then,

$$f(\mathbf{y}_t) - f(\mathbf{x}^*) < \frac{\epsilon^2}{2}$$

$$f(\mathbf{x}^*) > f(\mathbf{y}_t) - \frac{\epsilon^2}{2}$$

We now have two cases to consider:
Case 1: $f(\mathbf{y}_t) - \frac{\epsilon^2}{2} \geq 0$
Let $\mathbf{z}^*$ be the global minimum of the function h. Since $\mathbf{x}^*$ is a global minimum of function f

$$f(\mathbf{x}^*) > 0 \Rightarrow f(\mathbf{z}^*) > 0 \tag{46}$$

$$\sum_{i=1}^{n} f_i(\mathbf{z}^*) > 0 \tag{47}$$

As $f_i(\mathbf{x}) = \frac{1}{2}h_i(\mathbf{x})^2$, we can write:

$$\frac{1}{2}\sum_{i=1}^{n} h_i(\mathbf{z}^*)^2 > 0 \tag{48}$$

Since $h(\mathbf{x})$ represents the maximum value over all $h_i(\mathbf{x})$, we conclude that:

$$h(\mathbf{z}^*)^2 > 0 \Rightarrow h(\mathbf{z}^*) > 0 \tag{49}$$

as $h(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^d$
Since $\mathbf{z}^*$ is the global minimum of h, we conclude that $h(\mathbf{x}) > 0$ for every $\mathbf{x} \in \mathbb{R}^d$.

Case 2: $f(\mathbf{y}_t) - \frac{\epsilon^2}{2} < 0$. Then,

$$\sum_{i=1}^{n} f_i(\mathbf{y}_t) - \frac{\epsilon^2}{2} < 0 \tag{50}$$

$$\frac{1}{2}\sum_{i=1}^{n} h_i(\mathbf{y}_t)^2 - \frac{\epsilon^2}{2} < 0 \tag{51}$$

$$\sum_{i=1}^{n} h_i(\mathbf{y}_t)^2 - \epsilon^2 < 0 \tag{52}$$

$$h(\mathbf{y}_t)^2 - \epsilon^2 < 0 \tag{53}$$

$$h(\mathbf{y}_t) < \epsilon \tag{54}$$

where the same arguments as in the first case are used.
Then choose $\bar{\mathbf{x}} = \mathbf{y}_t$ such that $h(\bar{\mathbf{x}}) < \epsilon$.

Therefore, running accelerated gradient descent for $T > \frac{2n^{1/2}R}{\epsilon}$ steps suffices to solve the $\epsilon - intersection$ problem.

As each step of the algorithm requires taking a sum over n elements, the complexity is increased by an additional factor of n. Thus, the $\epsilon - intersection$ problem can be solved in $\mathcal{O}(n^{3/2}/\epsilon)$.

□