4 Clustering via Alternation with Weiszfeld Step

In this section we tackle the clustering problem with distance-like function being the Euclidean norm in $\mathbb{R}^n$, namely

$$\min_{x^1,x^2,\dots,x^k\in\mathbb{R}^n}\left\{\sum_{i=1}^{m}\min_{1\le l\le k}\cancel{d(x^l,a^i)}\right\}, \tag{4.1}$$

$\|x^l - a^i\|$

~~where $d(u,v)=\|u-v\|$.~~

In the previous sections we showed that (4.1) has the following equivalent form

$$\min\left\{\Psi(z) := H(w,x) + G(w) \mid z := (w,x) \in \mathbb{R}^{km}\times\mathbb{R}^{nk}\right\},$$

where $H(w,x) = \sum_{i=1}^{m}\langle w^i, d^i(x)\rangle = \sum_{i=1}^{m}\sum_{l=1}^{k}\cancel{w_l^i d(x^l,a^i)} = \sum_{l=1}^{k}\sum_{i=1}^{m}\cancel{w_l^i d(x^l,a^i)} = \sum_{l=1}^{k}\sum_{i=1}^{m} w_l^i\|x^l-a^i\|$, and $G(w) = \sum_{i=1}^{m}\delta_\Delta(w^i)$.

delete

4.1 Weiszfeld Algorithm Scheme

Short overview of the problem and the algorithm ...

4.2 Clustering with Weiszfeld Step

We introduce some useful notations that will be useful for this section. For $1\le l\le k$ denote ~~$L_l(w,x) = \sum_{i=1}^{m}\frac{w_l^i}{\|x_l-a^i\|}$ and~~ $H_l(w,x) = \sum_{i=1}^{m} w_l^i\|x^l-a^i\|$.

delete

Weiszfeld single iteration is defined for each $1\le l\le k$ via

$$T_l(w,x) = \frac{\sum_{i=1}^{m}\frac{w_l^i a^i}{\|x^l-a^i\|}}{\sum_{i=1}^{m}\frac{w_l^i}{\|x^l-a^i\|}}. \tag{4.2}$$

Next we present our algorithm for solving problem (4.1). The algorithm alternates between clusters assignment step, which is exactly as in KPALM, and centers update step that is based on a single Weizsfeld iteration.

(1) Initialization: Set $t=0$, and pick random vectors $(w(0),x(0))\in\Delta^m\times\mathbb{R}^{nk}$.

(2) For each $t=0,1,\dots$ generate a sequence $\{(w(t),x(t))\}_{t\in\mathbb{N}}$ as follows:

(2.1) Cluster Assignment: Take any $\alpha_i(t)>0$ and for each $i=1,2,\dots,m$ compute

$$w^i(t+1) = \arg\min_{w^i\in\Delta}\left\{\langle w^i, d^i(x(t))\rangle + \frac{\alpha_i(t)}{2}\|w^i-w^i(t)\|^2\right\}. \tag{4.3}$$

(2.2) Centers Update: For each $l=1,2,\dots,k$ compute $x^l\in\mathbb{R}^n$ via

$$x^l(t+1) = T_l(w(t+1),x(t)). \tag{4.4}$$

Assumption 3. *For any step $t \in \mathbb{N}$ and for all $1 \leq l \leq k$, we assume that $x^l(t) \notin \mathcal{A}$.*

Remark 2. *(i) Due to Assumption 3 it follows that the centers update step in (4.4) is well defined.*

(ii) It is easy to verify that for all $1 \leq l \leq k$ the following equations hold true:

$$\nabla_{x^l} H_l(w, x) = \sum_{i=1}^{m} w_l^i \frac{x^l - a^i}{\|x^l - a^i\|}, \quad \forall x^l \notin \mathcal{A}, \tag{4.5}$$

and that

$$T_l(w, x) = x^l - \frac{1}{L_l(w, x)} \nabla_{x^l} H_l(w, x), \quad \forall x^l \notin \mathcal{A}. \tag{4.6}$$

As in KPALM case, we aim to prove the the sufficient decrease property and subgradient lower bounds for iterates gap property. Note that

$$x^l(t+1) = T_l(w(t+1), x(t)) = \frac{\sum\limits_{i=1}^{m} \frac{w_l^i(t+1)a^i}{\|x^l(t)-a^i\|}}{\sum\limits_{i=1}^{m} \frac{w_l^i(t+1)}{\|x^l(t)-a^i\|}} = \sum_{i=1}^{m} \frac{\frac{w_l^i(t+1)}{\|x^l(t)-a^i\|}}{\sum\limits_{j=1}^{m} \frac{w_l^j(t+1)}{\|x^l(t)-a^j\|}} a^i \in Conv(\mathcal{A}),$$

hence the sequence generated by (Alg-Name) is bounded as well.

Proposition 4.1 (Sufficient decrease property). *Let $\{z(t)\}_{t \in \mathbb{N}} = \{(w(t), x(t))\}_{t \in \mathbb{N}}$ be the sequence generated by (Alg-Name), then there exists $\rho_1 > 0$ such that*

$$\rho_1 \|z(t+1) - z(t)\|^2 \leq \Psi(z(t)) - \Psi(z(t+1)) \quad \forall t \in \mathbb{N}.$$

Proof. In the proof of sufficient decrease property of KPALM we showed that

$$\frac{\alpha(t)}{2} \|w(t+1) - w(t)\|^2 \leq H(w(t), x(t)) - H(w(t+1), x(t)), \tag{4.7}$$

where $\alpha(t) = \min\limits_{1 \leq i \leq m} \{\alpha_i(t)\}$. This part was proven independently of the distance function $d(\cdot, \cdot)$, and since in (Alg-Name) the clusters assignment step is identical to KPALM, the claim is correct in the case of (Alg-Name) as well.

Applying Lemma 4.2 from [reference to Weizsfeld paper] yields

$$\begin{aligned} &H_l(w(t+1), x(t+1)) - H_l(w(t+1), x) \leq \\ &\quad \leq \frac{L_l(w(t+1), x(t))}{2} \left(\|x^l(t) - x^l\|^2 - \|x^l(t+1) - x^l\|^2 \right), \quad \forall x \in \mathbb{R}^{nk}, 1 \leq l \leq k. \end{aligned}$$

Setting $x = x(t)$ and rearranging yields

$$\frac{L_l(w(t+1), x(t))}{2} \|x^l(t+1) - x(t)\|^2 \leq H_l(w(t+1), x(t)) - H_l(w(t+1), x(t+1)), \quad \forall 1 \leq l \leq k. \tag{4.8}$$

Denote $L(t) = \min\limits_{1 \le l \le k} \{L_l(w(t+1), x(t))\}$. Summing (4.8) over $l = 1, 2, \ldots, k$ leads to

What is $L_\ell(t)$?

$$
\begin{aligned}
\frac{L(t)}{2}\|x(t+1) - x(t)\|^2 &= \frac{L(t)}{2}\sum_{l=1}^{k}\|x^l(t+1) - x^l(t)\|^2 \\
&\le \sum_{l=1}^{k}\frac{L_l(t)}{2}\|x^l(t+1) - x^l(t)\|^2 \\
&\le \sum_{l=1}^{k}\left(H_l(w(t+1), x(t)) - H_l(w(t+1), x(t+1))\right) \\
&= H(w(t+1), x(t)) - H(w(t+1), x(t+1)).
\end{aligned}
\tag{4.9}
$$

Set $\rho_1 = \frac{1}{2}\min\{\alpha(t), L(t)\}$, and then ~~again~~ Assumption 1 assures that $\rho_1 > 0$. ~~Combined with~~ Combining (4.7) and (4.9) ~~we have that~~ yields

$$
\begin{aligned}
\rho_1\|z(t+1) - z(t)\|^2 = \rho_1\left(\|w(t+1) - w(t)\|^2 + \|x(t+1) - x(t)\|^2\right) \le \\
\le [H(w(t), x(t)) - H(w(t+1), x(t))] + [H(w(t+1), x(t)) - H(w(t+1), x(t+1))] \\
= H(z(t)) - H(z(t+1)) = \Psi(z(t)) - \Psi(z(t+1)),
\end{aligned}
$$

which proves the desired result. $\square$

Next we prove the subgradient lower bounds for iterates gap property of the sequence generated by (Alg-Name), we start with a useful lemma.

Lemma 4.1.1. *Let $\{z(t)\}_{t\in\mathbb{N}} = \{(w(t), x(t))\}_{t\in\mathbb{N}}$ be the sequence generated by (Alg-Name), then*

$$\|d^i(x(t+1) - d^i(x(t))\| \le \|x(t+1) - x(t)\|, \quad \forall i = 1, 2, \ldots, m,\ t \in \mathbb{N}.$$

Proof. Since $d(u, v) = \|u - v\|$, we get that

$$
\begin{aligned}
\|d^i(x(t+1)) - d^i(x(t))\| &\overset{=}{\not\le} \left[\sum_{l=1}^{k}\left|\|x^l(t+1) - a^i\| - \|x^l(t) - a^i\|\right|^2\right]^{\frac{1}{2}} \\
&\le \left[\sum_{l=1}^{k}\left\|\left(x^l(t+1) - a^i\right) - \left(x^l(t) - a^i\right)\right\|^2\right]^{\frac{1}{2}} \\
&= \left[\sum_{l=1}^{k}\|x^l(t+1) - x^l(t)\|^2\right]^{\frac{1}{2}} = \|x(t+1) - x(t)\|.
\end{aligned}
$$

Should be larger

$\square$

Proposition 4.2 (Subgradient lower bound for iterates gap property). *Let $\{z(t)\}_{t\in\mathbb{N}} = \{(w(t), x(t))\}_{t\in\mathbb{N}}$ be the sequence generated by (Alg-Name), then there exists $\rho_2 > 0$ and $\gamma(t+1) \in \partial\Psi(z(t+1))$ such that*

$$\|\gamma(t+1)\| \le \rho_2\|z(t+1) - z(t)\|, \quad \forall t \in \mathbb{N}.$$

the

Proof. Repeating the steps of the proof in case of KPALM yields that

$$\gamma(t+1) := \Big(\big(d^i(x(t+1)) + u^i(t+1)\big)_{i=1,\dots,m}, \nabla_x H(w(t+1), x(t+1))\Big) \in \partial\Psi(z(t+1)), \quad (4.10)$$

where for all $1 \le i \le m$, $u^i(t+1) \in \partial\delta_\Delta(w^i(t+1))$ such that

$$d^i(x(t)) + \alpha_i(t)\left(w^i(t+1) - w^i(t)\right) + u^i(t+1) = \mathbf{0}. \quad (4.11)$$

Plugging (4.11) into (4.10), and taking norm yields

$$\begin{aligned}
\|\gamma(t+1)\| &\le \sum_{i=1}^{m} \|d^i(x(t+1)) - d^i(x(t)) - \alpha_i(t)\left(w^i(t+1) - w^i(t)\right)\| + \|\nabla_x H(w(t+1), x(t+1))\| \\
&\le \sum_{i=1}^{m} \|d^i(x(t+1)) - d^i(x(t))\| + \sum_{i=1}^{m} \alpha_i(t)\|w^i(t+1) - w^i(t)\| + \|\nabla_x H(w(t+1), x(t+1))\| \\
&\le m\|x(t+1) - x(t)\| + m\overline{\alpha}(t)\|w(t+1) - w(t)\| + \|\nabla_x H(w(t+1), x(t+1))\|,
\end{aligned}$$

delete

where the last inequality follows from Lemma 4.1.1, and the fact that $\overline{\alpha}(t) = \max_{1 \le i \le m} \alpha_i(t)$.

Applying ~~l~~ L emma 4.1 [reference to Weiszfeld paper] with respect to $H_l(w, \cdot)$ yields

$$H_l(w, T_l(w,x)) \le H_l(w,x) + \langle \nabla_{x^l} H_l(w,x)), H_l(w, T_l(w,x)) - H_l(w,x)\rangle + \frac{L_l(w,x)}{2}\|T_l(w,x) - x^l\|^2, \quad \forall x \in (\mathbb{R}\setminus\mathcal{A})^k.$$

X 2

Close review of the proof of ~~l~~ L emma 4.1 reveals that we can switch x with $T_l(w,x)$ and get

$$H_l(w,x) \le H_l(w, T_l(w,x)) + \langle \nabla_{x^l} H_l(w, T_l(w,x)), H_l(w,x) - H_l(w, T_l(w,x))\rangle + \frac{L_l(w, T_l(w,x))}{2}\|x^l - T_l(w,x)\|^2, \quad \forall x \in (\mathbb{R}\setminus\mathcal{A})^k.$$

Summing the last two inequalities and rearranging lead to

*

$$\begin{aligned}
&\langle \nabla_{x^l} H_l(w, T_l(w,x)) - \nabla_{x^l} H_l(w, T_l(w,x)), H_l(w, T_l(w,x)) - H_l(w,x)\rangle \le \\
&\le \frac{L_l(w,x) + L_l(w, T_l(w,x))}{2}\|T_l(w,x) - x^l\|^2, \quad \forall x \in (\mathbb{R}\setminus\mathcal{A})^k.
\end{aligned}$$

Thus, it follows that

$$\begin{aligned}
&\|\nabla_{x^l} H_l(w, T_l(w,x)) - \nabla_{x^l} H_l(w, T_l(w,x))\| \le \\
&\le \frac{L_l(w,x) + L_l(w, T_l(w,x))}{2}\|T_l(w,x) - x^l\|, \quad \forall x \in (\mathbb{R}\setminus\mathcal{A})^k. \quad (4.12)
\end{aligned}$$

will derive the following

. Indeed,

With the last result we shall bound ~~the~~ $\|\nabla_x H(w(t+1), x(t+1))\| \le c\|x(t+1 - x(t)\|$, for some constant $c > 0$, ~~as follows~~ we have

delete

x(t+1) - x(t)

$$\begin{aligned}
&\|\nabla_x H(w(t+1), x(t+1))\| \le \sum_{l=1}^{k} \|\nabla_{x^l} H_l(w(t+1), x(t+1))\| \\
&\le \sum_{l=1}^{k} \|\nabla_{x^l} H_l(w(t+1), x(t))\| + \sum_{l=1}^{k} \|\nabla_{x^l} H_l(w(t+1), x(t+1)) - \nabla_{x^l} H_l(w(t+1), x(t))\| \\
&\le \sum_{l=1}^{k} L_l(w(t+1), x(t))\|x^l(t+1) - x^l(t)\| + \sum_{l=1}^{k} \frac{L_l(w(t+1), x(t)) + L_l(w(t+1), x(t+1))}{2}\|x^l(t+1) - x^l(t)\|,
\end{aligned}$$