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

$$\min_{x^1,x^2,\ldots,x^k\in\mathbb{R}^n} \left\{ \sum_{i=1}^{m} \min_{1\le l\le k} \|x^l - a^i\| \right\}. \tag{4.1}$$

Before proceeding towards the algorithm that is based on PALM theory, we need to develop some useful ~~results~~.

tools

4.1 The Smoothed Fermat-Weber Problem

Solving the smoothed Fermat-Weber plays a significant role in the algorithm that addresses the clustering problem with Euclidean norm distance-like function. The Fermat-Weber problem is formulated as follows

$$\min_{x\in\mathbb{R}^n} \left\{ f(x) := \sum_{i=1}^{m} w_i \|x - a^i\| \right\}, \tag{4.2}$$

where $w_i > 0$, $i = 1, 2, \ldots, m$, are given positive weights and $\mathcal{A} = \{a^1, a^2, \ldots, a^m\} \subset \mathbb{R}^n$ are given vectors. As shown in [BS2015] this problem can be solved via the consecutive appliance of the operator $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$T(x) = \frac{\sum_{i=1}^{m} \frac{w_i a^i}{\|x-a^i\|}}{\sum_{i=1}^{m} \frac{w_i}{\|x-a^i\|}}.$$

$$\frac{1}{\sum \frac{w_i}{\|x-a^i\|}} \sum \frac{w_i a^i}{\|x-a^i\|}$$

It is easily noticed that $f(x)$ is not differentiable over $\mathcal{A}$. For our purposes we are interested in the smoothed Fermat-Weber problem, that can be formulated in the following manner

$$\min_{x\in\mathbb{R}^n} \left\{ f_\epsilon(x) := \sum_{i=1}^{m} w_i \left(\|x - a^i\|^2 + \epsilon^2\right)^{1/2} \right\}, \tag{4.3}$$

with $\epsilon > 0$ being some small perturbation constant. Next we introduce the operator $T_\epsilon : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$T_\epsilon(x) = \frac{\sum_{i=1}^{m} \frac{w_i a^i}{(\|x-a^i\|^2+\epsilon^2)^{1/2}}}{\sum_{i=1}^{m} \frac{w_i}{(\|x-a^i\|^2+\epsilon^2)^{1/2}}}.$$

$$\frac{1}{\sum \frac{w_i}{}} \sum \frac{w_i a^i}{}$$

This version of the sm~~o~~oth ~~T~~ operator together with its properties that are to be discussed below are the cornerstone to prove the properties needed by PALM, and in turn to show the convergence of the sequence generated by the algorithm proposed to tackle the smooth version of the clustering problem presented later on. In order to prove some properties of T_ϵ, which are the same as the properties of T described in [BS2015], we also will need an auxiliary function $h_\epsilon : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$h_\epsilon(x, y) = \sum_{i=1}^{m} \frac{w_i \left(\|x - a^i\|^2 + \epsilon^2\right)}{\left(\|y - a^i\|^2 + \epsilon^2\right)^{1/2}}.$$

Another useful function $L_\epsilon : \mathbb{R}^n \to \mathbb{R}$ that serves somewhat like Lipschitz function for the gradient of f_ϵ is defined by

$$L_\epsilon(x) = \sum_{i=1}^{m} \frac{w_i}{(\|x - a^i\|^2 + \epsilon^2)^{1/2}}.$$

It is easy to verify the following equality

$x - \frac{1}{L_\epsilon(x)} \nabla f_\epsilon(x)$

$$T_\epsilon(x) = x - \frac{\nabla f_\epsilon(x)}{L_\epsilon(x)}, \forall x \in \mathbb{R}^n. \tag{4.4}$$

Lemma 4.0.1 (Properties of the auxiliary function h_ϵ). *The following properties of h_ϵ hold.*

(i) For any $y \in \mathbb{R}^n$,

$$h_\epsilon(y, y) = f_\epsilon(y).$$

(ii) For any $x, y \in \mathbb{R}^n$,

$$h_\epsilon(x, y) \geq 2f_\epsilon(x) - f_\epsilon(y).$$

(iii) For any $y \in \mathbb{R}^n$,

$$T_\epsilon(y) = \arg\min_{x \in \mathbb{R}^n} h_\epsilon(x, y).$$

(iv) For any $x, y \in \mathbb{R}^n$,

$$h_\epsilon(x, y) = h_\epsilon(y, y) + \langle \nabla_x h_\epsilon(y, y), x - y \rangle + L_\epsilon(y)\|x - y\|^2.$$

Proof. (i) Follows by substituting $x = y$ in $h(x, y)$.

(ii) For any two numbers $a \in \mathbb{R}$ and $b > 0$ the inequality

$$\frac{a^2}{b} \geq 2a - b,$$

holds true. Thus for every $i = 1, 2, \ldots, m$ we have that

$$\frac{\|x - a^i\|^2 + \epsilon^2}{(\|y - a^i\|^2 + \epsilon^2)^{1/2}} \geq 2\left(\|x - a^i\|^2 + \epsilon^2\right)^{1/2} - \left(\|y - a^i\|^2 + \epsilon^2\right)^{1/2}.$$

Multiplying the last inequality by w_i and summing over $i = 1, 2, \ldots, m$, and the results follows.

h_ϵ

(iii) The function $x \mapsto h(x, y)$ is strongly convex and its unique minimizer is determined by the optimality equation

$$\nabla_x h_\epsilon(x, y) = \sum_{i=1}^{m} \frac{2w_i\,(x - a^i)}{(\|y - a^i\|^2 + \epsilon^2)^{1/2}} = 0.$$

Simple algebraic manipulation leads to the relation

$$x = T_\epsilon(y),$$

and the desired results follows.

(iv) The function $x \mapsto h_\epsilon(x, y)$ is quadratic with associated matrix $L_\epsilon(y)\mathbf{I}$. Therefore, its second-order taylor expansion around y leads to the desired result.

□

h_ϵ

The following lemma proofs are based on the properties of the auxiliary function $\cancel{H}_\epsilon$, and they are similar to the proofs in [BS2015], hence we will just state them here.

Lemma 4.0.2 (Monotonicity property of T_ϵ, similar to (BS2015, Lemma 3.2, page 7)). *For every $y \in \mathbb{R}^n$ we have*

$$f_\epsilon(T_\epsilon(y)) \le f_\epsilon(y).$$

Lemma 4.0.3 (Decent lemma for function f_ϵ, similar to (BS2015, Lemma 5.1, page 10)). *For every $y \in \mathbb{R}^n$ we have*

$$f_\epsilon(T_\epsilon(y)) \le f_\epsilon(y) + \langle \nabla f_\epsilon(y), T_\epsilon(y) - y\rangle + \frac{L_\epsilon(y)}{2}\|T_\epsilon(y) - y\|^2.$$

Lemma 4.0.4 (Similar to (BS2015, Lemma 5.2, page 12)). *For every $x, y \in \mathbb{R}^n$ we have*

$$f_\epsilon(T_\epsilon(y)) - f_\epsilon(x) \le \frac{L_\epsilon(y)}{2}\left(\|y - x\|^2 - \cancel{T_\epsilon(y) - x}\|^2\right).$$

$\|T_\epsilon(y) - x\|^2$

Lemma 4.0.5. *For all $y^0, y \in \mathbb{R}^n$ the following statement holds true*

$$\|\nabla f_\epsilon(y) - \nabla f_\epsilon(y^0)\| \le \frac{2L_\epsilon(y^0)L_\epsilon(y)}{L_\epsilon(y^0) + L_\epsilon(y)}\|y^0 - y\|.$$

Proof. Let $y^0 \in \mathbb{R}^n$ be a fixed vector. Define the following two functions

$$\widetilde{f}_\epsilon(y) = f_\epsilon(y) - \left\langle \nabla f_\epsilon(y^0), y\right\rangle,$$

and

$$\widetilde{h}_\epsilon(x, y) = h_\epsilon(x, y) - \left\langle \nabla f_\epsilon(y^0), x\right\rangle.$$

from Lemma 4.0.1(i)

It is clear that $x \mapsto \widetilde{h}_\epsilon(x, y)$ is still quadratic function with associated matrix $L_\epsilon(y)\mathbf{I}$. Therefore, we can write

$$\begin{aligned}\widetilde{h}_\epsilon(x, y) &= \widetilde{h}_\epsilon(y, y) + \left\langle \nabla_x \widetilde{h}_\epsilon(y, y), x - y\right\rangle + L_\epsilon(y)\|x - y\|^2 \\ &= \widetilde{f}_\epsilon(y) + \left\langle 2\nabla f_\epsilon(y) - \nabla f_\epsilon(y^0), x - y\right\rangle + L_\epsilon(y)\|x - y\|^2.\end{aligned} \tag{4.5}$$

Lemma 4.0.1(ii)

On the other hand, from ~~Equation (4.4) (ii)~~ we have that

$$\begin{aligned}\widetilde{h}_\epsilon(x, y) &= h_\epsilon(x, y) - \left\langle \nabla f_\epsilon(y^0), x\right\rangle \ge 2f_\epsilon(x) - f_\epsilon(y) - \left\langle \nabla f_\epsilon(y^0), x\right\rangle \\ &= 2\widetilde{f}_\epsilon(x) - \widetilde{f}_\epsilon(y) + \left\langle \nabla f_\epsilon(y^0), \cancel{y - x}\right\rangle,\end{aligned} \tag{4.6}$$

$x - y$

where the last equality follows from the definition of $\widetilde{f}_\epsilon$. Combining (4.5) and (4.6) yields

$$\begin{aligned}2\widetilde{f}_\epsilon(x) &\le 2\widetilde{f}_\epsilon(y) + 2\left\langle \nabla f_\epsilon(y) - \nabla f_\epsilon(y^0), x - y\right\rangle + L_\epsilon(y)\|x - y\|^2 \\ &= 2\widetilde{f}_\epsilon(y) + 2\left\langle \nabla \widetilde{f}_\epsilon(y), x - y\right\rangle + L_\epsilon(y)\|x - y\|^2.\end{aligned}$$

Dividing the last inequality by 2 leads to

$$\widetilde{f}_\epsilon(x) \le \widetilde{f}_\epsilon(y) + \left\langle \nabla \widetilde{f}_\epsilon(y), x - y\right\rangle + \frac{L_\epsilon(y)}{2}\|x - y\|^2. \tag{4.7}$$

using ~~from~~ (4.7) with $x = y - \frac{1}{L_\epsilon(y)} \nabla \widetilde{f}_\epsilon(y)$ yields ~~we obtain~~

It is clear that the optimal point of $\widetilde{f}_\epsilon$ is y^0 since $\nabla \widetilde{f}_\epsilon(y^0) = 0$, therefore ~~from~~ (4.7) ~~we obtain~~

$$\begin{aligned}\widetilde{f}_\epsilon(y^0) &\leq \widetilde{f}_\epsilon\left(y - \frac{1}{L_\epsilon(y)}\nabla \widetilde{f}_\epsilon(y)\right) \leq \widetilde{f}_\epsilon(y) + \left\langle \nabla \widetilde{f}_\epsilon(y), -\frac{1}{L_\epsilon(y)}\nabla \widetilde{f}_\epsilon(y)\right\rangle + \frac{L_\epsilon(y)}{2}\left\|\frac{1}{L_\epsilon(y)}\nabla \widetilde{f}_\epsilon(y)\right\|^2 \\ &= \widetilde{f}_\epsilon(y) - \frac{1}{2L_\epsilon(y)}\left\|\nabla \widetilde{f}_\epsilon(y)\right\|^2.\end{aligned}$$

Thus, using the definition of $\widetilde{f}_\epsilon$ and the fac~~x~~ that $\nabla \widetilde{f}_\epsilon(y) = \nabla f_\epsilon(y) - \nabla f_\epsilon(y^0)$, yields that

$$f_\epsilon(y^0) \leq f_\epsilon(y) + \langle \nabla f_\epsilon(y^0), y^0 - y\rangle - \frac{1}{2L_\epsilon(y)}\|\nabla f_\epsilon(y) - \nabla f_\epsilon(y^0)\|^2.$$

Now, following the same arguments we can show that

$$f_\epsilon(y) \leq f_\epsilon(y^0) + \langle \nabla f_\epsilon(y), y - y^0\rangle - \frac{1}{2L_\epsilon(y^0)}\|\nabla f_\epsilon(y^0) - \nabla f_\epsilon(y)\|^2$$

and combining last two inequalities yields that

$$\left(\frac{1}{2L_\epsilon(y^0)} + \frac{1}{2L_\epsilon(y)}\right)\|\nabla f_\epsilon(y) - \nabla f_\epsilon(y^0)\|^2 \leq \langle \nabla f_\epsilon(y^0) - \nabla f_\epsilon(y), y^0 - y\rangle,$$

that is

$$\|\nabla f_\epsilon(y) - \nabla f_\epsilon(y^0)\| \leq \frac{2L_\epsilon(y^0)L_\epsilon(y)}{L_\epsilon(y^0) + L_\epsilon(y)}\|y^0 - y\|,$$

for all $y^0, y \in \mathbb{R}^n$. □

4.2 Clustering with T_ϵ Operator

In the previous section 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

$$H(w,x) = \sum_{i=1}^{m}\langle w^i, d^i(x)\rangle = \sum_{i=1}^{m}\sum_{l=1}^{k} w_l^i\|x^l - a^i\|,$$

and

$$G(w) = \sum_{i=1}^{m}\delta_\Delta(w^i).$$

However, in order to be able to use the theory of PALM, we need the coupled function $H(w,x)$ to be smooth, and in our case it is not. Therefore, it leads us to the following smoothed form of the clustering problem

$$\min\left\{\Psi_\epsilon(z) := H_\epsilon(w,x) + G(w) \mid z := (w,x) \in \mathbb{R}^{km} \times \mathbb{R}^{nk}\right\}, \tag{4.8}$$

where

$$H_\epsilon(w,x) = \sum_{i=1}^{m}\langle w^i, d_\epsilon^i(x)\rangle = \sum_{i=1}^{m}\sum_{l=1}^{k} w_l^i\left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2},$$

with $d_\epsilon^i(x) = \left(\left(\|x^1 - a^i\|^2 + \epsilon^2\right)^{1/2}, \left(\|x^2 - a^i\|^2 + \epsilon^2\right)^{1/2}, \ldots, \left(\|x^k - a^i\|^2 + \epsilon^2\right)^{1/2} \right) \in \mathbb{R}^k$, for $i = 1, 2, \ldots, m$. Note that $\Psi_\epsilon(z)$ is a perturbed form of $\Psi(z)$ for some small $\epsilon > 0$.

Next we extend the notations of the previous subsection, so that the functions and operators defined there are to be dependent on the weights w. For each $1 \le l \le k$, denote $w_l = \left(w_l^1, w_l^2, \ldots, w_l^m\right) \in \mathbb{R}_+^m$ and define

$$L_\epsilon^{w_l}(x^l) = \sum_{i=1}^{m} \frac{w_l^i}{\left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2}},$$

and

$$T_\epsilon^{w_l}(x^l) = \frac{\sum_{i=1}^{m} \frac{w_l^i a^i}{\left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2}}}{L_\epsilon^{w_l}(x^l)}.$$

$\frac{1}{L_\varepsilon^{w_l}(x^l)} \sum \frac{w_l^i a^i}{}$

For all $1 \le l \le k$ we define $H_\epsilon^{w_l} : \mathbb{R}^n \to \mathbb{R}$ as follows

$$H_\epsilon^{w_l}(x^l) = \sum_{i=1}^{m} w_l^i \left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2},$$

thus we have

$$H_\epsilon(w, x) = \sum_{l=1}^{k} H_\epsilon^{w_l}(x^l).$$

Now we present our algorithm for solving problem (4.8), we call it ε-KPALM. The algorithm alternates between cluster assignment step, similar to that as in KPALM, and centers update step that is based on a T_ϵ operator.

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

(2) For each $t = 0, 1, \ldots$ 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 $i = 1, 2, \ldots, m$ compute

$$\begin{aligned} w^i(t+1) &= \arg\min_{w^i \in \Delta} \left\{ \langle w^i, d_\epsilon^i(x(t)) \rangle + \frac{\alpha_i(t)}{2} \|w^i - w^i(t)\|^2 \right\} \\ &= P_\Delta \left(w^i(t) - \frac{d_\epsilon^i(x(t))}{\alpha_i(t)} \right). \end{aligned} \tag{4.9}$$

(2.2) Center Update: For each $l = 1, 2, \ldots, k$ compute

$$x^l(t+1) = T_\epsilon^{w_l(t+1)}(x^l(t)). \tag{4.10}$$

Remark 2. *(i) Assumption 1 is still valid, hence the center update step in (4.10) is well defined.*

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

space

$$\nabla H_\epsilon^{w_l}(x^l) = \sum_{i=1}^{m} w_l^i \frac{x^l - a^i}{\left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2}}, \quad \forall x^l \in \mathbb{R}^n, \tag{4.11}$$

and that

$$T_\epsilon^{w_l}(x^l) = x^l - \frac{1}{L_\epsilon^{w_l}(x^l)} \nabla H_\epsilon^{w_l}(x^l), \quad \forall x^l \in \mathbb{R}^n. \tag{4.12}$$

As in KPALM case, the sequence that is generated by ε-KPALM is contained within the convex hull of $\mathcal{A}$. Indeed,

$$x^l(t+1) = T_\epsilon^{w_l(t+1)}(x^l(t)) = \frac{\sum_{i=1}^{m} \frac{w_l^i(t+1)a^i}{\left(\|x^l(t)-a^i\|^2+\epsilon^2\right)^{1/2}}}{\sum_{i=1}^{m} \frac{w_l^i(t+1)}{\left(\|x^l(t)-a^i\|^2+\epsilon^2\right)^{1/2}}} = \sum_{i=1}^{m} \left(\frac{\frac{w_l^i(t+1)}{\left(\|x^l(t)-a^i\|^2+\epsilon^2\right)^{1/2}}}{\sum_{j=1}^{m} \frac{w_l^j(t+1)}{\left(\|x^l(t)-a^j\|^2+\epsilon^2\right)^{1/2}}} \right) a^i \in Conv(\mathcal{A}),$$

hence the sequence generated by ε-KPALM is bounded as well.

Now we are finally ready to prove the properties needed by PALM, and deduce that the sequence that is generated by ε-KPALM converge to critical point of Ψ_ϵ.

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 ε-KPALM, then there exists $\rho_1 > 0$ such that*

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

Proof. Similar steps to the ones in the proof of sufficient decrease property of KPALM lead to

$$\frac{\alpha(t)}{2}\|w(t+1) - w(t)\|^2 \leq H_\epsilon(w(t), x(t)) - H_\epsilon(w(t+1), x(t)), \tag{4.13}$$

where $\alpha(t) = \min_{1\leq i\leq m} \{\alpha_i(t)\}$.

Applying Lemma 4.0.4 with respect to $H_\epsilon^{w_l(t+1)}(\cdot)$ yields

space

$$H_\epsilon^{w_l(t+1)}(x^l(t+1)) - H_\epsilon^{w_l(t+1)}(x^l) \leq \frac{L_\epsilon^{w_l(t+1)}(x^l(t))}{2}\left(\|x^l(t) - x^l\|^2 - \|x^l(t+1) - x^l\|^2\right), \quad \forall x^l \in \mathbb{R}^n,$$

for all $l = 1, 2, \ldots, k$. Setting $x^l = x^l(t)$ and rearranging yields

$$\frac{L_\epsilon^{w_l(t+1)}(x^l(t))}{2}\|x^l(t+1) - x^l(t)\|^2 \leq H_\epsilon^{w_l(t+1)}(x^l(t)) - H_\epsilon^{w_l(t+1)}(x^l(t+1)), \quad \forall 1 \leq l \leq k. \tag{4.14}$$

Denote $L(t) = \min_{1\leq l\leq k}\left\{L_\epsilon^{w_l(t+1)}(x^l(t))\right\}$. Summing (4.14) over $l = 1, 2, \ldots, k$ leads to

$$\begin{aligned}
\frac{L(t)}{2}\|x(t+1) - x(t)\|^2 &= \frac{L(t)}{2}\sum_{l=1}^{k}\|x^l(t+1) - x^l(t)\|^2 \\
&\leq \sum_{l=1}^{k}\frac{L_\epsilon^{w_l(t+1)}(x^l(t))}{2}\|x^l(t+1) - x^l(t)\|^2 \\
&\leq \sum_{l=1}^{k}\left(H_\epsilon^{w_l(t+1)}(x^l(t)) - H_\epsilon^{w_l(t+1)}(x^l\cancel{(t)})\right) \\
&= H_\epsilon(w(t+1), x(t)) - H_\epsilon(w(t+1), x(t+1)).
\end{aligned} \tag{4.15}$$

$x^l(t+1)$

$t\in N$

Set $\rho_1 = \frac{1}{2}\min\{\alpha(t), L(t)\}$, and note that Assumption 1 assures that $\rho_1 > 0$. Combining (4.13) and (4.15) yields

$$\begin{aligned}
\rho_1\|z(t+1) - z(t)\|^2 &= \rho_1\left(\|w(t+1) - w(t)\|^2 + \|x(t+1) - x(t)\|^2\right) \leq \\
&\leq [H_\epsilon(w(t), x(t)) - H_\epsilon(w(t+1), x(t))] + [H_\epsilon(w(t+1), x(t)) - H_\epsilon(w(t+1), x(t+1))] \\
&= H_\epsilon(z(t)) - H_\epsilon(z(t+1)) = \Psi_\epsilon(z(t)) - \Psi_\epsilon(z(t+1)),
\end{aligned}$$

Plugging (4.17) into (4.16), and taking norm yields

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

where the last inequality follows from Lemma 4.1.1 and the fact that $\overline{\alpha}(t) = \max_{1\le i\le m} \alpha_i(t)$.
Next we bound $\|\nabla_x H_\epsilon(w(t+1), x(t+1))\| \le c\|x(t+1) - x(t)\|$, for some constant $c > 0$. Indeed, we have

$$\begin{aligned}
\|\nabla_x H_\epsilon(w(t+1), x(t+1))\| &\le \sum_{l=1}^{k} \|\nabla H_\epsilon^{w_l(t+1)}(x^l(t+1))\| \\
&\le \sum_{l=1}^{k} \|\nabla H_\epsilon^{w_l(t+1)}(x^l(t))\| + \sum_{l=1}^{k} \|\nabla H_\epsilon^{w_l(t+1)}(x^l(t+1)) - \nabla H_\epsilon^{w_l(t+1)}(x^l(t))\|.
\end{aligned} \tag{4.18}$$

(4.10) and

From (4.12) we have

$$\nabla H_\epsilon^{w_l(t+1)}(x^l(t)) = L_\epsilon^{w_l(t+1)}(x^l(t))\left(x^l(t+1) - x^l(t)\right), \quad \forall 1 \le l \le k,$$

space

applying Lemma 4.0.5 with respect to $H_\epsilon^{w_l(t+1)}(\cdot)$ and plugging into (4.18) yields

$$\begin{aligned}
&\|\nabla_x H(w(t+1), x(t+1))\| \le \\
&\le \sum_{l=1}^{k} \left(L_\epsilon^{w_l(t+1)}(x^l(t)) + \frac{2L_\epsilon^{w_l(t+1)}(x^l(t)) L_\epsilon^{w_l(t+1)}(x^l(t+1))}{L_\epsilon^{w_l(t+1)}(x^l(t)) + L_\epsilon^{w_l(t+1)}(x^l(t+1))} \right) \|x^l(t+1) - x^l(t)\|.
\end{aligned}$$

Therefore, denote $\overline{L}(t) = \max_{1\le l\le k} \left\{ L_\epsilon^{w_l(t+1)}(x^l(t)) + \frac{2L_\epsilon^{w_l(t+1)}(x^l(t)) L_\epsilon^{w_l(t+1)}(x^l(t+1))}{L_\epsilon^{w_l(t+1)}(x^l(t)) + L_\epsilon^{w_l(t+1)}(x^l(t+1))} \right\}$, and set $\rho_2 = m + m\overline{\alpha}(t) + k\overline{L}(t)$, and the result follows. □

ρ_2 should be constant and therefore can't be depended on t.
Hence, you might define

$$\rho_2 = m + \max_{t\in\mathbb{N}} \{m\overline{\alpha}(t) + k\overline{L}(t)\}$$

(similar argument should be fixed in Prop. 3.2)

In addition, can we say something on the boundedness from above of $\overline{\alpha}(t)$ and $\overline{L}(t)$?
In the PALM paper, we assume it but maybe here we can have it from the specific structure?

which proves the desired result. ∎

The next lemma will be useful in proving the subgradient lower bounds for iterates gap property of the sequence generated by ε-KPALM.

Lemma 4.1.1. *For any* $x, y \in \mathbb{R}^n$ *such that* $x^l, y^l \in Conv(\mathcal{A})$ *for all* $1 \le l \le k$ *the following inequality holds* (space)

$$\|d_\epsilon^i(x) - d_\epsilon^i(y)\| \le \frac{d_{\mathcal{A}}}{\epsilon}\|x - y\|, \quad \forall i = 1, 2, \ldots, m,$$

with $d_{\mathcal{A}} = diam(Conv(\mathcal{A}))$.

the

Proof. Define $\psi(t) = \sqrt{t + \epsilon^2}$, for $t \ge 0$. Using Lagrange mean value theorem over $a > b \ge 0$ yields

$$\frac{\psi(a) - \psi(b)}{a - b} = \psi'(c) = \frac{1}{2\sqrt{c + \epsilon^2}} \le \frac{1}{2\epsilon},$$

where $c \in (b, a)$. Therefore, for all $i = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, k$ we have

$\frac{1}{2\epsilon}$

$$\begin{aligned}\left|\left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2} - \left(\|y^l - a^i\|^2 + \epsilon^2\right)^{1/2}\right| &\le \cancel{(1/2\epsilon)}\left|\|x^l - a^i\|^2 + \epsilon^2 - \left(\|y^l - a^i\|^2 + \epsilon^2\right)\right| \\ &= \cancel{(1/2\epsilon)}\left|\|x^l - a^i\|^2 - \|y^l - a^i\|^2\right| \\ &= (1/2\epsilon)\left|\|x^l - a^i\| + \|y^l - a^i\|\right| \cdot \left|\|x^l - a^i\| - \|y^l - a^i\|\right| \\ &\le \cancel{\frac{d_{\mathcal{A}}\|x^l - y^l\|}{\epsilon}}.\end{aligned}$$

$\frac{1}{\epsilon} d_{\mathcal{A}} \|x^l - y^l\|$

Hence,

$$\begin{aligned}\|d_\epsilon^i(x) - d_\epsilon^i(y)\| &= \left[\sum_{l=1}^{k}\left|\left(\|x - a^i\|^2 + \epsilon^2\right)^{1/2} - \left(\|y - a^i\|^2 + \epsilon^2\right)^{1/2}\right|^2\right]^{\frac{1}{2}} \\ &\le \left[\sum_{l=1}^{k}\left(\cancel{\frac{d_{\mathcal{A}}\|x^l - y^l\|}{\epsilon}}\right)^2\right]^{\frac{1}{2}} \\ &= \frac{d_{\mathcal{A}}}{\epsilon}\|x - y\|,\end{aligned}$$

$\frac{1}{\epsilon} d_{\mathcal{A}} \|x^l - y^l\|$

as asserted ∎

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* ε*-KPALM, then there exists* $\rho_2 > 0$ *and* $\gamma(t+1) \in \partial\Psi_\epsilon(z(t+1))$ *such that* (space)

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

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

$$\gamma(t+1) := \left(\left(d_\epsilon^i(x(t+1)) + u^i(t+1)\right)_{i=1,\ldots,m}, \nabla_x H_\epsilon(w(t+1), x(t+1))\right) \in \partial\Psi_\epsilon(z(t+1)), \quad (4.16)$$

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

$$d_\epsilon^i(x(t)) + \alpha_i(t)\left(w^i(t+1) - w^i(t)\right) + u^i(t+1) = \mathbf{0}. \quad (4.17)$$