$$\min_{x\in\mathbb{R}^{nk}} \left\{ F(x) := \sum_{i=1}^{m} \min_{1\le l\le k} d(x^l, a^i) \right\},$$

1 The Clustering Problem

Let $\mathcal{A} = \{a^1, a^2, \ldots, a^m\}$ be a given set of points in $\mathbb{R}^n$, and let $1 < k < m$ be a fixed given number of clusters. The clustering problem consists of partitioning the data $\mathcal{A}$ into k subsets $\{C^1, C^2, \ldots, C^k\}$, called clusters. For each $l = 1, 2, \ldots, k$, the cluster C^l is represented by its center x^l $\in\mathbb{R}^n$~~, and we want~~ to determine k cluster centers $\{x^1, x^2, \ldots, x^k\}$ such that the sum of certain proximity measures from each point $a^i, i = 1, 2, \ldots, m$, to a nearest cluster center x^l is minimized. We denote the vector of all centers by $x = (x^1, x^2, \ldots, x^k) \in \mathbb{R}^{nk}$.

The clustering problem is given by

$$\cancel{\min_{x^1,x^2,\ldots,x^k\in\mathbb{R}^n} F(x^1, x^2, \ldots, x^k) := \sum_{i=1}^{m} \min_{1\le l\le k} d(x^l, a^i),} \tag{1.1}$$

with $d(\cdot,\cdot)$ being a distance-like function.

2 Problem Reformulation and Notations

~~We introduce some notations that will be used throughout this document.~~

We begin with a reformulation of the clustering problem which will be the basis for our developments in this work. The reformulation is based on the following fact:

$$\min_{1\le l\le k} u^l = \min_{v} \{ \langle u, v\rangle : v\in\Delta \},$$

where Δ is the well-known simplex defined by

$$\Delta = \left\{ u\in\mathbb{R}^k : \sum_{l=1}^{k} u^l = 1,\ u\ge 0 \right\}.$$

Using this fact in Problem (1.1) and introducing new variables $w^i \in \mathbb{R}^k, i = 1, 2, \ldots, m$,

~~$a = (a^1, a^2, \ldots, a^m) \in \mathbb{R}^{nm}$, where $a^i \in \mathbb{R}^n, i = 1, 2, \ldots, m$.~~

~~$w = (w^1, w^2, \ldots, w^m) \in \mathbb{R}^{km}$, where $w^i \in \mathbb{R}^k, i = 1, 2, \ldots, m$.~~

~~$x = (x^1, x^2, \ldots, x^k) \in \mathbb{R}^{nk}$, where $x^l \in \mathbb{R}^n, l = 1, 2, \ldots, k$.~~

~~$d^i(x) = (d(x^1, a^i), d(x^2, a^i), \ldots, d(x^k, a^i)) \in \mathbb{R}^k, i = 1, 2, \ldots, m$.~~

$$\cancel{\Delta = \left\{ u\in\mathbb{R}^k \mid \sum_{l=1}^{k} u_l = 1,\ u_l \ge 0,\ l = 1, 2, \ldots, k \right\}.}$$

~~Let $S \subseteq \mathbb{R}^n$. The indicator function of S is defined and denoted as follows $\delta_S(p) = \begin{cases} 0, & \text{if } p\in S, \\ \infty, & \text{if } p\notin S. \end{cases}$~~

Us~~ing the fact that $\min_{1\le l\le k} u_l = \min\{\langle u, v\rangle \mid v\in\Delta\}$, and applying it over (1.1)~~, gives a smooth reformulation of the clustering problem

$$\min_{x\in\mathbb{R}^{nk}} \sum_{i=1}^{m} \min_{w^i\in\Delta} \langle w^i, d^i(x)\rangle, \tag{2.1}$$

where $d^i(x) = (d(x^1, a^i), d(x^2, a^i), \ldots, d(x^k, a^i)) \in \mathbb{R}^k$, $i = 1, 2, \ldots, m$.

Replacing further the constraint $w^i \in \Delta$ by adding the indicator function $\delta_\Delta(\cdot)$, which defined to be 0 in Δ and ∞ otherwise, to the objective function, results in a equivalent formulation

$$\min_{x\in\mathbb{R}^{nk}, w\in\mathbb{R}^{km}} \left\{ \sum_{i=1}^{m} \left(\langle w^i, d^i(x)\rangle + \delta_\Delta(w^i) \right) \right\}, \tag{2.2}$$

where $w = (w^1, w^2, \ldots, w^m) \in \mathbb{R}^{km}$.

Finally, i~~ntroducing several more useful notations is needed. For each $i = 1, 2, \ldots, m$, we denote~~ For the simplicity of the yet to come expositions, we define the following functions

\quad

$$H(w,x) := \sum_{i=1}^{m} H_i(w,x) = \sum_{i=1}^{m}\langle w^i, d^i(x)\rangle \text{ and } G(w) = \sum_{i=1}^{m} G(w^i) := \sum_{i=1}^{m}\delta_\Delta(w^i).$$

Replacing the terms in ~~(2.1)~~ (2.2) with the functions defined above gives a compact equivalent form of the original clustering problem

$$\min\left\{\Psi(z) := H(w,x) + G(w) \mid z := (w,x) \in \mathbb{R}^{km} \times \mathbb{R}^{nk}\right\}. \tag{2.3}$$

3 Clustering via PALM Approach

3.1 Introduction to PALM Theory

Presentation of PALM's requirements and of the algorithm steps ...

3.2 Clustering with PALM for Squared Euclidean Norm Distance Function

In this section we tackle the clustering problem, given in (2.3), with the classical distance function defined by $d(u,v) = \|u-v\|^2$. We devise a PALM-like algorithm, based on the discussion about PALM in the previous subsection. Since the clustering problem has a specific structure, we are ought to exploit it in the following manner. ~~First we notice that the function $w \mapsto H(w,x)$ is linear in w, so there is no need to linearize it. In addition, the function $x \mapsto H(w,x) = \sum_{i=1}^{m}\sum_{l=1}^{k} w_l^i\|x^l - a^i\|^2 = \sum_{l=1}^{k}\sum_{i=1}^{m} w_l^i\|x^l - a^i\|^2$ is convex and quadratic in x, hence we do not need to add a proximal term as in PALM algorithm.~~

~~Now we propose a PALM-like algorithm for the clustering problem, which we call KPALM.~~

(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

$$\cancel{w^i(t+1) = \arg\min_{w^i\in\Delta}\left\{\langle w^i, d^i(x(t))\rangle + \frac{\alpha_i(t)}{2}\|w^i - w^i(t)\|^2\right\}.} \tag{3.1}$$

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

$$\cancel{x(t+1) = \arg\min\left\{H(w(t+1), x) \mid x \in \mathbb{R}^{nk}\right\}.} \tag{3.2}$$

① The function $w \longrightarrow H(w,x)$, for fixed x, is linear and therefore there is no need to linearize it as suggested in PALM.

② The function $x \longrightarrow H(w,x)$, for fixed w, is quadratic and convex. Hence, there is no need to add a quadratic proximal term as suggested in PALM.

~~Based on these observations, we propose the following algorithm.~~ As in the PALM algorithm, our algorithm is based on alternating minimization, with the following adaptations which are motivated by the observations

Assumption 2. *For any step $t \in \mathbb{N}$, each $a \in \mathcal{A}$ belongs exclusively to single cluster $C^l(t)$.*

For any $x \in \mathbb{R}^{nk}$ we denote the super-partition of $\mathcal{A}$ with respect to x by $\overline{C^l}(x) = \{a \in \mathcal{A} \mid \|a - x^l\| \leq \|a - x^j\|, \quad \forall j \neq l\}$, for all $1 \leq l \leq k$, and the sub-partition of $\mathcal{A}$ by $\underline{C^l}(x) = \{a \in \mathcal{A} \mid \|a - x^l\| < \|a - x^j\|, \quad \forall j \neq l\}$. Moreover, denote $R_{lj}(t) = \min_{a \in C^l(t)} \{\|a - x^j(t)\| - \|a - x^l(t)\|\}$ for all $1 \leq l, j \leq k$, and $r(t) = \min_{l \neq j} R_{lj}$.
Due to Assumption 2 we have that $\overline{C^l}(x(t)) = \underline{C^l}(x(t)) = C^l(t+1)$, for all $1 \leq l \leq k$, $t \in \mathbb{N}$, we also have that $r(t) > 0$ for all $t \in \mathbb{N}$.

Proposition 3.3. *Let $(C(t), x(t))$ be the clusters and centers KMEANS returns. Denote an open neighbourhood of $x(t)$ by $U = B\left(x^1(t), \frac{r(t)}{2}\right) \times B\left(x^2(t), \frac{r(t)}{2}\right) \times \cdots \times B\left(x^l(t), \frac{r(t)}{2}\right)$, then for any $x \in U$ we have $\underline{C^l}(x) = C^l(t)$ for all $1 \leq l \leq k$. Let $(C(t), x(t))$ be the clusters and centers KMEANS returns. Denote by $U = B\left(x^1(t), \frac{r(t)}{2}\right) \times B\left(x^2(t), \frac{r(t)}{2}\right) \times \cdots \times B\left(x^l(t), \frac{r(t)}{2}\right)$ an open neighbourhood of $x(t)$, then for any $x \in U$ we have $C^l(t) = \underline{C^l}(x)$ for all $1 \leq l \leq k$.*

Proof. Pick some $a \in C^l(t)$, then $x^l(t-1)$ is the closest center among the centers of $x(t-1)$. Since KMEANS halts at step t, then from (3.12) we have $x(t) = x(t-1)$, thus $x^l(t)$ is the closest center to a among the centers of $x(t)$. Further we have

$$r(t) \leq \|x^j(t) - a\| - \|x^l(t) - a\| \quad \forall j \neq l. \tag{3.13}$$

Next, we show that $a \in \underline{C^l}(x)$, indeed

$$\begin{aligned}
\|a - x^l\| - \|a - x^j\| &\leq \|a - x^l(t)\| + \|x^l(t) - x^l\| - \left(\|a - x^j(t)\| - \|x^j(t) - x^j\|\right) \\
&= \|a - x^l\| - \|a - x^j(t)\| + \|x^l(t) - x^l\| + \|x^j(t) - x^j\| \\
&< \|a - x^l\| - \|a - x^j(t)\| + r(t) \\
&\leq -r(t) + r(t) = 0,
\end{aligned}$$

where the second inequality holds since $x^l \in B\left(x^l(t), \frac{r(t)}{2}\right)$ and $x^j \in B\left(x^j(t), \frac{r(t)}{2}\right)$, and the third inequality follows from (3.13), and we get that $C^l(t) \subseteq \underline{C^l}(x)$. By definition of $\underline{C^l}(x)$ we have that for any $l \neq j$, $\underline{C^l}(x) \cap \underline{C^j}(x) = \emptyset$, and for all $1 \leq l \leq k$, $\underline{C^l}(x) \subseteq \mathcal{A}$. Now, since $C(t)$ is a partition of $\mathcal{A}$, then $C^l(t) = \underline{C^l}(x)$ for all $1 \leq l \leq k$. □

Proposition 3.4 (KMEANS converges to local minimum). *Let $(C(t), x(t))$ be the clusters and centers KMEANS returns, then $x(t)$ is local minimum of F in $U = B\left(x^1(t), \frac{r(t)}{2}\right) \times B\left(x^2(t), \frac{r(t)}{2}\right) \times \cdots \times B\left(x^l(t), \frac{r(t)}{2}\right) \subset \mathbb{R}^{nk}$.*

Proof. The minimum of F in U is

$$\min_{x \in U} F(x) = \min_{x \in U} \sum_{l=1}^{k} \sum_{a \in \underline{C^l}(x)} \|a - x^l\|^2 = \min_{x \in U} \sum_{l=1}^{k} \sum_{a \in C^l(t)} \|a - x^l\|^2,$$

where the last equality follows from Proposition 3.3.
The function $x \mapsto \sum_{l=1}^{k} \sum_{a \in C^l(t)} \|a - x^l\|^2$ is strictly convex, separable in x^l for all $1 \leq l \leq k$, and reaches its minimum at $\left(x^l\right)^* = \frac{1}{|C^l(t)|} \sum_{a \in C^l(t)} a = mean(C^l(t)) = x^l(t)$, and the result follows. □

3.3 Similarity to KMEANS

Move to the one!

The famous KMEANS algorithm has close proximity to KPALM algorithm. KMEANS alternates between cluster assignments and center updates as well. In detail, we can write its steps in the following manner

(1) Initialization: Set $t = 0$, and pick random centers $y(0) \in \mathbb{R}^{nk}$.

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

(2.1) Cluster Assignment: For $i = 1, 2, \dots, m$ compute

$$v^i(t+1) = \arg\min_{v^i \in \Delta} \left\{ \langle v^i, d^i(y(t)) \rangle \right\}. \tag{3.8}$$

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

$$y^l(t+1) = \frac{\sum_{i=1}^m v_l^i(t+1) a^i}{\sum_{i=1}^m v_l^i(t+1)}. \tag{3.9}$$

The KMEANS algorithm obviously resemble KPALM algorithm. Denote $\overline{\alpha}(t) = \max_{1 \le i \le m} \alpha_i(t)$. Assuming same starting point $x(0) = y(0)$ and by taking $\overline{\alpha}(t) \to 0$, we have

$$v(t) = \lim_{\overline{\alpha}(t) \to 0} w(t), \quad y(t) = \lim_{\overline{\alpha}(t) \to 0} x(t),$$

meaning, both algorithms converge to the same result.

3.4 KMEANS Convergence Proof

We start with rewriting the KMEANS algorithms, in its most familiar form

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

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

(2.1) Cluster Assignment: For $i = 1, 2, \dots, m$ compute

$$C^l(t+1) = \left\{ a \in \mathcal{A} \mid \|a - x^l(t)\| \le \|a - x^j(t)\|, \quad \forall 1 \le l \le k \right\}. \tag{3.10}$$

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

$$x^l(t+1) = mean(C^l(t)) := \frac{1}{|C^l(t)|} \sum_{a \in C^l(t)} a. \tag{3.11}$$

(2.3) Stopping criteria: Halt if

$$\forall 1 \le l \le k \quad C^l(t+1) = C^l(t) \tag{3.12}$$

As in KPALM, KMEANS needs Assumption 1 for step (3.11) to be well defined. In order to prove the convergence of KMEANS to local minimum, we will need to following assumption.

$$= \left[\sum_{l=1}^{k}(4M)^2\|x^l(t+1) - x^l(t)\|^2\right]^{\frac{1}{2}} = 4M\|x(t+1) - x(t)\|,$$

this proves the desired result. $\square$

Proposition 3.2 (Subgradient lower bound for the iterates gap property). *Let* $\{z(t)\}_{t\in\mathbb{N}}$ ~~$= \{(w(t), x(t))\}_{t\in\mathbb{N}}$~~ *be the sequence generated by KPALM.* ~~*then*~~ Then, *there exists* $\rho_2 > 0$ *and* $\gamma(t+1) \in \partial\Psi(z(t+1))$ *such that*

space

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

Proof. By the definition of Ψ (see (2.3)) we get

$$\partial\Psi = \nabla H + \partial G = \left((\nabla_{w^i}H_i + \partial_{w^i}\delta_\Delta)_{i=1,2,\ldots,m}, \nabla_x H\right).$$

Evaluating the last relation at $z(t+1)$ yields

$$\begin{aligned}
&\partial\Psi(z(t+1)) = \\
&\quad = \left((\nabla_{w^i}H_i(w(t+1), x(t+1)) + \partial_{w^i}\delta_\Delta(w^i(t+1)))_{i=1,\ldots,m}, \nabla_x H(w(t+1), x(t+1))\right) \\
&\quad = \left((d^i(x(t+1)) + \partial_{w^i}\delta_\Delta(w^i(t+1)))_{i=1,\ldots,m}, \nabla_x H(w(t+1), x(t+1))\right) \\
&\quad = \left((d^i(x(t+1)) + \partial_{w^i}\delta_\Delta(w^i(t+1)))_{i=1,\ldots,m}, \mathbf{0}\right),
\end{aligned}$$

where the last equality follows from (3.2), that is, the optimality condition of $x(t+1)$. ~~Taking the norm of the last equality yields~~

$$\cancel{\|\partial\Psi(z(t+1))\| \leq \sum_{i=1}^{m}\|d^i(x(t+1)) + \partial_{w^i}\delta_\Delta(w^i(t+1))\|.} \qquad \cancel{(3.6)}$$

The optimality condition of $w^i(t+1)$ ~~that is~~ Which derived from (3.1), yields that for all $i = 1, 2, \ldots, m$ there exists $u^i(t+1) \in \partial\delta_\Delta(w^i(t+1))$ such that

$$d^i(x(t)) + \alpha_i(t)\left(w^i(t+1) - w^i(t)\right) + u^i(t+1) = \mathbf{0}. \qquad (3.7)$$

~~therefore~~ By using (3.7) we obtain

Setting $\gamma(t+1) := \left((d^i(x(t+1)) + u^i(t+1))_{i=1,2,\ldots,m}, \mathbf{0}\right) \in \partial\Psi(z(t+1))$, ~~and plugging (3.7) into (3.6) we have~~

Hence

$$\begin{aligned}
\|\gamma(t+1)\| &\leq \sum_{i=1}^{m}\|d^i(x(t+1)) - d^i(x(t)) - \alpha_i(t)\left(w^i(t+1) - w^i(t)\right)\| \\
&\leq \sum_{i=1}^{m}\|d^i(x(t+1)) - d^i(x(t))\| + \sum_{i=1}^{m}\alpha_i(t)\|w^i(t+1) - w^i(t)\| \\
&\leq \sum_{i=1}^{m}4M\|x(t+1) - x(t)\| + m\overline{\alpha}\cancel{(t)}\|z(t+1) - z(t)\| \\
&\leq m\left(4M + \overline{\alpha}\cancel{(t)}\right)\|z(t+1) - z(t)\|,
\end{aligned}$$

where the third inequality follows from Lemma 3.1.1, and $\overline{\alpha}\cancel{(t)} = \max_{1\leq i\leq m} \alpha_i$ ~~$\alpha_i(t)$~~. Define $\rho_2 = \cancel{\max_{t\in\mathbb{N}}}\, m\left(4M + \overline{\alpha}\cancel{(t)}\right)$, ~~due to Remark 1(iii) it follows that ρ_2 is bounded from above,~~ and the result follows. $\square$

$$\gamma(t+1) = \left(\left(d^i(x(t+1)) - d^i(x(t)) - \alpha_i(t)\left(w^i(t+1) - w^i(t)\right)\right)_{i=1,2,\ldots,m}, 0\right),$$

From Assumption ~~1~~ (ii) we have that $\beta(w(t)) = 2 \min_{1 \le l \le k} \left\{ \sum_{i=1}^{m} w_l^i(t) \right\} \geq \cancel{\beta}\ \beta$, and from Lemma 3.0.2 it follows that the function $x \mapsto H(w(t), x)$ is strongly convex with parameter $\beta(w(t))$, hence it follows that

$$
\begin{aligned}
H(w(t+1), x(t)) - H(w(t+1), x(t+1)) &\geq \\
&\geq \langle \nabla_x H(w(t+1), x(t+1)), x(t) - x(t+1) \rangle + \frac{\beta(w(t))}{2} \|x(t) - x(t+1)\|^2 \\
&= \frac{\beta(w(t))}{2} \|x(t+1) - x(t)\|^2,
\end{aligned}
$$

$\geq \frac{\beta}{2} \|x(t+1) - x(t)\|^2$

where the ~~last~~ equality follows from (3.2), since $\nabla_x H(w(t+1), x(t+1)) = 0$. Set $\rho_1 = \frac{1}{2} \min_{\cancel{t \in \mathbb{N}}} \{\alpha\cancel{(t)}, \cancel{\beta(w(t))}\ \beta\}$, combined with the previous inequalities, we have

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

where the last equality follows from ~~Remark 1(i). Note that due to Remark 1(iii) and Assumption 1 it follows that $\rho_1 > 0$.~~ the fact that $G(w(t)) = 0$ for all $t \in \mathbb{N}$ and therefore $H(z(t)) = \Psi(z(t))$, $t \in \mathbb{N}$. ∎

~~Next~~ Now, we aim to prove the subgradient lower bound for the iterates gap ~~property~~. The following lemma will be essential in our proof.

Lemma 3.1.1. *Let $\{z(t)\}_{t \in \mathbb{N}}$ ~~$= \{(w(t), x(t))\}_{t \in \mathbb{N}}$~~ be the sequence generated by KPALM, then*

$$
\|d^i(x(t+1) - d^i(x(t))\| \leq 4M \|x(t+1) - x(t)\|, \quad \forall i = 1, 2, \ldots, m, \ t \in \mathbb{N},
$$

where $M = \max_{1 \le i \le m} \|a^i\|$.

space

Proof. Since $d(u, v) = \|u - v\|^2$, we get that

$$
\begin{aligned}
\|d^i(x(t+1) - d^i(x(t))\| &= \left[\sum_{l=1}^{k} \left| \|x^l(t+1) - a^i\|^2 - \|x^l(t) - a^i\|^2 \right|^2 \right]^{\frac{1}{2}} \\
&= \left[\sum_{l=1}^{k} \left| \|x^l(t+1)\|^2 - 2\left\langle x^l(t+1), a^i \right\rangle + \|a^i\|^2 - \|x^l(t)\|^2 + 2\left\langle x^l(t), a^i \right\rangle - \|a^i\|^2 \right|^2 \right]^{\frac{1}{2}} \\
&\leq \left[\sum_{l=1}^{k} \left(\left| \|x^l(t+1)\|^2 - \|x^l(t)\|^2 \right| + \left| 2\left\langle x^l(t) - x^l(t+1), a^i \right\rangle \right| \right)^2 \right]^{\frac{1}{2}} \\
&\leq \left[\sum_{l=1}^{k} \left(\left| \|x^l(t+1)\| - \|x^l(t)\| \right| \cdot \left| \|x^l(t+1)\| + \|x^l(t)\| \right| + 2\|x^l(t) - x^l(t+1)\| \cdot \|a^i\| \right)^2 \right]^{\frac{1}{2}} \\
&\leq \left[\sum_{l=1}^{k} \left(\|x^l(t+1) - x^l(t)\| \cdot 2M + 2\|x^l(t+1) - x^l(t)\| M \right)^2 \right]^{\frac{1}{2}}
\end{aligned}
$$

The following assumption will be crucial for the coming analysis.

Assumption A. (i) The chosen sequences of parameters $\{\alpha_i(t)\}_{t\in\mathbb{N}}$, $i=1,2,\dots,m$, are bounded, that is, there exist $\underline{\alpha_i} > 0$ and $\overline{\alpha_i} < \infty$ for all $i=1,2,\dots,m$, such that

$$\underline{\alpha_i} \le \alpha_i(t) \le \overline{\alpha_i}, \quad \forall t \in \mathbb{N}.$$

2

(iii) The sequence $\{w(t)\}_{t\in\mathbb{N}}$ is bounded, since $w^i(t) \in \Delta$ for all $i = 1,2,\dots,m$ and $t \in \mathbb{N}$. Combined with the previous item, the result follows.

□

~~Suppose that Assumption A(ii) holds true~~

Lemma 3.0.2 (Strong convexity of $H(w,x)$ in x). *The function $x \mapsto H(w,x)$ is strongly convex with parameter* $\beta(w) := 2\min_{1\le l\le k}\left\{\sum_{i=1}^m w_l^i\right\}$, ~~*whenever $\beta(w) > 0$.*~~ Okay

Proof. Since the function $x \mapsto H(w(t),x) = \sum_{l=1}^k\sum_{i=1}^m w_l^i\|x^l - a^i\|^2$ is C^2, it is strongly convex if and only if the smallest eigenvalue of the corresponding Hessian matrix is positive. ~~Thus~~ Indeed, the Hessian is given by

$$\nabla_{x^j}\nabla_{x^l}H(w,x) = \begin{cases} 0 & \text{if } j \neq l, \quad 1 \le j,l \le k, \\ 2\sum_{i=1}^m w_l^i & \text{if } j = l, \quad 1 \le j,l \le k. \end{cases}$$

Since the Hessian is a diagonal matrix, the smallest eigenvalue is $\min_{1\le l\le k} 2 \cancel{\sum_{i=1}^m w_l^i = \beta(w)}$ $\beta(w) := 2\min \sum_{i=1}^m w_l^i$, $1\le l\le k$, and the result follows. □

Now we are ready to prove the decrease property of the KPALM algorithm.

Proposition 3.1 (Sufficient decrease property). Suppose that Assumption 1 holds true and let ~~Let~~ $\{z(t)\}_{t\in\mathbb{N}}$ ~~$= \{(w(t),x(t))\}_{t\in\mathbb{N}}$~~ *be the sequence generated by KPALM*. ~~*then*~~ Then, *there exists $\rho_1 > 0$ such that*

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

Proof. From the step (3.1) we derive the following inequality

$$\begin{aligned} H_i(w(t+1),x(t)) + \frac{\alpha_i(t)}{2}\|w^i(t+1) - w^i(t)\|^2 &= \langle w^i(t+1), d^i(x(t))\rangle + \frac{\alpha_i(t)}{2}\|w^i(t+1) - w^i(t)\|^2 \\ &\le \langle w^i(t), d^i(x(t))\rangle + \frac{\alpha_i(t)}{2}\|w^i(t) - w^i(t)\|^2 \\ &= \langle w^i(t), d^i(x(t))\rangle \\ &= H_i(w(t),x(t)). \end{aligned}$$

Hence, we obtain

$$\frac{\alpha_i(t)}{2}\|w^i(t+1) - w^i(t)\|^2 \le H_i(w(t),x(t)) - H_i(w(t+1),x(t)). \tag{3.5}$$

Denote ~~$\alpha(t) = \min_{1\le i\le m}\{\alpha_i(t)\}$~~ $\underline{\alpha} = \min_{1\le i\le m} \underline{\alpha_i}$. Summing inequality (3.5) over $i = 1,2,\dots,m$ yields

$$\begin{aligned} \frac{\alpha\cancel{(t)}}{2}\|w(t+1) - w(t)\|^2 &= \frac{\alpha\cancel{(t)}}{2}\sum_{i=1}^m\|w^i(t+1) - w^i(t)\|^2 \\ &\le \sum_{i=1}^m\frac{\alpha_i(t)}{2}\|w^i(t+1) - w^i(t)\|^2 \\ &\le \sum_{i=1}^m\left[H_i(w(t),x(t)) - \cancel{\sum_{i=1}^m} H_i(w(t+1),x(t))\right] \\ &= H(w(t),x(t)) - H(w(t+1),x(t)), \end{aligned}$$

where the first inequality follows from Assumption 1(i).

~~From Assumption 1(ii) which implies that $\beta(w) > 0$~~

mentioned above. More precisely, with respect to w we suggest to regularize the subproblem with proximal term as follows:

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

On the other hand, with respect to x we perfor exact minimization

$$x(t+1) = \arg\min_{x\in\mathbb{R}^{nk}} H(w(t+1), x).$$

~~At each step $t \in \mathbb{N}$, the KPALM algorithm alternates between cluster assignment and centers update. The explicit formulas, at step t, are given below~~ It is easy to check that all subproblems, with respect to w^i, $i=1,2,\ldots,m$, and x, can be simplified as follow:

$$w^i(t+1) = P_\Delta\left(w^i(t) - \frac{d^i(x(t))}{\alpha_i(t)}\right), \quad i = 1,2,\ldots,m, \tag{3.3}$$

$$x^l(t+1) = \frac{\sum_{i=1}^m w_l^i(t+1)a^i}{\sum_{i=1}^m w_l^i(t+1)}, \quad l = 1,2,\ldots,k, \tag{3.4}$$

where P_Δ is the orthogonal projection onto the set Δ, and

~~**Assumption 1.** *We assume that* $\inf_{t\in\mathbb{N}}\left\{\min_{1\le l\le k}\sum_{i=1}^m w_l^i(t)\right\} > 0$.~~

~~**Remark 1.** *(i) Since for all $t \in \mathbb{N}$ we have that $w(t) \in \Delta^m$ then $G(w(t)) = 0$ and therefore $\Psi(z(t)) = H(w(t), x(t))$.*~~

~~*(ii) For any choice of distance like function $d(\cdot,\cdot)$, the function $x \mapsto H(w,x)$ is separable in x^l for all $l = 1,2,\ldots,k$. Thus, regardless the choice of distance-like function $d(\cdot,\cdot)$, the centers update step can be done in parallel over all centers, that is, $x^l(t+1) = \arg\min_{x^l\in\mathbb{R}^k}\left\{\sum_{i=1}^m w_l^i(t)d(x^l,a^i)\right\}$, $l = 1,2,\ldots,k$, and in the case of the squared Euclidean norm the result is given in (3.4).*~~

~~*(iii) Note that in the cluster assignment step we can bound the choice of $\alpha_i(t)$ out of some interval $[\alpha_{min},\alpha_{max}]$, where $\alpha_{min} > 0$ and $\alpha_{max} < \infty$.*~~

We begin our analysis of the KPALM algorithm with the following boundedness property of the generated sequence.

For simplicity, from now on, we denote $z(t) := (w(t), x(t))$, $t\in\mathbb{N}$.

Lemma 3.0.1 (Boundedness of KPALM sequence). *Let $\{z(t)\}_{t\in\mathbb{N}}$ ~~$= \{(w(t), x(t))\}_{t\in\mathbb{N}}$~~ be the sequence generated by KPALM. Then, the following statements hold true.*

(i) For all $l = 1,2,\ldots,k$, the sequence $\{x^l(t)\}_{t\in\mathbb{N}}$ is contained in $Conv(\mathcal{A})$, ~~*where $Conv(\mathcal{A})$ is the convex hull of $\mathcal{A}$.*~~ the convex hull of $\mathcal{A}$, and therefore bounded by $M := \max_{1\le i\le m}\|a^i\|$.

~~*(ii) For all $l = 1,2,\ldots,k$, the sequence $\{x^l(t)\}_{t\in\mathbb{N}}$ is bounded by $M = \max_{1\le i\le m}\|a^i\|$.*~~

(iii) The sequence $\{z(t)\}_{t\in\mathbb{N}}$ is bounded in $\mathbb{R}^{km}\times\mathbb{R}^{nk}$.

Proof. (i) Set $\lambda_i = \frac{w_l^i(t)}{\sum_{j=1}^m w_l^j(t)}$, $i = 1,2,\ldots,m$, then $\lambda_i \ge 0$ and $\sum_{i=1}^m \lambda_i = 1$. From (3.4) we have

$$x^l(t) = \frac{\sum_{i=1}^m w_l^i(t)a^i}{\sum_{i=1}^m w_l^i(t)} = \sum_{i=1}^m\left(\frac{w_l^i(t)}{\sum_{j=1}^m w_l^j(t)}\right)a^i = \sum_{i=1}^m \lambda_i a^i \in Conv(\mathcal{A}).$$

Add number (num) to this formula

Hence $x^l(t)$ is in the convex hull of $\mathcal{A}$, for all $l = 1,2,\ldots,k$ and $t \in \mathbb{N}$.

~~(ii)~~ Taking the norm of $x^l(t)$ and using (num) yields that ~~yields again from (3.4) that~~

$$\cancel{\|x^l(t)\| = \left\|\sum_{i=1}^m\left(\frac{w_l^i(t)}{\sum_{j=1}^m w_l^j(t)}\right)a^i\right\| \le \sum_{i=1}^m\left(\frac{w_l^i(t)}{\sum_{j=1}^m w_l^j(t)}\right)\|a^i\|} \le \sum_{i=1}^m \lambda_i \max_{1\le i\le m}\|a^i\| = M.$$

$$\|x^l(t)\| = \left\|\sum \lambda_i a^i\right\| \le \sum \lambda_i\|a^i\|$$

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

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

We are about to develop an algorithm that is based on PALM theory to treat this problem. However, first we need to discuss the Fermat-Weber problem that bears close relation with the algorithm that we will present later, and develop some useful 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 $\mathcal{A}=\left\{a^1,a^2,\ldots,a^m\right\}\subset\mathbb{R}^n$ 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

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

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

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

This version of the 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

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

It is easy to verify the following equality

$$T_\epsilon(x) = x - \frac{1}{L_\epsilon(x)} \nabla f_\epsilon(x), \quad \forall x \in \mathbb{R}^n. \tag{4.4}$$

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 $x, y \in \mathbb{R}^n$,

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

(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

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

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

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

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

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.

□

The following proofs are based on the properties of the auxiliary function h_ϵ, and they are similar to the proofs in [BS2015], hence we will just state them here. Lemma 4.0.5 does not appear in that paper, and its proof is given 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*

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

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*

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

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

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

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

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

$y^0 \to z$

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

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

and

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

$\sum \frac{w_i(\|x - b^i\|^2 + \epsilon^2)}{(\|y - b^i\|^2 + \epsilon^2)^{1/2}}$

It is clear that $x \mapsto \widetilde{h}_\epsilon(x, y)$ is ~~still~~ also a quadratic function with associated matrix $L_\epsilon(y)\mathbf{I}$. Therefore, ~~from 4.0.1(i)~~ we can write

okay

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

On the other hand, from 4.0.1(ii) we have that

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

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

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

Dividing the last inequality by 2 leads to

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

It is clear that the optimal point of $\widetilde{f}_\epsilon$ is y^0 since $\nabla \widetilde{f}_\epsilon(y^0) = 0$, therefore using (4.7) with $x = y - \frac{1}{L_\epsilon(y)} \nabla \widetilde{f}_\epsilon(y)$ yields

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

Thus, using the definition of $\widetilde{f}_\epsilon$ and the fact 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 $y^0, y \in \mathbb{R}^n$. □

Start from here!

4.2 Algorithm to the Smoothed Clustering Problem

In the previous section we showed that Problem (2.1) ~~(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

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

mentioned in Section 3.1

However, in order to be able to use the theory ~~of PALM~~, we need the coupled function $H(w, x)$ to be smooth, which is not the case now ~~and in our case it is not~~. Therefore, for any $\epsilon > 0$, 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

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

Please use \varepsilon and not \epsilon!

and $d_\varepsilon^i(x) = \Bigl(\qquad \Bigr) \in \mathbb{R}^k,\quad i = 1, 2, \ldots, m.$

~~with $d_\epsilon^i(x) = \left((\|x^1 - a^i\|^2 + \epsilon^2)^{1/2}, (\|x^2 - a^i\|^2 + \epsilon^2)^{1/2}, \ldots, (\|x^k - a^i\|^2 + \epsilon^2)^{1/2} \right) \in \mathbb{R}^k$, for $i = 1, 2, \ldots, m$.~~ Note that $\Psi_\epsilon(z)$ is a perturbed form of $\Psi(z)$ for some small $\epsilon > 0$, and $\Psi_0(z) = \Psi(z)$.

(3) N~~ext 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 $1 \leq l \leq k$, denote $w_l = (w_l^1, w_l^2, \ldots, w_l^m) \in \mathbb{R}_+^m$ and define~~

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

~~and~~

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

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

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

~~thus we have~~

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

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.~~ certain ~~step~~ gradient step

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

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

(2.1) Cluster ~~A~~ assignment: ~~Take any~~ Choose certain $\alpha_i(t) > 0$, ~~and for each~~ $i = 1, 2, \ldots, m$ and compute

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

(2.2) Center ~~U~~ update: ~~F~~ for each $l = 1, 2, \ldots, k$ compute

$$x^l(t+1) = \cancel{T_\epsilon^{w_l(t+1)}(x^l(t))}\; x^l - \frac{1}{L_\epsilon(w(t+1), x^l(t))} \ldots \tag{4.10}$$

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

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

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

~~*and that*~~

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

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

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

~~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*

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

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

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

where $\underline{\alpha}(t) = \min\limits_{1\le i\le m} \{\alpha_i(t)\}$.

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

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

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

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

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

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

Set $\rho_1 = \frac{1}{2}\min_{t\in\mathbb{N}}\{\underline{\alpha}(t), \underline{L}(t)\}$, and note that since $x^l(t) \in Conv(\mathcal{A})$ for all $1 \le l \le k$, then

$$L_\epsilon^{w_l(t+1)}(x^l(t)) = \sum_{i=1}^{m} \frac{w_l^i(t+1)}{\left(\|x^l(t) - a^i\|^2 + \epsilon^2\right)^{1/2}} \ge \frac{\sum\limits_{i=1}^{m} w_l^i(t+1)}{\left({d_{\mathcal{A}}}^2 + \epsilon^2\right)^{1/2}},$$

where $d_{\mathcal{A}} = diam(Conv(\mathcal{A}))$, hence together with Remark 1(iii) and Assumption 1 assures that $\rho_1 > 0$. Combining (4.13) and (4.15) yields

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

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 $x, y \in \mathbb{R}^{nk}$ such that $x^l, y^l \in Conv(\mathcal{A})$ for all $1 \leq l \leq k$ the following inequality holds*

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

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

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

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

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

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

Hence,

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

as asserted. □

Lemma 4.1.2 (Upper bound of the sequence $\{\overline{L}(x(t))\}_{t\in\mathbb{N}}$). *Let $\{z(t)\}_{t\in\mathbb{N}} = \{(w(t), x(t))\}_{t\in\mathbb{N}}$ be the sequence generated by ε-KPALM, then for any $t \in \mathbb{N}$ we have*

$$\overline{L}(x(t)) = \max_{1\leq l\leq k}\left\{L_\epsilon^{w_l(t+1)}(x^l(t)) + \frac{2L_\epsilon^{w_l(t+1)}(x^l(t))L_\epsilon^{w_l(t+1)}(x^l(t+1))}{L_\epsilon^{w_l(t+1)}(x^l(t)) + L_\epsilon^{w_l(t+1)}(x^l(t+1))}\right\} \leq \frac{2m}{\epsilon}.$$

Proof. For any $w_l \in [0,1]^m$ and $x^l \in \mathbb{R}^n$ we have

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

Therefore,

$$\overline{L}(x(t)) = \max_{1 \leq l \leq k} \left\{ L_\epsilon^{w_l(t+1)}(x^l(t)) + \frac{2}{\frac{1}{L_\epsilon^{w_l(t+1)}(x^l(t))} + \frac{1}{L_\epsilon^{w_l(t+1)}(x^l(t+1))}} \right\} \leq \frac{m}{\epsilon} + \frac{2}{\frac{2\epsilon}{m}} = \frac{2m}{\epsilon},$$

this proves the desired result. □

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*

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

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 $1 \leq i \leq m$, $u^i(t+1) \in \partial\delta_\Delta(w^i(t+1))$ such that

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

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

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

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

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

From (4.10) and (4.12) we have

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

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

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

Therefore, denote $\overline{L}(x(t)) = \max_{1 \leq l \leq k} \left\{ L_\epsilon^{w_l(t+1)}(x^l(t)) + \frac{2L_\epsilon^{w_l(t+1)}(x^l(t)) L_\epsilon^{w_l(t+1)}(x^l(t+1))}{L_\epsilon^{w_l(t+1)}(x^l(t)) + L_\epsilon^{w_l(t+1)}(x^l(t+1))} \right\}$, and set $\rho_2 = m\left(\frac{d_{\mathcal{A}}}{\epsilon} + \overline{\alpha}(t)\right) + k\overline{L}(x(t))$, note that Lemma 4.1.2 together with Assumption 1 imply that ρ_2 is bounded from above, and the result follows. $\square$

The following lemma shows that the smoothed function indeed $H_\epsilon(w, x)$ approximates $H(w, x)$.

Lemma 4.2.1 (Closeness of smooth). *For any $(w, x) \in \Delta^m \times \mathbb{R}^{nk}$ and $\epsilon > 0$ the following inequalities hold true*

$$H(w, x) \leq H_\epsilon(w, x) \leq H(w, x) + m\epsilon.$$

Proof. Applying the inequality

$$(a + b)^\lambda \leq a^\lambda + b^\lambda, \quad \forall\, a, b \geq 0,\ \lambda \in (0, 1],$$

with $a = \|x^l - a^i\|^2$, $b = \epsilon^2$ and $\lambda = \frac{1}{2}$, yields

$$\left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2} \leq \|x^l - a^i\| + \epsilon, \quad \forall\, 1 \leq l \leq k,\ 1 \leq i \leq m.$$

Together with the fact that

$$\|x^l - a^i\| \leq \left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2},$$

yields the following inequality

$$\|x^l - a^i\| \leq \left(\|x^l - a^i\|^2 + \epsilon^2\right)^{1/2} \leq \|x^l - a^i\| + \epsilon,$$

for all $l = 1, 2, \ldots, k$, $i = 1, 2, \ldots, m$. Multiplying each inequality by w_l^i and summing over $l = 1, 2, \ldots, k$, $i = 1, 2, \ldots, m$ we obtain

$$H(w, x) \leq H_\epsilon(w, x) \leq H(w, x) + \sum_{i=1}^{m} \sum_{l=1}^{k} w_l^i \epsilon.$$

Since for all $i = 1, 2, \ldots, m$, $w^i \in \Delta$, the result follows. $\square$

Delete for now!

5 Clustering via ADMM Approach

Introducing some new variable into the problem leads to the following clustering problem notation

$$\min_{x\in\mathbb{R}^{nk}} \min_{w\in\mathbb{R}^{km}} \left\{ \sum_{i=1}^{m}\sum_{l=1}^{k} w_l^i d(x^l, a^i) \mid w^i \in \Delta, i = 1, 2, \ldots, m \right\}$$
$$= \min_{x\in\mathbb{R}^{nk}, w\in\mathbb{R}^{km}, z\in\mathbb{R}^{km}} \left\{ \sum_{i=1}^{m}\sum_{l=1}^{k} w_l^i z_l^i \;\middle|\; \begin{array}{lll} w^i \in \Delta, & i = 1, 2, \ldots, m, & \\ z_l^i = d(x^l, a^i), & i = 1, 2, \ldots, m, & l = 1, 2, \ldots, k \end{array} \right\}.$$

The augmented Lagrangian that is associated with this problem is

$$L_\rho(w, x, z, y) = \sum_{i=1}^{m}\sum_{l=1}^{k} w_l^i z_l^i + \sum_{i=1}^{m}\sum_{l=1}^{k} y_l^i (z_l^i - d(x^l, a^i)) + \frac{\rho}{2}\sum_{i=1}^{m}\sum_{l=1}^{k} \left(z_l^i - d(x^l, a^i)\right)^2. \tag{5.1}$$

Thus the ADMM formulas for (4.2) are as follows

$$\begin{aligned}
& w(t+1) = \arg\min_{w\in\Delta^m} L_\rho(w, x(t), z(t), y(t)), \\
\Rightarrow\quad & w^i(t+1) = \arg\min_{w^i\in\Delta} \sum_{l=1}^{k} w_l^i z_l^i(t) = \arg\min_{w^i\in\Delta} \left\langle w^i, z^i(t)\right\rangle, \quad 1 \le i \le m, \\
& x(t+1) = \arg\min_{x\in\mathbb{R}^{nk}} L_\rho(w(t+1), x, z(t), y(t)), \\
\Rightarrow\quad & x^l(t+1) = \arg\min_{x^l\in\mathbb{R}^{nk}} -\sum_{i=1}^{m} y_l^i(t) d(x^l, a^i) + \frac{\rho}{2}\sum_{i=1}^{m}\left(z_l^i - d(x^l, a^i)\right)^2, \quad 1 \le l \le k, \\
& z(t+1) = \arg\min_{z\in\mathbb{R}^{km}} L_\rho(w(t+1), x(t+1), z, y(t)), \\
\Rightarrow\quad & z^i(t+1) = \arg\min_{z^i\in\mathbb{R}^{km}} \left\langle w^i(t+1), z^i\right\rangle + \left\langle y^i(t), z^i\right\rangle + \frac{\rho}{2}\left\| z^i - \left(d(x^l(t+1), a^i)\right)_{l=1,\ldots,k}\right\|^2 \\
& \qquad = \left(d(x^l(t+1), a^i)\right)_{l=1,\ldots,k} - \frac{1}{\rho}\left(w^i(t+1) + y^i(t)\right), \quad 1 \le i \le m, \\
& y_l^i(t+1) = y_l^i(t) + \rho(z_l^i(t+1) - d(x^l(t+1), a^i)), \quad 1 \le i \le m,\ 1 \le l \le k.
\end{aligned}$$

Therefore we can record now the suggested KPALM algorithm (1)

> **KPALM**
>
> (1) Initialization: $(w(0), x(0)) \in \Delta^m \times \mathbb{R}^{nk}$.
>
> (2) General step $(t = 0, 1, 2, \ldots)$:
>
> (2.1) Cluster assignment: choose certain $\alpha_i(t) > 0$, $i = 1, 2, \ldots, m$, and compute
>
> ~~(2.2) Centers update~~
> $$w^i(t+1) = P_\Delta\left(w^i(t) - \frac{d^i(x(t))}{\alpha_i(t)}\right).$$
>
> (2.2) Centers update: for each $\ell = 1, 2, \ldots, k$ compute
> $$x^\ell(t+1) = \frac{\sum_{i=1}^{m} w_\ell^i(t+1)\, a^i}{\sum_{i=1}^{m} w_\ell^i(t+1)}.$$

~~In the previous result we proved that the generated sequence $\{w(t)\}_{t\in\mathbb{N}}$ is bounded and therefore there is a subsequence which converges to a certain limit w^*. We will need the following assumption~~ (2)

(ii) For all $t \in \mathbb{N}$, there exists $\beta > 0$ such ~~we have~~ that

$$\min_{1\le \ell \le k} \sum_{i=1}^{m} w_\ell^i(t) \geq \beta.$$

It should be noted that Assumption 1(i) is very mild since the parameters $\alpha_i(t)$, $1 \le i \le m$ and $t \in \mathbb{N}$, can be chosen aribtrarly by the user and therefore it can be controlled such that the boundedness property holds true. Assumption 1(ii) is essential since if it is not true then $w_{\ell_*}^i(t) = 0$ for all $1 \le i \le m$, which means that the center x^{ℓ_*} does not ~~appear~~ involved in the objective function.

---

(3) Now we would like to develope an algorithm which based on the methodology of PALM to solve Problem (4.8). It is easy to see that with respect to w, the objective function Ψ_ε keeps on the same structure as Ψ and therefore we apply the same step as in KPALM. More precisely, for all $i = 1, 2, \ldots, m$, we have

$$w^i(t+1) = \underset{w^i \in \Delta}{\operatorname{argmin}} \left\{ \langle w^i, d_\varepsilon^i(x(t)) \rangle + \frac{\alpha_i(t)}{2} \|w^i - w^i(t)\|^2 \right\}$$
$$= P_\Delta\left(w^i(t) - \frac{d_\varepsilon^i(x(t))}{\alpha_i(t)}\right), \qquad \forall\, t \in \mathbb{N},$$

where $\alpha_i(t)$, $i = 1, 2, \ldots, m$, is arbitrarly chosen. On the other hand, with respect to x we tackle the subproblem differently than KPALM. Here, we ~~[illegible]~~ follow exactly the idea of PALM that is

$$x^\ell(t+1) = \underset{x^\ell}{\arg\min} \left\{ \langle x^\ell - x^\ell(t), \nabla_{x^\ell} H_\varepsilon(w(t+1), x(t)) \rangle + \frac{L_\varepsilon(w(t+1), x^\ell(t))}{2} \| x^\ell - x^\ell(t) \|^2 \right\},$$

where

$$L_\varepsilon(w(t+1), x(t)) = \sum_{i=1}^{m} \frac{w_\ell^i(t+1)}{(\|x^\ell(t) - a^i\|^2 + \varepsilon^2)^{1/2}}.$$

---

(4) Similarly to the KPALM algorithm, the sequence generated by ε-KPALM is also bounded, since here we also have that

use \align

$$x^\ell(t+1) = x^\ell(t) - \frac{1}{L_\varepsilon(w(t+1), x^\ell(t))} \nabla_{x^\ell} H(w(t+1), x(t))$$

$$= x^\ell(t) - \frac{1}{L_\varepsilon(w(t+1), x^\ell(t))} \sum w_\ell^i(t+1) \cdot \frac{x^\ell(t) - a^i}{(\|x^\ell(t) - a^i\|^2 + \varepsilon^2)^{1/2}}$$

$$= \frac{1}{L_\varepsilon(w(t+1), x^\ell(t))} \sum \frac{w_\ell^i(t+1)\, a^i}{(\|x^\ell(t) - a^i\|^2 + \varepsilon^2)^{1/2}} \in \operatorname{conv}(\mathcal{A}).$$

Before we will be able to prove the two properties needed for global convergence of the sequence $\{z(t)\}_{t\in\mathbb{N}}$ generated by ε-KPALM, we will need several auxiliary results. For the simplicity of the expositions we define the following function

$$f_\varepsilon(x) = \sum_{i=1}^{m} w_i \left(\|x - b^i\|^2 + \varepsilon^2 \right)^{1/2},$$

for fixed $\overbrace{w_1, w_2, \ldots, w_m}^{\text{positive}} \in \mathbb{R}$ and $b^i \in \mathbb{R}^n$, $i = 1, 2, \ldots, m$.

Lemma. The gradient of $f_\varepsilon(\cdot)$ is Lipschitz continuous, that is

$$\| \nabla f_\varepsilon(x) - \nabla f_\varepsilon(\tilde{x}) \| \le \frac{2 L_\varepsilon(x) L_\varepsilon(\tilde{x})}{L_\varepsilon(x) + L_\varepsilon(\tilde{x})} \| x - \tilde{x} \|, \quad \forall\, x, \tilde{x} \in \mathbb{R}^n,$$

where

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

Proof. ...

We also need the following function

$$h_\varepsilon(x, y) = \ldots$$

Lemma 4.0.1 ...