The problem

$\mathcal{A} = \{a^1, \dots, a^m\} \subset \mathbb{R}^n$, $k \in (1, m)$ is a given fixed number of clusters.

Find $\{x^1, \dots, x^k\}$ the clusters centers:

$$(C) \quad \min \left\{ \sum_{i=1}^{m} \min_{1 \le l \le k} d(x^l, a^i) : x^1, \dots, x^k \in \mathbb{R}^n \right\}$$

where $d(\cdot, \cdot)$ is some given distance.

Reformulation & Notations

$a = (a^1, \dots, a^m) \in (\mathbb{R}^n)^m$; $a^i \in \mathbb{R}^n \quad i = 1, \dots, m$

$w = (w^1, \dots, w^m) \in (\mathbb{R}^k)^m$; $w^i \in \mathbb{R}^k \quad i = 1, \dots, m$

$x = (x^1, \dots, x^k) \in (\mathbb{R}^n)^k$; $x^l \in \mathbb{R}^n \quad l = 1, \dots, k$

$d_i(x) := \left(d(x^1, a^i), d(x^2, a^i), \dots, d(x^k, a^i) \right) \in \mathbb{R}^k \quad i = 1, \dots, m$

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

$\langle u, v \rangle = \sum_{l=1}^{k} u_l v_l \qquad \forall u, v \in \mathbb{R}^k$

using the fact that: $\min_{1 \le l \le k} u_l = \min \{ \langle u, v \rangle : v \in \Delta \}$

$\Rightarrow$ Smooth Reformulation of (C):

$$(C) \quad \min_{x \in (\mathbb{R}^n)^k} \sum_{i=1}^{m} \min \{ \langle w^i, d^i(x) \rangle : w^i \in \Delta \}$$

$$\Leftrightarrow \quad (C) \quad \min_{\substack{x \in (\mathbb{R}^n)^k \\ w \in (\mathbb{R}^k)^m}} \sum_{i=1}^{m} \left\{ \langle w^i, d^i(x) \rangle + \delta_\Delta(w^i) \right\}$$

Define for each $i=1,\dots,m$:

$$H_i(W,X) = \langle W^i, d^i(X) \rangle$$

$$G(W^i) = \delta_\Delta(w^i)$$

and

$$H(W,X) = \sum_{i=1}^m H_i(W,X) \quad ; \quad G(W) = \sum_{i=1}^m G(W^i)$$

Then $(C) \Leftrightarrow \min \{ G(W) + H(W,X) : W \in \mathbb{R}^{km}, \ X \in \mathbb{R}^{nk} \}$

Setting $d(u,v) := \|u-v\|^2$

PALM for (C):

0. Initialization: $t=0$, and pick $(W(0), X(0))$

1. Cluster Assignment: For $i=1,\dots,m$ solve (parallelly):

$$w^i(t+1) = \underset{w^i \in \Delta}{\operatorname{argmin}} \{ \langle w^i, d^i(x(t)) \rangle + \frac{\alpha(t)}{2} \| w^i - w^i(t) \|^2 \} =$$

$$= \Pi_\Delta \left(w^i(t) - \frac{d^i(x(t))}{\alpha(t)} \right)$$

Here $\alpha(t) > 0$ is the stepsize that is choosen to be:

$$\alpha(t) = \min_{1 \le l \le k} \sum_{i=1}^m w_l^i$$

2. Cluster Centers: Compute $x^l \in \mathbb{R}^n$ $l=1,\dots,k$ via

$$X(t+1) = \operatorname{argmin} \{ H(W(t+1), X) : X \in \mathbb{R}^{nk} \} \Rightarrow$$

$$x^l(t+1) = \frac{\sum_{i=1}^m w_l^i(t+1) \cdot a^i}{\sum_{i=1}^m w_l^i(t+1)}$$

Setting $\Psi(W,X) := H(W,X) + G(W)$ and $z := (W,X)$

You should remind that $G(w(t)) = 0$ for any $t > 0$ and thus

$\Psi(z(t)) = H(z^*(t))$ for all $t > 0$.

2

There is missing power 2 here

Sufficient decrease property

$\exists \rho_1 > 0 : \; \| z(t+1) - z(t) \| \leq \Psi(z(t)) - \Psi(z(t+1))$

proof: from step 1.:

$$\langle w^i(t+1), d^i(x(t)) \rangle + \frac{\alpha(t)}{2} \| w^i(t+1) - w^i(t) \|^2 \leq$$
$$\langle w^i(t), d^i(x(t)) \rangle + \frac{\alpha(t)}{2} \underbrace{\| w^i(t) - w^i(t) \|^2}_{0}$$

$$\Rightarrow \frac{\alpha(t)}{2} \| w^i(t+1) - w^i(t) \|^2 \leq H_i(w(t), x(t)) - H_i(w(t+1), x(t))$$

$$\Rightarrow \frac{\alpha(t)}{2} \| w(t+1) - w(t) \|^2 = \frac{\alpha(t)}{2} \sum_{i=1}^{m} \| w^i(t+1) - w^i(t) \|^2 \leq$$
$$\leq \sum_{i=1}^{m} H_i(w(t), x(t)) - \sum_{i=1}^{m} H_i(w(t+1), x(t)) = H(w(t), x(t)) - H(w(t+1), x(t))$$

why just there exists $m>0$? The function $H(w,x)$ is explicitly given so you can compute the parameter!

~~H(w,x) strongly convex~~

Since $x \mapsto H(w,x)$ is strongly convex then $\exists m > 0$:

$$H(w(t+1), x(t)) - H(w(t+1), x(t+1)) \geq$$
$$\geq \underbrace{\nabla_x H(w(t+1), x(t+1))^T}_{\vec{0}} (x(t) - x(t+1)) + \frac{m}{2} \| x(t) - x(t+1) \|^2 =$$

$x(t+1)$ is $\underset{x \in (\mathbb{R}^n)^k}{\operatorname{argmin}}$ of $\{H(w(t+1), x\}$

$$= \frac{m}{2} \| x(t) - x(t+1) \|^2$$

Taking $\cancel{\frac{\alpha(t)}{2} = \rho_1 =} \frac{m}{2}$ then: =

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

$w(t), w(t+1) \in \Delta$

This is not true, you should define $\rho_1 = \min\{\frac{\alpha(t)}{2}, \frac{m}{2}\}$.

Lemma: At step t, $x \to H(w(t),x)$ is strongly convex if and only if $\min_{1\le l\le k} \sum_{i=1}^{m} w_l^i > 0$.

proof: Since $x \to H(w(t),x)$ is C^2, it is strongly convex iff its Hesian matrix smallest e.v. is positive.

$$\nabla_{x_s} \partial_{x_l} H(w(t),x) = \begin{cases} 0 & s \neq l \\ 2\sum_{i=1}^{m} w_l^i(t) & s = l \end{cases}$$

You may write all the needed properties of H before you start your proofs and thus you will use them in the sequel.

$\Rightarrow$ $x \to H(w(t),x)$ is strongly convex iff $m := \min_{1\le l\le k} \sum_{i=1}^{m} w_l^i > 0$

Remark: if at step t $\min_{1\le l\le k} \sum_{i=1}^{m} w_l^i(t) = 0$ then we can conclude that the number of clusters can be reduced by 1. Hence w.l.o.g. we assume that at each step t $\min_{1\le l\le k} \sum_{i=1}^{m} w_l^i > 0$

You might give more details here, like that in this case ~~$w_l^i(t)=0$~~ for all i.

Subgradient lower bound for iterates property:

$\exists \rho_2 > 0$: $\|\gamma(t+1)\| \le \rho_2 \| z(t+1) - z(t) \|$

where $\gamma(t) \in \partial \Psi(z(t))$, $z := (w,x)$

proof: $\Psi = H + G$

You dont need to write z here.

$$\partial_{\cancel{z}} \Psi = \nabla_{\cancel{z}} H + \partial_{\cancel{z}} G = (\nabla_w H, \nabla_x H) + (\partial_{w^1} \delta_\Delta, \dots, \partial_{w^m} \delta_\Delta, 0, \dots, 0)$$

$$= \left((\nabla_{w^i} H_i + \partial_{w^i} \delta_\Delta)_{i=1,\dots,m}, \nabla_x H \right)$$

also here

$$\Rightarrow \partial_{\cancel{z}} \Psi(z(t+1)) = \Big(\big(\nabla_{w^i} H(w(t+1), x(t+1)) + \partial_{w^i} \delta_\Delta (w^i(t+1))\big)_{i=1,\dots,m}, \underbrace{\nabla_x H(w(t+1), x(t+1))}_{=0} \Big)$$

This is true but you should give reference for that

4

You might add here another equality saying that since $\nabla_{w^i} H_i(z(t+1)) = d^i(x(t+1))$ we get that

$$\partial \Psi(z(t+1)) = \left(d^i(x(t+1)) + \partial_{w_i} \delta_\Delta(w^i(t+1)) \right)_{i=1,\dots,m}, 0)$$

according to my previous remark $\rightarrow$ $d^i(x(t+1))$

$$\Rightarrow \|\partial\Psi(z(t+1))\| \le \sum_{i=1}^{m} \|\nabla_{w^i} \cancel{H(w(t+1), x(t+1))} + \partial_{w^i}\delta_\Delta(w^i(t+1))\|$$

(i) Now, since $w^i(t+1) = \underset{w^i\in\mathbb{R}^k}{\operatorname{argmin}} \left\{ \langle w^i, d^i(x(t))\rangle + \frac{\alpha(t)}{2}\|w^i - w^i(t)\|^2 + \delta_\Delta(w^i) \right\}$

$$\Rightarrow \exists\, u^i(t+1) \in \partial\delta_\Delta(w^i(t+1)) : \; d^i(x(t)) + \alpha(t)\left(w^i(t+1) - w^i(t)\right) + u^i(t+1) = 0$$

$$\Rightarrow \|\partial\Psi(z(t+1))\| \le \sum_{i=1}^{m} \|d^i(x(t+1)) - d^i(x(t)) - \alpha(t)\left(w^i(t+1) - w^i(t)\right)\| \le$$

$$\le \sum_{i=1}^{m} \|d^i(x(t+1)) - d^i(x(t))\| + m\alpha(t)\|z(t+1) - z(t)\|$$

You can bound it by a better constant $\sqrt{m}$ instead of m. Show it!

Hence, it is sufficient to show that $d^i(\cdot)$ is Lipchitz Continious

i.e. $\exists M \ge 0 : \|d^i(x) - d^i(y)\| \le M\|x - y\| \quad x, y \in (\mathbb{R}^n)^k$

<u>Proposition</u>: if $\|\nabla f\|$ is bounded on Ω, $M := \sup_{z\in\Omega}\|\nabla f(z)\|$

then $\|f(x) - f(y)\| \le M\|x - y\| \quad \forall x, y \in \Omega$

<u>Lemma</u>: $\|\nabla_x d^i(\cdot)\|$ is bounded on $\Omega = \{x \in (\mathbb{R}^n)^k \mid \|x^l\| \le \sum_{j=1}^{m}\|a^j\|\}$

<u>proof</u>: $\|\nabla_x d^i(x)\| = \|2(x^1 - a^i), \dots, (x^k - a^i)\| \le 2\sum_{l=1}^{k}\|x^l - a^i\| \le$

$$\le 2\left(\sum_{l=1}^{k}\|x^l\| + k\|a^i\|\right) \underset{x\in\Omega}{\le} 2\left(k\sum_{j=1}^{m}\|a^j\| + k\|a^i\|\right) \le 2(k+1)\sum_{j=1}^{m}\|a^j\|$$

<u>Lemma</u>: $x(t) \in \Omega \quad \forall t \ge 1$

<u>proof</u>: $\|x^l(t+1)\| = \left\|\sum_{j=1}^{m} \frac{w_l^j(t+1)\, a^j}{\sum_{j=1}^{m} w_l^j(t+1)}\right\| \le \sum_{j=1}^{m} \frac{w_l^j(t+1)}{\sum_{j=1}^{m} w_l^j(t+1)}\|a^j\| \le \sum_{j=1}^{m}\|a^j\|$

$\Rightarrow$ It follows that $\rho_2 = m(M + \alpha(t))$ is an appropriate constant, ie $\|\gamma(t+1)\| \le \rho_2\|z(t+1) - z(t)\|$, $\gamma(t) \in \partial\Psi(z(t))$

You can improve it according to what I have mentioned above!

You don't need result as is because anyway you have to prove this result another the assumption that the generated sequence is bounded and in this case it is clear that $\|\nabla_x d^i(x(t+1))\|$ is bounded for all $t > 0$.