1 The Clustering Problem

Let $\mathcal{A} = \{a^1, \ldots, a^m\}$ be a given set of point in the subset $S \subset \mathbb{R}^n$, 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 $\{A^1, \ldots, A^k\}$, called clusters. For each $l = 1, \cdots, k$, the cluster A_l is represented by its center x^l, and we want to determine k cluster centers $\{x_1, \cdots, x_k\}$ such that the sum of proximity measures from each point a^i to a nearest cluster center x^l is minimized. In hard clustering problem, we demand the clusters are mutually exclusive, and in soft clustering case, non-empty intersection of clusters is allowed.

2 Clustering via PALM approach

The clustering problem formulation is given by

$$\min_{x^1,\ldots,x^k \in S} F(x^1, \ldots, x^k) := \sum_{i=1}^{m} v_i \min_{1 \le l \le k} d(x^l, a^i), \tag{2.1}$$

with $d(\cdot, \cdot)$ being a distance-like function, and v_i are positive weights such that $\sum\limits_{i=1}^{m} v_i = 1$.

An equivalent smooth formulation to the clustering problem

$$\min_{x^1,\ldots,x^k \in S} \min_{w^1,\ldots,w^m \in \mathbb{R}^k} \left\{ \sum_{i=1}^{m} v_i \sum_{l=1}^{k} w_l^i d(x^l, a^i) \mid w^i \in \Delta^i, i = 1, \ldots, m \right\}, \tag{2.2}$$

where Δ^i is the unit simplex in $\mathbb{R}^k$ given by

$$\Delta^i = \left\{ w^i \in \mathbb{R}^k \mid \sum_{l=1}^{k} w_l^i = 1, w_l^i \ge 0, l = 1, \ldots, k \right\}. \tag{2.3}$$

PALM algorithms addresses nonconvex-nonsmooth problems of the form

$$minimize_{x,y} \Psi(x, y) := f(x) + g(y) + H(x, y), \tag{2.4}$$

and in the extended form for p blocks

$$minimize \left\{ \Psi(x_1, \ldots, x_p) := \sum_{i=1}^{p} f_i(x_i) + H(x_1, \ldots, x_p) : x_i \in \mathbb{R}^{n_i} \right\}, \tag{2.5}$$

where $H : \mathbb{R}^N \to \mathbb{R}$ with $N = \sum_{i=1}^{p} n_i$ is assumed to be C^1 and each $f_i, i = 1, \ldots, p$, is proper and lower-semicontinuous function.

Applying the PALM notations to the clustering problem formulation (1.2), with distance-like function $d(u,v) = \|u-v\|^2$, setting $f_l(x^l) = \delta_S(x^l)$, $l = 1,\ldots,k$, $g_i(w^i) = \delta_{\Delta^i}(w^i)$, $i = 1,\ldots,m$ and $H(x^1,\ldots,x^k,w^1,\ldots,w^m) = \sum_{i=1}^{m} v_i \sum_{l=1}^{k} w_l^i d(x^l,a^i)$.

Next, we confirm all requirements of f_l, g_i and H as listed in Assumptions 1 and 2 at (reference to PALM article). For simplicity, we introduce some notations $\mathbf{x} = (x^1,x^2,\ldots,x^k)$ and similarly $\mathbf{w} = (w^1,w^2,\ldots,w^m)$. Also $\mathbf{x}^{-l} = (x^1,\ldots,x^{l-1},x^{l+1},\ldots,x^k)$ and similarly $\mathbf{w}^{-i} = (w^1,\ldots,w^{i-1},w^{i+1},\ldots,w^m)$.

(i) Since $f_l, g_i, H \geq 0$ they all are proper. g_i and H are lower semicontinuous since Δ_i is closed and H in C^2. As for lower semicontinuity of f_l it requires S to be closed.

(ii) The partial gradient $\nabla_{x^l} H(\mathbf{x},\mathbf{w})$ is globally Lipschitz with moduli $L_{x^l}(\mathbf{x}^{-l},\mathbf{w}) = 2\sum_{i=1}^{m} v_i w_l^i \leq$ $\leq 2w_l^{max} \sum_{i=1}^{m} v_i = 2w_l^{max}$, for $l = 1,\ldots,k$, where $w_l^{max} := \max_{i=1,\cdots,m} w_l^i$.

(iii) H is linear with respect to $\mathbf{w}$ thus $\nabla_{x^l} H(\mathbf{x},\mathbf{w})$ is globally Lipschitz with moduli $L_{w^i}(\mathbf{x},\mathbf{w}^{-i}) = 0$, for $i = 1,\ldots,m$. For PALM's proximal steps remain always well-defined, we set $L_{w^i}(\mathbf{x},\mathbf{w}^{-i}) = \mu_i > 0$, for $i = 1,\ldots,m$ (see Remark 3 (iii)). Similarly, in case $L_{x^l}(\mathbf{x}^{-l},\mathbf{w})$ is too close to 0, we set $L_{x^l}(\mathbf{x}^{-l},\mathbf{w}) = \nu_l > 0$, for $l = 1,\cdots,k$.

(iv) $\inf\left\{L_{w^i}(\mathbf{x},\mathbf{w}^{-i})\right\} = \sup\left\{L_{w^i}(\mathbf{x},\mathbf{w}^{-i})\right\} = \mu_i, i = 1,\cdots,m$ and $\sup\left\{L_{x^l}(\mathbf{x}^{-l},\mathbf{w})\right\} \leq 2w_l^{max}$, $\inf\left\{L_{x^l}(\mathbf{x}^{-l},\mathbf{w})\right\} \geq \nu_l, l = 1,\cdots,k$.

(v) ∇H is Lipschitz continuous on bounded subset, since H in C^2 (see Remark 3 (iv)).

(vi) PALM requires Ψ to be KL function. H is real polynomial function, thus satisfies the KL property. Δ_i is semi-algebraic set, and we require S to be semi-algebraic set.

Next, we formulate PALM's steps for the clustering problem, and explicitly compute the proximal formulas.

PALM-Clustering

(1) Initialization: Select random vectors $x^{l,0} \in S, l = 1,\cdots,k$ and $w^{i,0} \in \Delta^i, i = 1\cdots,m$.

(2) For each $t = 0,1,\cdots$ generate a sequence $\left\{(x^{1,t},\cdots,x^{k,t},w^{1,t},\cdots,w^{m,t})\right\}_{t\in\mathbb{N}}$ as follows:

(2.1) For each $l = 1,\cdots,k$ compute:

(2.1.1) Take $\gamma_l > 1$, set $c_l^t = \gamma_l L_{x^l}(x^{1,t+1},\cdots,x^{l-1,t+1},x^{l+1,t},\cdots,x^{k,t},w^{1,t},\cdots,w^{m,t})$ and compute

$$\begin{aligned}
x^{l,t+1} &\in prox_{c_l^t}^{f_l}\left(x^{l,t} - \tfrac{1}{c_l^t}\nabla_{x^l} H(x^{1,t+1},\cdots,x^{l-1,t+1},x^{l,t},x^{l+1,t},\cdots,x^{k,t},w^{1,t},\cdots,w^{m,t})\right) \\
&= \Pi_S\left(x^{l,t} - \frac{\sum_{i=1}^{m} v_i w_l^{i,t} 2(x^{l,t}-a^i)}{\gamma_l \max\left\{\nu_l, 2\sum_{i=1}^{m} v_i w_l^{i,t}\right\}}\right) = \Pi_S\left(x^{l,t}\left(1 - \frac{\sum_{i=1}^{m} v_i w_l^{i,t}}{\gamma_l \max\left\{\frac{\nu_l}{2}, \sum_{i=1}^{m} v_i w_l^{i,t}\right\}}\right) + \frac{\sum_{i=1}^{m} v_i w_l^{i,t} a^i}{\gamma_l \max\left\{\frac{\nu_l}{2}, \sum_{i=1}^{m} v_i w_l^{i,t}\right\}}\right)
\end{aligned}$$

(2.2) For each $i = 1,\cdots,m$ compute:

(2.2.1) Take $\beta_i > 1$, set $d_i^t = \beta_i L_{w^i}(x^{1,t+1}, \cdots, x^{k,t+1}, w^{1,t+1}, \cdots, w^{i-1,t+1}, w^{i+1,t}, \cdots, w^{m,t})$ and compute

$$\begin{aligned} w^{i,t+1} \in prox_{d_i^t}^{g_i}(w^{i,t} - \tfrac{1}{d_i^t} \nabla_{w^i} H(x^{1,t+1}, \cdots, x^{k,t+1}, w^{1,t+1}, \cdots, w^{i-1,t+1}, w^{i,t}, w^{i+1,t}, \cdots, w^{m,t}) \\ = \Pi_{\Delta^i}(w^{i,t} - \tfrac{v_i}{\beta_i \mu_i}(w_1^{i,t}\|x^{1,t+1} - a^i\|^2, \cdots, w_k^{i,t}\|x^{k,t+1} - a^i\|^2)^T) \\ = \Pi_{\Delta^i}((w_l^{i,t}(1 - \tfrac{v_i\|x^{l,t+1} - a^i\|^2}{\beta_i \mu_i}))_{1 \le l \le k}) \end{aligned}$$

3 Clustering via ADMM approach

First we add new variables $z^l, l = 1, \cdots, k$, and formulate an equivalent problem to the clustering problem (see (1.2)):

$$\min_{x^1,\ldots,x^k \in \mathbb{R}^n} \min_{w^1,\ldots,w^m \in \mathbb{R}^k} \min_{z^1,\ldots,z^k \in S} \left\{ \sum_{i=1}^{m} v_i \sum_{l=1}^{k} w_l^i d(x^l, a^i) \mid w^i \in \Delta^i, i = 1, \ldots, m, x^l = z^l, l = 1, \ldots, k \right\} \tag{3.1}$$

We present the augmented Lagrangian associated with the clustering problem

$$L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}, \mathbf{w}) = H(\mathbf{x}, \mathbf{w}) + \sum_{l=1}^{k} (y^l)^T (x^l - z^l) + \frac{\rho}{2} \sum_{l=1}^{k} \|x^l - z^l\|^2 \tag{3.2}$$

ADMM update:

$$x^{l,t+1} := \frac{\rho z^{l,t} - y^{l,t} + 2 \sum_{i=1}^{m} v_i w_l^{i,t} a^i}{\rho + 2 \sum_{i=1}^{m} v_i w_l^{i,t}}$$

$$z^{l,t+1} := \Pi_S(x^{l,t+1} + \tfrac{y^{l,t}}{\rho})$$

$$y^{l,t+1} := y^{l.t} + \rho(x^{l,t+1} - z^{l,t+1})$$

$$w^{i,t+1} \in \left\{ w \in \mathbb{R}^k \mid w \in \Delta^i, \text{ such that if } l \notin Nearest(\mathbf{x}^{t+1}, a^i) \text{ then } w_l^i = 0 \right\}$$
$$\text{where } Nearest(\mathbf{x}, a^i) := \left\{ 1 \leq l \leq k \mid \|x^l - a^i\| = \min_{1 \leq j \leq k} \|x^j - a^i\| \right\}$$

$\|\gamma(t+1)\| \leq \rho_2 \|z(t+1) - z(t)\|$
∇H