Extended Abstract: Hierarchical Data Analysis

Wendy Di, Stefan Wild

June 6, 2017

1 Notations

Goal: Given a halo with N_p particles, find its MBP.

- D: dimension of the space, $D = 2$ for now
- $\mathbf{X} = [X_i] \in \mathbb{R}^{N_p \times 2}$: collection of particles
- $\mathbf{m} = [m_i] \in \mathbb{R}^{N_p}$: collection of each particle's mass, normally $m_i = 1, \forall i$
- $d(x, y)$: Euclidean distance between points x and y
- $\bar{m} \in R$: mass of super-particle as a specific collection of particles.

Algorithm 1 Naive

1: $MBP = \min_i \sum_{j \neq i} \frac{m_j}{d(X_i, X_j)}$

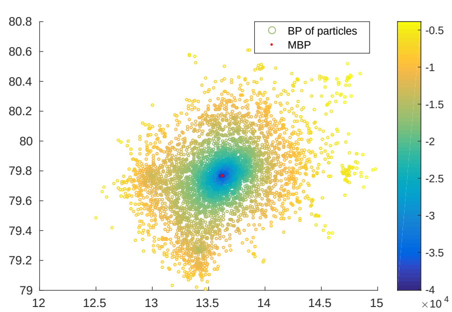

Figure 1: MBP Illustration

Algorithm 2 Mixed Particle/Super-particle Hierarchy

procedure $MBP = mixed_k means(X, m)$
  $[IDX, c] = kmeans(X, N_c)$, where N_c is the number of clusters, $IDX \in \mathbb{N}^{N_p}$ is the index function to indicate which cluster particle X_i belongs to, $\mathbf{c}_i \in \mathbb{R}^2$, $i = 1, \ldots, N_c$ is the centroid of each cluster.
  $\bar{m}_i = |\{j|IDX(j) = i\}|, \forall i = 1 : N_c$
  $SBP_i = \sum_{\{j|IDX(i) \neq j\}} \frac{\bar{m}_j}{d(X_i, c_j)} + \sum_{\{j|IDX(j)=IDX(i), i \neq j\}} \frac{m_j}{d(X_i, X_j)}, \forall i$
  $MBP = \min_i SBP_i$
end procedure

Algorithm 3 Super-particle Hierarchy

procedure $MBP = sp_k means(X, m)$
  $[IDX, c] = kmeans(X, N_c)$
  $\bar{m}_i = |\{j|IDX(j) = i\}|, \forall i = 1 : N_c$
  $MBP = \min_i \sum_{\{j|IDX(i) \neq j\}} \frac{\bar{m}_j}{d(X_i, c_j)}$
  $n_c = N_c$
  $k = 1$
  while $k \leq n_k$ **do**
    $MBP_{old} = MBP$
    $v = \{j \,|\, IDX(j) = IDX(MBP)\}$
    $[I\tilde{D}X, \tilde{c}] = kmeans(X(i), \tilde{N}_c$
    $c = \{c_1, \ldots, c_{MBP-1}\} \cup \tilde{c}_1 \cup \{c_{MBP+1}, \ldots, c_{N_c}\} \cup \{\tilde{c}_2, \ldots, \tilde{c}_{\tilde{N}_c}\}\}$
    $IDX(v) = I\tilde{D}X + kN_c - 1$
    $n_c = n_c - 1 + \tilde{N}_c$
    $\bar{m}_i = \frac{1}{|\{j|IDX(j)=i\}|}, \forall i = 1 : n_c$
    $MBP = \min_i \sum_{\{j|IDX(i) \neq j\}} \frac{\bar{m}_j}{d(X_i, c_j)}$
    if $MBP = MBP_{old}$ **then**
      Stop
    end if
    $k = k + 1$
  end while
end procedure

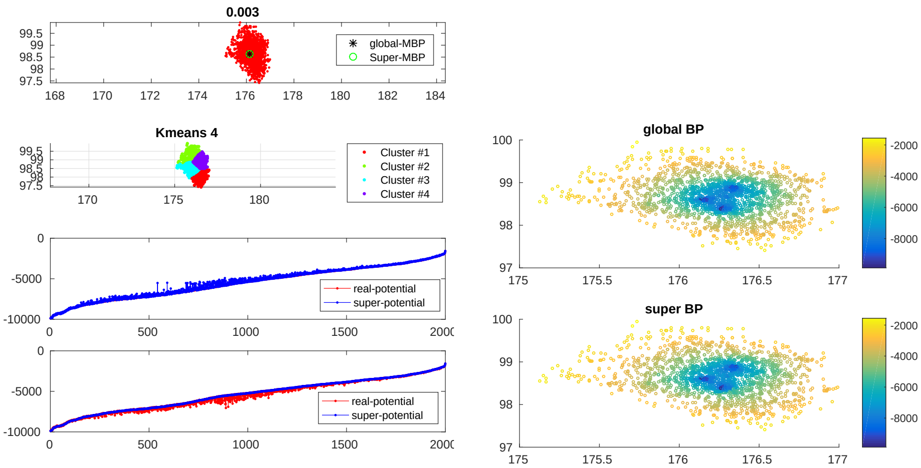

Figure 2: Result of Algorithm 2

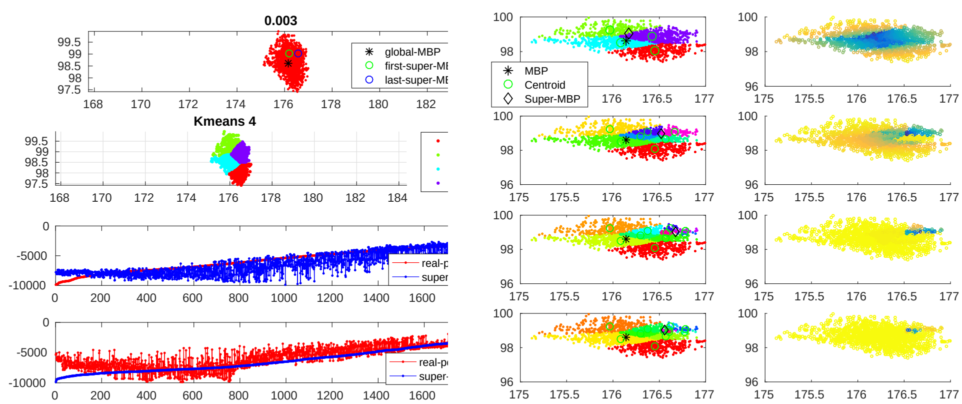

Figure 3: Result of Algorithm 3

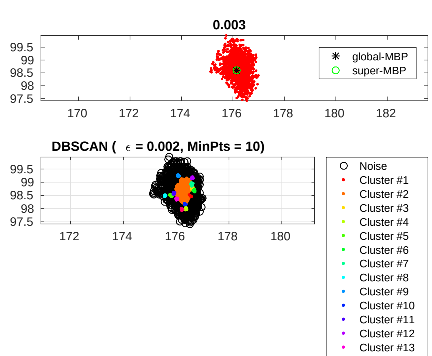

Figure 5: Result of Algorithm 4

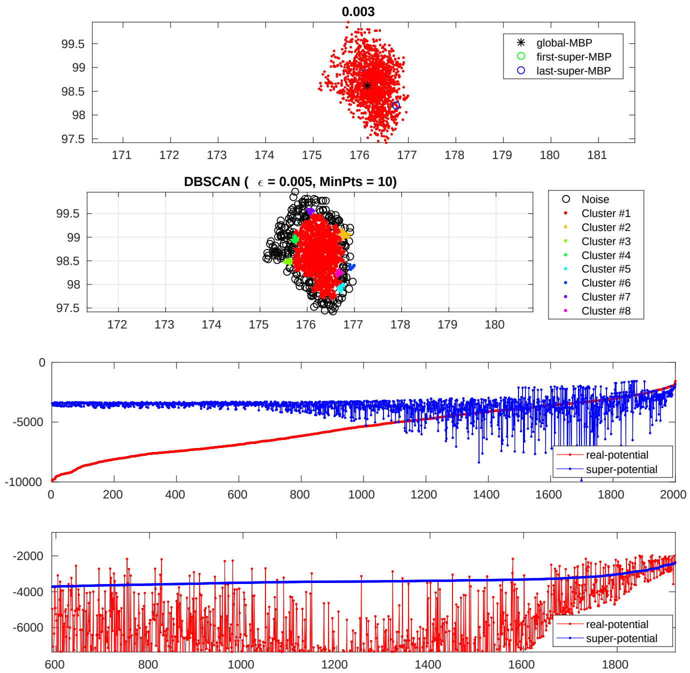

Figure 4: Result of replacing Kmeans by DBSCAN in Algorithm 3

Algorithm 4 Locate MBP from a subset of particles which forms the most dense super-particle via DBSCAN

procedure $MBP = mbp_d bscan_m ax(X, m)$

  $[IDX, c] = dbscan(X, \varepsilon)$, where ε is the linkage length provided for DBSCAN.

  $\bar{m}_i = |\{j|IDX(j) = i\}|$, $\forall i = 1 : N_c$, where N_c is the resulted number of clusters given ε

  $i_s = \max_i \bar{m}_i$

  $V = \{j \,|\, IDX(j) = i_s\}$

  $SBP_i = \sum_{\{j|IDX(i)\neq j\}} \frac{\bar{m}_j}{d(X_i, c_j)} + \sum_{\{j|IDX(j)=IDX(i), i\neq j\}} \frac{m_j}{d(X_i, X_j)}$, $\forall i \in V$

  $MBP = \min_{i\in V} SBP_i$

end procedure