

“クラスタ同定アルゴリズム”の数式解説 (“compute_cluster_atom.cpp”)

Open DEM Japan

2025年6月30日

離散要素系において、各粒子（原子） i の位置を $\mathbf{r}_i = (x_i, y_i, z_i)^T \in \mathbb{R}^3$ で表す。クラスタ同定は、隣接距離 $r_c > 0$ （コード中 cutoff）を用い、距離

$$d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|_2 \quad (1)$$

が

$$d_{ij} < r_c \quad (2)$$

を満たす粒子対 (i, j) をグラフの辺として定義し、グラフ $G = (\mathcal{V}, \mathcal{E})$ の連結成分を求める問題に帰着する。ここで $\mathcal{V} = \{i \mid i \text{ は計算対象粒子}\}$ 、 $\mathcal{E} = \{(i, j) \mid d_{ij} < r_c\}$ 。

■初期ラベル付け 各粒子には一意な整数 ID（コード中 tag）が与えられている。クラスタ ID ベクトル $\mathbf{c}^{(0)} = (c_1^{(0)}, \dots, c_N^{(0)})$ を

$$c_i^{(0)} = \begin{cases} \text{tag}_i, & i \in \text{計算対象グループ} \\ 0, & \text{その他} \end{cases} \quad (3)$$

で初期化する。

■反復ラベル伝搬 連結成分の最小 ID へ収束させるため、以下の更新を全粒子に対して同時に行う。

$$c_i^{(k+1)} = \min\left(c_i^{(k)}, \min_{j \in \mathcal{N}(i)} c_j^{(k)}\right), \quad \mathcal{N}(i) = \{j \mid d_{ij} < r_c, j \text{ 有効}\}, \quad (4)$$

ただし $\mathcal{N}(i)$ は粒子 i の隣接集合である。式(4)はラベル伝搬の一形態であり、各連結成分内で最小 ID が共有ラベルとして伝播する。

■収束判定 粒子数 N 、世界サイズ（MPI プロセス数）を P とし、反復毎に

$$\Delta^{(k)} = \sum_{i=1}^N |c_i^{(k+1)} - c_i^{(k)}| \quad (5)$$

を各プロセスで評価し、MPI Allreduce により

$$\Delta_{\text{glob}}^{(k)} = \max_{p=1, \dots, P} \Delta_p^{(k)} \quad (6)$$

を得る。 $\Delta_{\text{glob}}^{(k)} = 0$ となった時点で $\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)}$ が全並列領域で成立し、アルゴリズムは収束する。

■通信モデル 並列計算では境界粒子（ゴースト粒子）のクラスタ ID を交換する必要がある．通信バッファ \mathbf{b} への打包は

$$\mathbf{b} = (c_{j_1}, c_{j_2}, \dots, c_{j_m}), \quad (7)$$

解凍は

$$c_{j_\ell} \leftarrow b_\ell, \quad \ell = 1, \dots, m, \quad (8)$$

で表される（式 (??),(??) は `pack_comm`, `unpack_comm` に対応）．

■メモリ使用量 クラスタ ID 配列は倍精度実数型で保持され、

$$M = N_{\max} \times \text{sizeof}(\text{double}) \quad (9)$$

のメモリ（バイト）を消費する．ここで N_{\max} は局所粒子数とゴースト粒子数の最大値である．

■まとめ アルゴリズムは距離条件 (??) に基づくグラフの連結成分を反復的ラベル伝搬 (??) により求め、最短 ID に正規化されたクラスタラベル (??) が一致するとき (??) - (??) により収束を判定する．この手法は局所通信のみを用いつつ全系のクラスタ情報を求めるため、大規模並列 DEM 計算で安定して動作する．