第7章

正則化による線形ファジィクラスタリング

本章では，第5章で提案した線形ファジィクラスタリング法の次元係数の決定に，第I部の第3章，第4章で用いた正則化の概念を導入した新たな定式化による手法を提案する。1990年にDaveによって提案された適応型(adaptive)手法では，次元係数の決定を各クラスタのファジィ散布行列の固有値を用いて決定する。これによってクラスタの構造をとらえたよい結果を得ることができる。第5章で提案した手法もこの考えを応用したものである。しかしこれはファジィクラスタリングの2段階アルゴリズムの考え方に合致しない面があり，目的関数の単調現象性が失われていることを指摘する。また，その具体的な例を示す。そして次元係数の決定に正則化の概念を導入することにより，その問題を解消する。また，本手法によるいくつかの数値例によるクラスタリング結果を紹介する。

7.1　はじめに

第5章で見てきたように，ファジィクラスタリングで広く用いられているfuzzy c-means法は，距離を線形多様体との距離とすることにより，多次元空間上のデータの線形構造をとらえたクラスタリングへと応用されている。そのような手法の多くは，1990年にDaveによって提案された「適応型(Adaptive)」手法または，それに類似した手法を用いているもの

が多い。本章では，このような「適応型(Adaptive)」手法には問題点があることを指摘し，新たに正則化の概念を導入した定式化による手法の提案を行う。

これまで fuzzy c-means 法 [11, 12, 10] は，fuzzy c-varieties 法 [41] や fuzzy c-elliptotypes 法 [42] などに広く応用され，線形構造によるクラスタリング手法が提案されている。さらに，クラスタの形状により適応的に目的関数を決める adaptive fuzzy c-elliptotypes(AFC) 法 [43] が Dave により提案され，最近は2つの目的関数だけでなく，より多くの目的関数を適合的に結合させたクラスタリング手法も提案されている [29, 51]。

しかし「適応的」に固有値を用いてクラスタの形状を決定する手法では，クラスタリング過程で目的関数が変化するため，目的関数の値が増加してしまうという問題点がある。本論文では，線形多様体の次元係数の決定に正則化の概念を導入することにより，クラスタリング過程を通して目的関数の減少が保証されたクラスタリング手法を提案する。

まず7.2節において，これまでの手法の問題点について考察を行う。次に7.3節で線形構造をあらわすパラメータの計算に正則化を用いた新たな定式化によるクラスタリング法を提案し，最後に節で，数値例による簡単なクラスタリング結果について紹介する。

7.2 適応型手法の問題点

ファジィクラスタリングには，本質的に次の2つの問題が存在する。

まず目的関数の妥当性である。ファジィクラスタリングは，クラスタリングの状態に応じた評価値を与える目的関数を定め，その値が最小となる分割が最もよい分割であると仮定する。もちろん，最良の分割結果は，パラメータの与え方によって変化するが，実際に期待している結果のときに最小値を与える目的関数である必要がある。

次にそのような目的関数が得られたときに，目的関数の最小値を与える分割をいかに発見するかである。目的関数が複雑なものになれば，大域的な最適解を得ることは難しくなる。

ここで提案する新しい目的関数に対しても，最初の問題は依然として存在している。多次元空間上の線形構造をもつクラスタ分割を1つの目的関数で比較することは難しく，まだまだ十分満足できるものではない。

しかし2番目の問題は，正則化の考え方を導入することにより改善することができる。これまでの AFC 法などの適応型手法では，α_i の値はクラスタの形状に応じて決められる。これはパラメータの意味を考えると穏当な決め方である。しかし，この α_i は目的関数の最適化とは独立して決められていることに注意が必要である。実際，AFC 法などの適応型手法ではクラスタリング過程において，目的関数が増加することがある。クラスタリングのアルゴリズムは二段階法により各ステップで最適解を求めている。しかし，パラメータ α_i の値は目的関数の最適化とは独立にクラスタリング過程で変化するため，目的関数が増加することがあるのである。

図 7.1 は，2 次元上のある人工的なデータに対する AFC 法によるクラスタリング過程での目的関数の値の変化を表すグラフである。目的関数の値は 7 ステップ目のときに最小の値

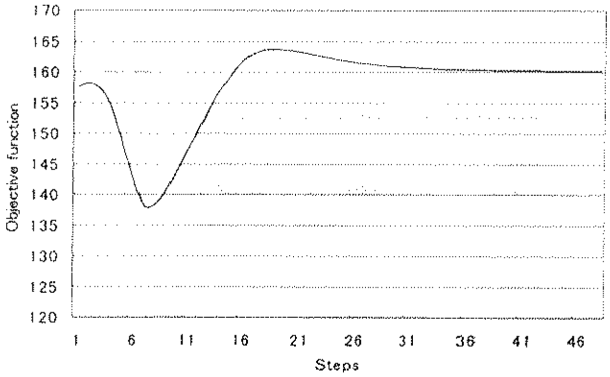

図 7.1: AFC 法によりクラスタリング過程において目的関数が増加している例

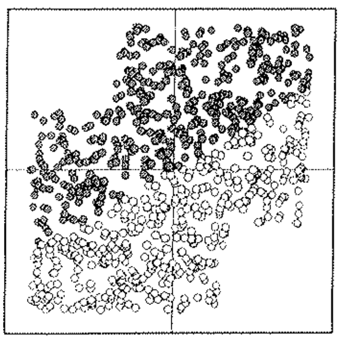

図 7.2: 7 ステップ後の目的関数の値が最小となる分割

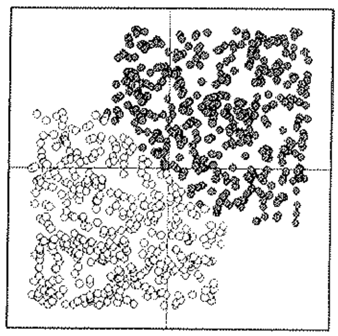

図 7.3: AFC 法による最終的なクラスタリング結果

となり，その後 α_i の変化によってその値はクラスタリングを始めたときの値をこえるまで増加している。図 7.2，図 7.3 はそれぞれ目的関数の値が最小となった 7 ステップ目とメンバシップが収束した最終結果のクラスタリング結果を示しており，各点はメンバシップ値の大きなクラスタへ色分けして示している。

このように目的関数が増加する手法では，目的関数を最小化する分割でクラスタリングが停止する保証がない。新しく提案するクラスタリングではメンバシップ値，クラスタの重心，および次元の係数に対して 3 段階の最適化をおこなう。この各段階で目的関数が単調に減少する本手法では，初期値を変えることにより，最小の目的関数を与える結果を得ることができる。

7.3 新しい目的関数

メンバシップ U やクラスタの中心 V と同様に次元係数 $B=(\beta_i^h)$ も変数であるとすると，FVD 法の目的関数はつぎのようになる。

$$J(U,V,B)=\sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{h=0}^{p-1}(u_{ik})^q\beta_i^h\frac{D_{ik}^h}{p-h} \tag{7.1}$$

この J を U,V を固定して，B について最適化を行うと，crisp k-means 法の u_{ik} 同様 β_i^h は 0,1 のクリスプ解となる。(補足 1 参照) さらに，$h\geq h'$ のとき，

$$\sum_{k=1}^{n}(u_{ik})^q\frac{D_{ik}^h}{p-h}\ \leq\ \sum_{k=1}^{n}(u_{ik})^q\frac{D_{ik}^{h'}}{p-h'} \tag{7.2}$$

となることが示される。(補足 2 参照) したがって，$B=(\beta_i^h)$ の最適解は，

$$\beta_i^{(p-1)}=1,\quad \beta_i^h=0\ \ (0\leq h\leq p-2). \tag{7.3}$$

すなわち，データ X によらず最も次元の高い線形多様体との係数だけが常に 1 となる。これでは B を変数として導入した意味がないため，D_{ik}^h を $(p-h)$ で割るかわりに，$(p-h)^2$ で割ることにする。ただし，目的関数にこれを採用するのが最良であるということではない。D_{ik}^h や $D_{ik}^h/(p-h)$ を用いてもだめなためにすぎない。あるいは，全く異なった関数

$f(D^h_{ik}, (p-h))$ を採用する事も考えることができる。この関数をどういったものにすべきかは，今後の研究としたい。

しかし，目的関数 J の最適化を考えれば，距離の一番小さい線形多様体の次元をそのクラスタの次元として採用することになり，その解はクリスプ解となる。しかし，それではクラスタの次元をうまく表現することができない。この問題は，crisp k-means 法の点のクラスタへの割り当て（補足 1 参照）と同じように考えることができる。クラスタへのクリスプな点の割り当てを，問題を正則化することによりファジィ化したように，クラスタの次元の割り当てを正則化するのである。それによって，クラスタは複数の次元をうまくトレードオフする係数を決定することができ，本手法では目的関数の単調に減少することが保証される。

具体的には，crisp k-means 法から fuzzy c-means 法への目的関数の変形と同様に，次元係数 β^h_i に対し新たなパラメータ l を導入し β^h_i を $(\beta^h_i)^l$ とする。これにより，β^h_i の値を目的関数を最小化するように柔軟に決定することができる。

以上により，最終的に

$$J_{efc} = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{h=0}^{p-1}(u_{ik})^q(\beta^h_i)^l\frac{D^h_{ik}}{(p-h)^2} \tag{7.4}$$

という目的関数を得る。各パラメータは他のパラメータを固定し，最適化することにより求めることができる。fuzzy c-means 法が二段階法によって目的関数を単調減少させていたが，これを三段階にすることによって，どの段階でも目的関数を単調減少させることができる。

7.4 新しいアルゴリズム

fuzzy c-means 法の 2 段階アルゴリズムを，$B = (\beta^h_i)$ についても同様に適用した，次のような 3 段階アルゴリズムにより目的関数を最小化する。

Step 1. $\bar{U}, \bar{V}, \bar{B}$ の初期値を適当に与える。。

Step 2. $\min_{V \in \boldsymbol{R}^{pc}} J_{efc}(\bar{U}, V, \bar{B})$ の最適解を $\bar{V}$ とする。

Step 3. $\min_{U \in M_{fcm}} J_{efc}(U, \bar{V}, \bar{B})$ の最適解を $\bar{U}$ とする。

Step 4. $\min_{B \in \boldsymbol{R}^{pc}} J_{efc}(\bar{U}, \bar{V}, B)$ の最適解を $\bar{B}$ とする。

Step 5. 最適解 $(\bar{U}, \bar{V}, \bar{B})$ が収束していれば終了，そうでなければ **Step 2** に戻る。

このとき，**Step 2** の解 $\bar{V}$ は，fuzzy c-means 法と同様に

$$\bar{\boldsymbol{v}}_i = \frac{\displaystyle\sum_{k=1}^{n} (\bar{u}_{ik})^q \boldsymbol{x}_k}{\displaystyle\sum_{k=1}^{n} (\bar{u}_{ik})^q} \tag{7.5}$$

により求められる。**Step 3** の解 $\bar{U}$ は，ラグランジュ乗数法により，

$$\bar{u}_{ik} = \left[\sum_{j=1}^{c} \left(\frac{\displaystyle\sum_{h=0}^{p-1} (\beta_i^h)^l \frac{D_{ik}^h}{(p-h)^2}}{\displaystyle\sum_{h=0}^{p-1} (\beta_j^h)^l \frac{D_{jk}^h}{(p-h)^2}} \right)^{\frac{1}{q-1}} \right]^{-1} \tag{7.6}$$

により求められる。**Step 4** の解 $\bar{B}$ も同様に，

$$\bar{\beta}_i^h = \left[\sum_{\tilde{h}=0}^{p-1} \left(\frac{\displaystyle\sum_{k=1}^{n} (u_{ik})^q \frac{D_{ik}^h}{(p-h)^2}}{\displaystyle\sum_{k=1}^{n} (u_{ik})^q \frac{D_{ik}^{\tilde{h}}}{(p-\tilde{h})^2}} \right)^{\frac{1}{l-1}} \right]^{-1} \tag{7.7}$$

と得られる。(補足4参照)

7.5 数値例

図 7.4 は各クラスタを別々に作成し，合成した人工的なデータのクラスタリング結果である。クラスタはそれぞれ，塊状・直線状・平面状に作成されており，クラスタリング結

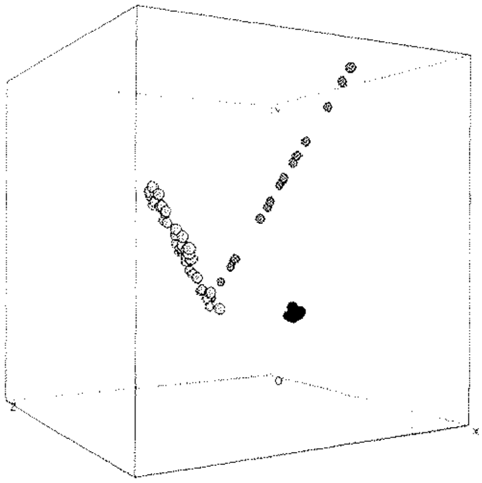

図 7.4: 3 次元上のクラスタリング例 (a)

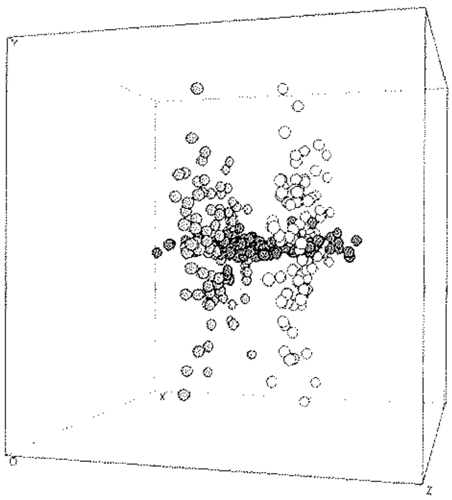

図 7.5: 3 次元上のクラスタリング例 (b)

果はその構造による分割となっている。次に図 7.5 は，2 つの平面状のクラスタが平行にあり，その 2 つのクラスタに直交するように直線状のクラスタを別々に作成して合成したデータである。このデータは完全に直線上・平面上にあるのでなく，ある程度の分散を持たせた正規乱数によって作成した。クラスタリング結果は，良好に 3 つのクラスタの構造をとらえたものとなっている。なお，どちらのクラスタリングについても，パラメータは，$c = 3$, $q = 1.8$, $l = 2.0$, $\varepsilon = 0.0001$ を用いた。

7.6 おわりに

これまでになされてきた線形クラスタリング手法における，目的関数の最小化に関する問題点を指摘し，新たなクラスタリング手法を提案した。線形多様体の次元係数の決定問題を，2 段階アルゴリズムと同様に扱い，目的関数を最小化するよう最適化により計算することにより，本手法はクラスタリング過程を通し，目的関数が単調減少することを保証するものである。

メンバシップ値のファジィ化については，他にもエントロピー正則化法 [23, 24]，2 次正則化法 [25, 26] が提案されている。本手法のメンバシップ，次元係数についてこれらの正則化手法を組み合わせることが考えられる。

7.7 補足

補足 1. 正則化について

ファジィクラスタリングがどのような意味で正則化といえるのかについては，宮本 [8] に述べられている。要約すれば，クリスプなクラスタリングを特異であるかのように見なし，パラメータを導入してクラスターをファジィ化することを正則化と呼んでいる。あるファ

ジィクラスタリングの方法が正則化と呼ばれるためには，パラメータを調節することによって正則化されたファジィなクラスタがクリスプなクラスタを近似することができる必要がある。fuzzy c-means 法では，q が正則化パラメータであり，$q \to +1$ となるとき，クラスタはクリスプなクラスタに近づくことが Bezdek[10] によって示されている。crisp k-means 法の fuzzy c-means 法への正則化について簡単に述べる。

crisp k-means 法では，個体 x_k がクラスタ i に所属するか否かを変数 u_{ik} によって，

$$u_{ik} = \begin{cases} 1, & (x_k \text{ がクラスタ } i \text{ に所属する}) \\ 0, & (x_k \text{ がクラスタ } i \text{ に所属しない}) \end{cases} \tag{7.8}$$

と決める。すべての個体は，ただ1つのクラスタに所属するので，$\sum_{i=1}^{c} u_{ik} = 1$ が成り立つ。

そこで，crisp k-means 法は，行列 $U = (u_{ik})$ とクラスタの中心 $V = (v_1, v_2, \cdots, v_c)$ を適切に選ぶことにより

$$J_{km}(U, V) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} \, \|v_i - x_k\|^2 \tag{7.9}$$

の最小化問題として定式化される。ここで U の許容集合

$$M_{km} = \left\{ (u_{ik}) \,\middle|\, \sum_{i=1}^{c} u_{ik} = 1,\ u_{ik} \in \{0, 1\},\ k \leq n \right\}$$

を用いれば，この最適化問題は次のように書ける。

$$\min_{U \in M_{km},\, V \in R^{pc}} J_{km}(U, V) \tag{7.10}$$

ただし，R^{pc} は p 次元空間 R^p の c 個の直積である。

次に fuzzy c-means 法では，crisp k-means 法をファジィ化するために2値マトリックス U をファジィに一般化する。u_{ik} は0から1の任意の値をとれるので，許容集合は，

$$M_{fcm} = \left\{ (u_{ik}) \,\middle|\, \sum_{i=1}^{c} u_{ik} = 1,\ u_{ik} \in [0, 1],\ k \leq n \right\}$$

となる。しかし，一般化はまだ十分ではない。この手続きで，J_{km}, M_{fcm} を用いたとしよう。このとき目的関数は u_{ik} に対して線形であり，最適な u_{ik} を求める問題は線形計画問題となる。そして線形計画問題の最適解は端点，すなわち u_{ik} が0か1となり，メンバシップはクリスプ解となる。

そのため，Dunn [11, 12] と Bezdek [10] はパラメータ $q(>1)$ を導入し，次のような評価規範を用いている。

$$J_{fcm}(U, V) = \sum_{i=1}^{c}\sum_{k=1}^{n} (u_{ik})^q \, \|\boldsymbol{v}_i - \boldsymbol{x}_k\|^2 \tag{7.11}$$

この J_{fcm} を用いることにより，u_{ik} はファジィ化され，$\boldsymbol{x}_k$ が $\boldsymbol{v}_i$ に一致しなければ，$0 < u_{ik} < 1$ となる。

$\bar{V}$ が与えられたとき，ステップ CM3 の解 $\bar{U}$ は

$$\bar{u}_{ik} = \left[\sum_{j=1}^{c} \left(\frac{\|\bar{\boldsymbol{v}}_i - \boldsymbol{x}_k\|^2}{\|\bar{\boldsymbol{v}}_i - \boldsymbol{x}_k\|^2} \right)^{\frac{1}{q-1}} \right]^{-1} \tag{7.12}$$

となり，$\bar{U}$ が与えられたとき，ステップ CM2 の解 $\bar{V}$ は

$$\bar{\boldsymbol{v}}_i = \frac{\displaystyle\sum_{k=1}^{n} (\bar{u}_{ik})^q \boldsymbol{x}_k}{\displaystyle\sum_{k=1}^{n} (\bar{u}_{ik})^q} \tag{7.13}$$

となる。

これを正則化の観点から見ると，メンバシップの許容集合 M_{fcm} を仮定したとき，2値解を極値的，つまり特異であると考える。正則化パラメータは q であり，目的関数はこのパラメータによって解がファジィ化されるよう変形されている。ここでいう正則化は，正則化された関数は元の関数に似ていて，正則化された問題の解は元の解に近いという意味である。

crisp k-means 法の正則化として，他にもエントロピー正則化法 [23, 24] や 2 次正則化法 [25, 26] が提案されており，正則化の方法によりクラスタリング結果が異なった性質を持つことが分類関数を用いて示されている。

補足 2. 目的関数の選択について

目的関数の D_{ik}^h を $(p-h)^2$ で割る理由について述べる。次元係数 β_i^h をクリスプに決定する問題を考える。目的関数は，各クラスタごとに独立であるから，

$$J_{\alpha}^{(i)} = \sum_{k=1}^{n}\sum_{h=0}^{p-1} (u_{ik})^q \beta_i^h \frac{D_{ik}^h}{(p-h)^\alpha} \tag{7.14}$$

とおくと，

$$J_\alpha = \sum_{i=1}^{c} J_\alpha^{(i)} \tag{7.15}$$

とあらわされ，$J_\alpha^{(i)}$ は，

$$\begin{aligned} J_\alpha^{(i)} &= \sum_{h=0}^{p-1} \beta_i^h \left(\sum_{k=1}^{n} (u_{ik})^q \frac{D_{ik}^h}{(p-h)^\alpha} \right) \\ &= \min_{h(\leq p-1)} \left(\sum_{k=1}^{n} (u_{ik})^q \frac{D_{ik}^h}{(p-h)^\alpha} \right) \end{aligned} \tag{7.16}$$

となり，これは各項の中で最小となる h の項を選択する問題となる。

(a) $\alpha = 0$ の場合

定義より，D_{ik}^h は，h に関して単調減少しているから，$h = p-1$ の項が最小値となり，

$$J_0^{(i)} = \sum_{k=1}^{n} (u_{ik})^q D_{ik}^{(p-1)} \tag{7.17}$$

$\beta_i^{(p-1)} = 1, \beta_i^h = 0 \quad (0 \leq h < p-1)$ が解となる。

(b) $\alpha = 1$ の場合

補足 3 より，ファジィ散布行列の固有値は，対応する固有ベクトル方向のファジィ分散であるので，クラスタ i のファジィ散布行列の第 j 固有値を $\lambda_{i,j}$ とすると

$$\begin{aligned} &\sum_{k=1}^{n} (u_{ik})^q \frac{D_{ik}^h}{p-h} \\ &= \frac{1}{p-h} \sum_{k=1}^{n} \sum_{j=h+1}^{p} (u_{ik})^q |\langle x_k - v_i, e_{ij} \rangle|^2 \\ &= \frac{1}{p-h} \sum_{j=h+1}^{p} \lambda_{i,j} \end{aligned} \tag{7.18}$$

また，$\lambda_{i,1} \geq \lambda_{i,2} \geq \cdots \geq \lambda_{i,p}$ より，$h \geq h'$ のとき，

$$\sum_{k=1}^{n} (u_{ik})^q \frac{D_{ik}^h}{p-h} \leq \sum_{k=1}^{n} (u_{ik})^q \frac{D_{ik}^{h'}}{p-h'} \tag{7.19}$$

となり，目的関数は $h = p-1$ のとき最小値をとり，

$$J_1^{(i)} = \lambda_{i,p} \tag{7.20}$$

$\beta_i^{(p-1)} = 1, \beta_i^h = 0 \ \ (0 \le h < p-1)$ が解となる。

よって，(a),(b) どちらの場合も，複数の次元の線形多様体を考慮する意味がない。

補足 3．ファジィ散布行列の固有値とファジィ分散

ファジィ散布行列の固有値が，その固有ベクトル方向のファジィ分散となることを示す。

いま，p 個の変数を $x_1, x_2, \cdots, x_p$ とし，個体 $\boldsymbol{x}_k$ の第 j 番目の変数の値を x_{jk} とする。このとき，クラスタ i の次のような合成変量

$$z_i = \sum_{j=1}^{p} a_{ij} x_j \tag{7.21}$$

を考える。また，変数 x_j のファジィ平均 $\bar{x}_j$，ファジィ分散 s_j，および変数 x_{j_1}, x_{j_2} のファジィ共分散 $s_{j_1 j_2}$ を，それぞれ

$$\bar{x}_j = \frac{\displaystyle\sum_{k=1}^{n} (u_{ik})^q x_{jk}}{\displaystyle\sum_{k=1}^{n} (u_{ik})^q} \tag{7.22}$$

$$s_j = \sum_{k=1}^{n} (u_{ik})^q (x_{jk} - \bar{x}_j)^2 \tag{7.23}$$

$$s_{j_1 j_2} = \sum_{k=1}^{n} (u_{ik})^q (x_{j_1 k} - \bar{x}_{j_1})(x_{j_2 k} - \bar{x}_{j_2}) \tag{7.24}$$

とする。このとき，合成変量 z_i のファジィ平均，ファジィ分散はそれぞれ，

$$\begin{aligned} \bar{z}_i &= \frac{\displaystyle\sum_{k=1}^{n} (u_{ik})^q \left[\sum_{j=1}^{p} a_{ij} x_{jk} \right]}{\displaystyle\sum_{k=1}^{n} (u_{ik})^q} \\ &= \sum_{j=1}^{p} \frac{a_{ij} \displaystyle\sum_{k=1}^{n} (u_{ik})^q x_{jk}}{\displaystyle\sum_{k=1}^{n} (u_{ik})^q} \\ &= \sum_{j=1}^{p} a_{ij} \bar{x}_j \end{aligned} \tag{7.25}$$

$$
\begin{aligned}
V(z_i) &= \sum_{k=1}^{n}(u_{ik})^q \left[\left(\sum_{j=1}^{p} a_{ij}x_{jk}\right) - \bar{z}_i\right]^2 \\
&= \sum_{k=1}^{n}(u_{ik})^q \left[\left(\sum_{j=1}^{p} a_{ij}x_{jk}\right) - \left(\sum_{j=1}^{p} a_{ij}\bar{x}_j\right)\right]^2 \\
&= \sum_{k=1}^{n}(u_{ik})^q \left[\sum_{j=1}^{p} a_{ij}(x_{jk} - \bar{x}_j)\right]^2 \\
&= \sum_{j_1=1}^{p}\sum_{j_2=1}^{p} s_{j_1 j_2} a_{ij_1} a_{ij_2} \qquad (7.26)
\end{aligned}
$$

となる。このとき，このファジィ分散を $\sum_{j=1}^{p} a_{ij}^2 = 1$ の制約のもとで最大化する解を求めるために

$$
F(a_{i1}, a_{i2}, \cdots, a_{ip}) = \sum_{j_1=1}^{p}\sum_{j_2=1}^{p} s_{j_1 j_2} a_{ij_1} a_{ij_2} - \lambda\left(\sum_{j_1=1}^{p} a_{ij_1}^2 - 1\right) \qquad (7.27)
$$

という Lagrange 関数を考える。これを，各 $a_{ij_1}, j_1 = 1, 2, \cdots, p$ で偏微分して以下の式が得られる。

$$
\frac{1}{2}\frac{\partial F}{\partial a_{ij_1}} = \sum_{j_2=1}^{p} s_{j_1 j_2} a_{ij_2} - \lambda a_{ij_1} = 0 \qquad (7.28)
$$

いま，クラスタ i のファジィ散布行列を S_i,

$$
\boldsymbol{a}_i = (a_{i1}, a_{i2}, \cdots, a_{ip})^\top \qquad (7.29)
$$

とおくと，これは

$$
S_i \boldsymbol{a}_i = \lambda \boldsymbol{a}_i \qquad (7.30)
$$

となり，S_i の固有ベクトルがその解となっていることがわかる。また，式 (7.28) に a_{ij_1} $(j_1 = 1, 2, \cdots, p)$ をかけて加え合わせると,

$$
\sum_{j_1=1}^{p}\sum_{j_2=1}^{p} s_{j_1 j_2} a_{ij_1} a_{ij_2} = \lambda \sum_{j_1=1}^{p} a_{ij_1}^2 = \lambda \qquad (7.31)
$$

となり，その固有値が固有ベクトル方向の合成変量のファジィ分散となる。

補足 4. 最適化による係数の決定

クラスタの中心 v_i，メンバシップ値 u_{ik}，および次元係数 β_i^h が，それぞれ他の値を定めたとき，目的関数の最適化により求められることを示す。

目的関数を

$$J_{efc} = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{h=0}^{p-1}(u_{ik})^q(\beta_i^h)^l f(h) D_{ik}^h \tag{7.32}$$

とする。ここで $f(h)$ は，距離 D_{ik}^h を線形多様体の次元数で補正するためのもので，

$$f(h) = (p-h)^{-2} \tag{7.33}$$

を本文中では用いている。このとき，v_i, u_{ik}, β_i^h に関する最適解を導出しよう。

補足 4(a) クラスタの中心 v_i に関する最適化

メンバシップ u_{ik} および，クラスタの次元係数 β_i^h が与えられたとき，目的関数を最小化するクラスタの中心 v_i を求める。目的関数は，各 v_i について独立だから，各クラスタごとに

$$J_i = \sum_{k=1}^{n}\sum_{h=0}^{p-1}(u_{ik})^q(\beta_i^h)^l f(h) D_{ik}^h \tag{7.34}$$

を最小化する v_i を求めればよい。$A_{ih} = I - \sum_{j=1}^{h} e_{ij}e_{ij}^{\mathsf{T}}$ とおくと，

$$\begin{aligned}
\frac{\partial D_{ik}^h}{\partial \boldsymbol{v}_i} &= -2\Big[(\boldsymbol{x}_k - \boldsymbol{v}_i) - \sum_{j=1}^{h} \boldsymbol{e}_{ij}^{\mathsf{T}}(\boldsymbol{x}_k - \boldsymbol{v}_i)\boldsymbol{e}_{ij}\Big] \\
&= -2A_{ih}(\boldsymbol{x}_k - \boldsymbol{v}_i)
\end{aligned} \tag{7.35}$$

したがって，

$$\begin{aligned}
\frac{\partial J_i}{\partial \boldsymbol{v}_i} &= \sum_{k=1}^{n}\sum_{h=0}^{p-1}(u_{ik})^q(\beta_i^h)^l f(h)\Big[-2A_{ih}(\boldsymbol{x}_k - \boldsymbol{v}_i)\Big] \\
&= -2\Big[\sum_{h=0}^{p-1} f(h)(\beta_i^h)^l A_{ih}\Big]\sum_{k=1}^{n}(u_{ik})^q(\boldsymbol{x}_k - \boldsymbol{v}_i)
\end{aligned}$$

$\frac{\partial J_i}{\partial \boldsymbol{v}_i} = \mathbf{0}$ の十分条件から，最終的に，式 (7.5) を得る。

補足 4(b) メンバシップ u_{ik} に関する最適化

クラスタの中心 v_i および，クラスタの次元係数 β_i^h が与えられたとき，目的関数を最小化する u_{ik} を求める。制約条件を考慮すれば，ラグランジュ関数は，

$$L_1 = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{h=0}^{p-1}(u_{ik})^q(\beta_i^h)^l f(h)D_{ik}^h + \sum_{k=1}^{n}\lambda_k(\sum_{i=1}^{c}u_{ik} - 1) \tag{7.36}$$

となる。ただし，λ_k はラグランジュ乗数である。

$$\frac{\partial L_1}{\partial u_{ik}} = q\cdot(u_{ik})^{q-1}\sum_{h=0}^{p-1}f(h)(\beta_i^h)^l D_{ik}^h + \lambda_k = 0$$

これを u_{ik} について解いて，

$$u_{ik} = \left[\frac{-\lambda_k}{q\cdot\sum_{h=0}^{p-1}f(h)(\beta_i^h)^l D_{ik}^h}\right]^{\frac{1}{q-1}} \tag{7.37}$$

制約条件 $\sum_{i=1}^{c}u_{ik} = 1$ より，

$$\lambda_k = -\left[\sum_{i=1}^{c}\left(q\cdot\sum_{h=0}^{p-1}f(h)(\beta_i^h)^l D_{ik}^h\right)^{-\frac{1}{q-1}}\right]^{-(q-1)} \tag{7.38}$$

となり最終的に，式 (7.6) を得る。

補足 4(c) 次元係数 β_i^h に関する最適化

メンバシップ u_{ik} および，クラスタの中心 v_i が与えられたとき，目的関数を最小化するクラスタの次元係数 β_i^h を求める。ここでも，メンバシップのときと同様に制約条件を考慮すれば，ラグランジュ関数は，

$$L_2 = \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{h=0}^{p-1}(u_{ik})^q(\beta_i^h)^l f(h)D_{ik}^h + \sum_{i=1}^{c}\lambda_i(\sum_{h=0}^{p-1}\beta_i^h - 1) \tag{7.39}$$

となる。ただし，λ_i はラグランジュ乗数である。

$$\frac{\partial L_2}{\partial \beta_i^h} = l\cdot(\beta_i^h)^{l-1}\sum_{k=1}^{n}f(h)(u_{ik})^q D_{ik}^h + \lambda_i = 0$$

これを β_i^h について解いて，

$$\beta_i^h = \left[\frac{-\lambda_i}{l \cdot \sum_{k=1}^{n} f(h)(u_{ik})^q D_{ik}^h} \right]^{\frac{1}{l-1}} \tag{7.40}$$

制約条件 $\sum_{h=0}^{p-1} \beta_i^h = 1$ より，

$$\lambda_i = -\left[\sum_{h=0}^{p-1} \left(l \cdot \sum_{k=1}^{n} f(h)(u_{ik})^q D_{ik}^h \right)^{-\frac{1}{l-1}} \right]^{-(l-1)} \tag{7.41}$$

となり最終的に，式 (7.7) を得る。