# 数量化IV類と計量多次元尺度構成法(計量MDS)について

(Notes on the Quantification Method IV and the Metric-MDS)

慶応義塾大学理工学部
竹内寿一郎

## １．はじめに

近年心理学の分野で多次元尺度法という分析手法が幅をきかすようになってきた。コンピュータの発達により、複雑な計算が短時間でできるようになってきたからである。フラクタルやカオス、ニューラルネットワークなども、まさにコンピュータの発達無しでは採用されなかったであろう。多次元尺度法とは、多次元空間に存在するであろう対象について、相互間の類似度・親近度 (あるいは非類似度・非親近度) に基づいて、それらの座標を決定する手法である。そこで思い起こされるのは数量化 IV 類である。数量化 IV 類も相互間の類似度情報に基づいて、多次元の空間座標を求める手法である。但し、多次元尺度法は類似度として計量値だけではなく、順序情報しかなくても解析が可能で、より広い状況下で使用することができ、それらは非計量 MDS と呼ばれている。しかしここでは比較のために計量 MDS と数量化 IV 類を使用してみて、その違いを検討することにする。

## ２．数量化 IV 類とは

$n$ 個の個体の間で何等かの親近性を表す量を $r_{ij}(i,j=1,2,\cdots,n)$ とし、個体が近い程大きい値をとるものとする。但し、必ずしも $r_{ij}=r_{ji}$ でなくともよい。これらの個体に対し位置ファクタ (スコア)$x_i$ を与え、親近性の大きいものほど近くに、小さいものほど遠くになるようにしたい。

$$\varphi=\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}(x_i-x_j)^2 \quad \longrightarrow \quad min$$

この問題を解くと最終的には固有値問題に帰着でき、最小固有値に対する固有ベクトルが求める解になる。しかしながら実際、最小固有値を求めるにあたり、ゼロ固有値とか誤差などと見間違う可能性もあるので、一般にはマイナスをつけて最大固有値問題に変える必要がある。そこでここでは $\varphi$ のマイナスを付けた最大化問題として扱うことにする。

親近性のあるものは $r_{ij}$ が大きいので $x_i-x_j$ がなるべく小さくなるようにスコアを与え、そうでないものは $r_{ij}$ が小さいので $x_i-x_j$ は適当な値になるようにスコアを与えるということで、$-\varphi$ を最大にする $x_i$ を求めるのが目的である。

ところで、

$$\begin{aligned}
-\varphi &= -\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}(x_i-x_j)^2 = -\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}(x_i^2-2x_ix_j+x_j^2)\\
&= 2\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}x_ix_j-\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}r_{ij}+\sum_{j=1}^{n}x_j^2\sum_{i=1}^{n}r_{ij}\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(r_{ij}+r_{ji})x_ix_j-\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}(r_{ij}+r_{ji})\\
&\qquad a_{ij}=r_{ij}+r_{ji} \quad \text{とおくと、}\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j-\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}a_{ij} \qquad \text{この式から } a_{ii} \text{の項は消えるので、}\\
&= \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}a_{ij}x_ix_j-\sum_{i=1}^{n}x_i^2\left\{\sum_{j=1,j\neq i}^{n}a_{ij}\right\}=\boldsymbol{x}'\mathbf{B}\boldsymbol{x}
\end{aligned}$$

$$\mathbf{B}=\begin{pmatrix} -\sum_{j=1,j\neq 1}^{n} a_{1j} & a_{12} & \cdots & a_{1n} \\ a_{21} & -\sum_{j=1,j\neq 2}^{n} a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & -\sum_{j=1,j\neq n}^{n} a_{nj} \end{pmatrix}, \qquad \boldsymbol{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$\boldsymbol{x}'\mathbf{B}\boldsymbol{x}$ を $\boldsymbol{x}'\boldsymbol{x}=const$ の下で最大にする。

$$\varphi^*=\boldsymbol{x}'\mathbf{B}\boldsymbol{x}-\lambda(\boldsymbol{x}'\boldsymbol{x}-c)$$

$$\frac{\partial\varphi^*}{\partial\boldsymbol{x}}=\mathbf{B}\boldsymbol{x}-\lambda\boldsymbol{x}=\mathbf{0}$$

の固有方程式を解き、その最大固有値に対する固有ベクトルが解となる。

多次元への拡張

2 次元の場合を述べればすぐに多次元への拡張が可能となることが分かる。

$n$ 個の対象の 2 次元座標を $(x_i,y_i),\ i=1,2,\cdots,n$ とすると、最大とすべき $-\varphi$ は次のように書ける。

$$-\varphi=-\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\{(x_i-x_j)^2+(y_i-y_j)^2\} \quad\longrightarrow\quad max$$

すなわち、

$$\begin{aligned} -\varphi &= -\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\{(x_i-x_j)^2+(y_i-y_j)^2\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(r_{ij}+r_{ji})x_ix_j-\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}(r_{ij}+r_{ji})+\sum_{i=1}^{n}\sum_{j=1}^{n}(r_{ij}+r_{ji})y_iy_j-\sum_{i=1}^{n}y_i^2\sum_{j=1}^{n}(r_{ij}+r_{ji}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j-\sum_{i=1}^{n}x_i^2\sum_{j=1}^{n}a_{ij}+\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}y_iy_j-\sum_{i=1}^{n}y_i^2\sum_{j=1}^{n}a_{ij} \\ &= \boldsymbol{x}'\mathbf{B}\boldsymbol{x}+\boldsymbol{y}'\mathbf{B}\boldsymbol{y} \end{aligned}$$

これを $\boldsymbol{x}'\boldsymbol{x}=const$、$\boldsymbol{y}'\boldsymbol{y}=const$ の条件下で、最大にする $\boldsymbol{x}$、$\boldsymbol{y}$ を求めることになる。

$$\varphi^*=\boldsymbol{x}'\mathbf{B}\boldsymbol{x}-\lambda(\boldsymbol{x}'\boldsymbol{x}-c)+\boldsymbol{y}'\mathbf{B}\boldsymbol{y}-\mu(\boldsymbol{y}'\boldsymbol{y}-d)$$

を偏微分してゼロとおくと、

$$\begin{cases} \dfrac{\partial\varphi^*}{\partial\boldsymbol{x}}=\mathbf{B}\boldsymbol{x}-\lambda\boldsymbol{x}=\mathbf{0} \\ \dfrac{\partial\varphi^*}{\partial\boldsymbol{y}}=\mathbf{B}\boldsymbol{y}-\mu\boldsymbol{y}=\mathbf{0} \end{cases}$$

両式の左からそれぞれ $\boldsymbol{y}'$、$\boldsymbol{x}'$ をかけて整理すると、

$$\begin{cases} \boldsymbol{y}'\mathbf{B}\boldsymbol{x}-\lambda\boldsymbol{y}'\boldsymbol{x}=\mathbf{0} \\ \boldsymbol{x}'\mathbf{B}\boldsymbol{y}-\mu\boldsymbol{x}'\boldsymbol{y}=\mathbf{0} \end{cases}$$

$$(\lambda-\mu)\boldsymbol{x}'\boldsymbol{y}=0$$

が得られ、固有値が異なれば $\boldsymbol{x}'\boldsymbol{y}=0$ であることが分かる。

また、$\mathbf{B}$ はランクおちしているので、少なくとも 1 つのゼロ固有値が存在する。それに対応する固有ベクトルは $\mathbf{1}=(1,1,\cdots,1)'$ であることも分かる。従ってゼロ以外の固有値に対応するベクトルは全て $\mathbf{1}$

に直交するベクトルとなる (すなわち要素の合計がゼロである)。

以上のことを考え合わせると、多次元の場合結局、

$$(\mathbf{B}-\lambda\mathbf{I})\boldsymbol{x}=\mathbf{0}$$

の固有値・固有ベクトル問題に帰着されることが分かる。解は大きい固有値に対応する固有ベクトルから順に採択すれば良いことになる。

ただし、これは固有値すべてが正である場合に言えることで、実際には固有値の順番が絶対値の大きさの順に並んでいることが多いので、固有値の順番を考えるときは注意しなければならない問題である。

計算のための注釈

$\mathbf{R}=\begin{pmatrix} & r_{ij} & \\ & & \end{pmatrix}$ とすると、$(r_{ij}+r_{ji})=\mathbf{R}+\mathbf{R}'=\mathbf{A}$

$\mathbf{B}$ の対角要素は、

$$\begin{aligned} b_{11} &= a_{11}-\sum_{j=1}^{n}a_{1j} \\ b_{22} &= a_{22}-\sum_{j=1}^{n}a_{2j} \\ &\vdots \\ b_{nn} &= a_{nn}-\sum_{j=1}^{n}a_{nj} \end{aligned}$$

であるから次の式で計算するとよい。

$$\mathbf{B}=\mathbf{A}-\mathbf{C}$$

但し、$\mathbf{C}=diag(a_{1\cdot},a_{2\cdot},\cdots,a_{n\cdot})$

$$a_{i\cdot}=\sum_{j=1}^{n}a_{ij}$$

## ３．計量 MDS とは

類似度・親近度が対象間の距離で与えられたときの多次元尺度法を計量 MDS という。$n$ 個の対象間の距離を $d_{ij}$ と書くことにし、例えばそれが $p$ 次元ユークリッド距離で与えられた場合は、

$$d_{ij}=\sqrt{\sum_{\alpha=1}^{p}(x_{i\alpha}-x_{j\alpha})^2}=||\boldsymbol{x}_i-\boldsymbol{x}_j||$$

と書ける。ここで、$\boldsymbol{x}_i=(x_{i1},x_{i2},\cdots,x_{ip}),\boldsymbol{x}_i'=(x_{j1},x_{j2},\cdots,x_{jp})$ は $i$ 番目、$j$ 番目の対象の座標、$||\cdot||$ はノルムを表す。通常ベクトルは縦で定義するが、ここではサンプルの添え字が先頭にくるので横ベクトルで定義し、これが縦に並んだ行列を $\boldsymbol{X}$ で定義する。このとき $\boldsymbol{x}_i$ の中の一点を原点にとって、任意の 2 点間の内積を定義すると、

$$(z_{ij})=(\boldsymbol{x}_i\boldsymbol{x}_j')=(\sum_{\alpha=1}^{p}x_{i\alpha}x_{j\alpha})=\boldsymbol{X}\boldsymbol{X}'$$

となり、この内積を点間の距離で置き換え、三角形の余弦定理 (ピタゴラスの定理の拡張)、

$$d_{ij}=d_{ik}^2+d_{jk}^2-2cos\theta d_{ik}d_{jk}$$

を使えば、内積 $z_{ij}$ は

$$(z_{ij}) = (cos\theta d_{ik} d_{jk}) = \left(\frac{1}{2}(d_{ik}^2 + d_{jk}^2 - d_{ij}^2)\right) = \boldsymbol{Z} = \boldsymbol{X}\boldsymbol{X}'$$

から、点間の距離が分かれば内積を計算することができる。しかし $n$ 個の座標の中一つの座標は原点 $(0, 0, \cdots, 0)$ としたから $\boldsymbol{Z}$ は実質 $(n-1) \times (n-1)$ の行列である。

そこで、$\boldsymbol{Z}$ から $\boldsymbol{X}$ を求めるには次のようにする。$\boldsymbol{Z}$ は対称行列かつ半正値定符号 (positive semi-definite) なので、それを対角化する直交行列 $\boldsymbol{T}$ が存在し、その固有値 (対角要素) $\lambda_1, \lambda_2, \cdots, \lambda_{n-1}$ は非負であることが分かっている。したがって、

$$\boldsymbol{T}'\boldsymbol{Z}\boldsymbol{T} = \boldsymbol{\Lambda} \quad \Rightarrow \quad \boldsymbol{Z} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}' = \boldsymbol{T}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{T}' = \boldsymbol{X}\boldsymbol{X}'$$

のように書くことができる。ここで、$\boldsymbol{\Lambda}$ は $\boldsymbol{Z}$ の固有値を対角要素とする $(n-1) \times (n-1)$ の行列、$\boldsymbol{\Lambda}^{\frac{1}{2}}$ は固有値の平方根を対角要素とする行列である。

すなわち $\boldsymbol{X} = \boldsymbol{T}\boldsymbol{\Lambda}^{\frac{1}{2}}$ が求める解である。

この話は $n$ 個の点が $p$ 次元空間内の $r$ 次元空間のみに存在すれば、お互いの点間の距離からそれらの座標が求められる (相対位置だけれど) ことを示している。しかし実際には距離の測定に誤差が入ることが普通で、そのときこれらの点は $r$ 次元内に治まるという保証はどこにもない。それどころか原点をどの点に選ぶかによって結果が変わってしまうことにも成りかねない。そこで結果の安定化のために、原点として $n$ 個の点の重心を採用し、その座標系に基づく内積から $\boldsymbol{X}$ を求めることにする。このときの内積は、

$$z_{ij} = \frac{1}{2}\left(\sum_{i=1}^{n}\frac{d_{ij}^2}{n} + \sum_{j=1}^{n}\frac{d_{ij}^2}{n} - \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{d_{ij}^2}{n^2} - d_{ij}^2\right)$$

となるのでこれを使うことにする。

## ３．Ｊのスクリプト

●数量化 IV 類

```
NB. 0!:0<'c:\j503\temp\jacobi.ijs'
NB.  これは固有値・固有ベクトルを求めるための関数
NB.  eigen を読み込むためのものです。
NB.  数量化 IV 類　最小固有値から求める
   su4_min=:3 : 0
b=.((+/y.)*=/~i.#y.)-y.
e0=.0{E0=.E#"1~(1e_10<|0{E=:eigen b)
if.1=*/1=*e0 do.
xx=:2{."1}.|."1 E0
else.
print 'Add an adoptive constant!!'
end.
)
NB.  数量化 IV 類　最大固有値から求める
   su4=:3 : 0
b=:y.-c=:(+/y.)*=/~i.#y.
e0=.0{E0=.E#"1~(1e_10<|0{E=:eigen b)
if. 1=*/1=*e0 do.
xx=:2{."1}.E0
```

```
else.
print 'Subtract an adoptive constant!!'
end.
)
```

NB.　使用法

```
Z=:(+|:)readtable<'c:\j\Jsympo2005\suryo4.dat'
NB. print z=:su4_min Z
NB. print z=:su4 Z-0.35
```

**【例題】**

2.5cm 離れたスクリーン上に、直径 1.5m の円形の 2 つの窓が互いに 1cm 離して用意されている。14 色から選んでその 2 つの窓に映される 2 つの色をみて、被験者は 2 つの色の似ている程度にグレードをつける。全く似ていないときは 0、全く同じ色に見えるときは 4 として、0 から 4 まで 5 段階のスコアを与える。

31 人の被験者に対して実験を行い、スコアの平均をとって 0 から 1 の間の数に変換したのが次の表である。

**色の類似度の実験（波長は実測値である）**

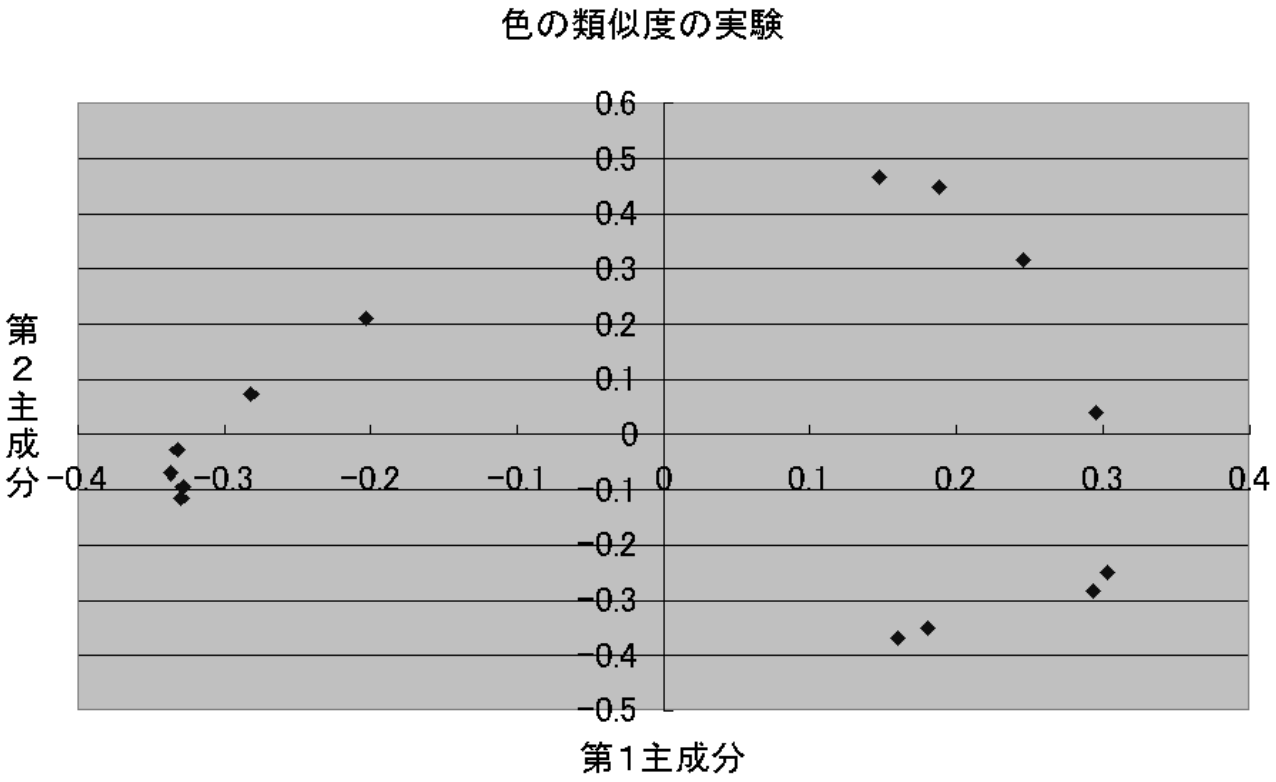
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 波長 | 434 | 445 | 465 | 472 | 490 | 504 | 537 | 555 | 584 | 600 | 610 | 628 | 651 | 674 | |
| 1 | 434 | | .86 | .42 | .42 | .18 | .06 | .07 | .04 | .02 | .07 | .09 | .12 | .13 | .16 |
| 2 | 445 | .86 | | .50 | .44 | .22 | .09 | .07 | .07 | .02 | .04 | .07 | .11 | .13 | .14 |
| 3 | 465 | .42 | .50 | | .81 | .47 | .17 | .10 | .08 | .02 | .01 | .02 | .01 | .05 | .03 |
| 4 | 472 | .42 | .44 | .81 | | .54 | .25 | .10 | .09 | .02 | .01 | .00 | .01 | .02 | .04 |
| 5 | 490 | .18 | .22 | .47 | .54 | | .61 | .31 | .26 | .07 | .02 | .02 | .01 | .01 | .00 |
| 6 | 504 | .06 | .09 | .17 | .25 | .61 | | .62 | .45 | .14 | .08 | .02 | .02 | .02 | .01 |
| 7 | 537 | .07 | .07 | .10 | .10 | .31 | .62 | | .73 | .22 | .14 | .05 | .02 | .02 | .00 |
| 8 | 555 | .04 | .07 | .08 | .09 | .26 | .45 | .73 | | .33 | .19 | .04 | .03 | .02 | .02 |
| 9 | 584 | .02 | .02 | .02 | .02 | .07 | .14 | .22 | .33 | | .58 | .37 | .27 | .20 | .23 |
| 10 | 600 | .07 | .04 | .01 | .01 | .02 | .08 | .14 | .19 | .58 | | .74 | .50 | .41 | .28 |
| 11 | 610 | .09 | .07 | .02 | .00 | .02 | .02 | .05 | .04 | .37 | .74 | | .76 | .62 | .55 |
| 12 | 628 | .12 | .11 | .01 | .01 | .01 | .02 | .02 | .03 | .27 | .50 | .76 | | .85 | .68 |
| 13 | 651 | .13 | .13 | .05 | .02 | .02 | .02 | .02 | .02 | .20 | .41 | .62 | .85 | | .76 |
| 14 | 674 | .16 | .14 | .03 | .04 | .00 | .01 | .00 | .02 | .23 | .28 | .55 | .68 | .76 | |

```
   print 15.5":su4 Z-10
```

NB. 計算された 2 軸までの座標

```
      0.15970        0.36918
      0.18021        0.35026
      0.29308        0.28246
      0.30251        0.24959
      0.29499       _0.04080
      0.24513       _0.31607
```

```
 0.18798       _0.44848
 0.14755       _0.46719
_0.20333       _0.21154
_0.28157       _0.07433
_0.33212        0.02809
_0.33655        0.06997
_0.32827        0.09411
```




●計量 MDS のスクリプト

```
NB. Metric-MDS
NB.   0!:0<'c:\j503\temp\jacobi.ijs'
   Distance=:[:+/[:*:-/"1~@|:                NB. 各点間距離を求める

NB.  Usage:   MT_MDS D2       D2:点間距離
   MT_MDS=:3 : 0
Z=.-:(+/~(+/D2)%n)-D2+(+/+/D2)%*:n=.#D2=.y. NB. 重心を原点とした内積の計算
X=.(}.E)*"1 lam=:%:0{E=:eigen Z             NB. 固有ベクトルを求める
)
   data=:4 2$1 _3 _3 4 4 1 _2 _2
   D2=:Distance data
   X=:MT_MDS D2
```

```
NB.   plot <"1 |:2{."1 ".15.5":X
```

【例題】

横山氏の例題はもともと重心が原点となっている。

```
   data
 1 _3
_3  4
 4  1
_2 _2
   +/data
0 0
   Distance data
 0 65 25 10
65  0 58 37
25 58  0 45
10 37 45  0

  11.5":MT_MDS data
   _2.82843   _1.41421    0.00000    0.00000
    4.94975    0.70711    0.00000    0.00000
   _2.12132    3.53553    0.00000    0.00000
    0.00000   _2.82843    0.00000    0.00000
```

NB. 上が結果のベクトル、下が計算途中の固有値・固有ベクトル

NB. 先頭行が固有値、２行目以降が直交行列。

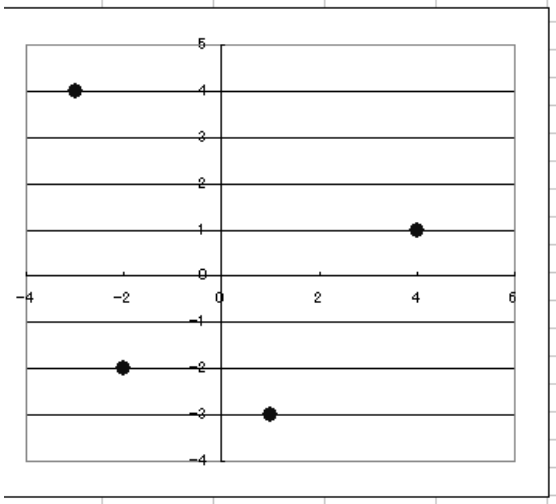
```
      11.5":E
   37.00000   23.00000    0.00000    0.00000
   _0.46499   _0.29488    0.82681   _0.11493
    0.81373    0.14744    0.53447    0.17447
   _0.34874    0.73721    0.14469    0.56033
    0.00000   _0.58977   _0.09893    0.80149

   Distance dd=:2{."1 MT_MDS data
 0 65 25 10
65  0 58 37
25 58  0 45
10 37 45  0
   +/dd
1.21645e_11 9.59166e_12
```

NB. 互いの距離は元データに等しく、求めた dd も原点が重心になっている。

**元データと計算から求めた座標**

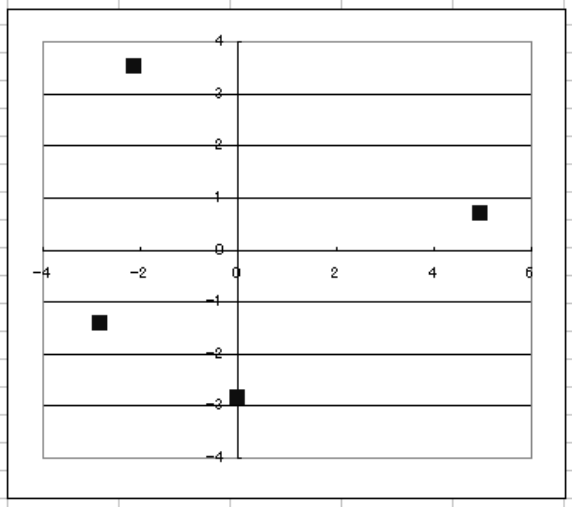
| | |
|---|---|
| 1 | −3 |
| −3 | 4 |
| 4 | 1 |
| −2 | −2 |

元データ



| | |
|---|---|
| −2.82843 | −1.41421 |
| 4.94975 | 0.707107 |
| −2.12132 | 3.53553 |
| 6.08E−12 | −2.82843 |

計算後



## ４．数量化 IV 類と計量 MDS

ここでは両者の関係を調べるための実験を行ってみる。それぞれの点間に適当なウェイトが与えられている場合、そのウェイトをある種のパーセンテージで与えるか、絶対値を加えて表すかによって、おそらく結果は大きくことなるであろう。そこで、最も安定していると思われるのは等ウェイトのときである。そこで試しに等ウェイトの場合について実験してみた。まず、２次元空間であれば３点が必要十分で、数量化 IV 類では等ウェイト、計量 MDS では等距離で最適座標を求めてみる。

```
   W=:3 3$1
   W
1 1 1
1 1 1
1 1 1
   su4 W
Subtract an adoptive constant!!
Subtract an adoptive constant!!
   su4 W-2
_0.707107 _0.408248
 0.707107 _0.408248
    1e_12  0.816497
```

最小問題として解くとすぐ解ける。

```
   su4_min W
_0.408248 _0.707107
_0.408248  0.707107
```

```
 0.816497     1e_12
```

始めのベクトルを２個とって距離を調べてみる。

```
   x1=.2{."1 su4 W-2
   x2=.2{."1 su4_min W
   Distance x1
0 2 2
2 0 2
2 2 0
   Distance x2
0 2 2
2 0 2
2 2 0
```

一方、計量 MDS では等距離行列を与えると、

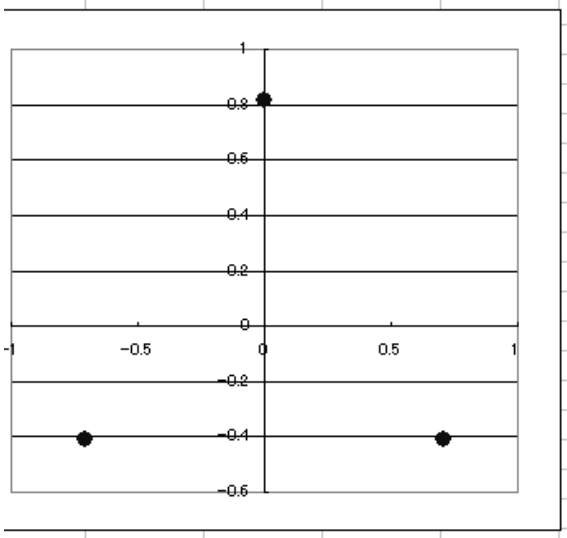
```
   X=:MT_MDS Distance x1
   15.5":X
        0.70711       _0.40825        0.00000
       _0.70711       _0.40825        0.00000
        0.00000        0.81650        0.00000
   xx1=:x1
   xx2=:2{."1 X
```

として図を描くと

## 等ウェイトから求めた座標

| | |
|---|---|
| −0.70711 | −0.40825 |
| 0.707107 | −0.40825 |
| 1.00E−12 | 0.816497 |

数量化 ×1




| | |
|---|---|
| 0.707107 | −0.40825 |
| −0.70711 | −0.40825 |
| 1.00E−12 | 0.816497 |

計量MDS