Bulletin of the Faculty of Education,
Nagoya University (Educational Psychology)
1987, Vol. 34, 235−253.

複数の変数集合の主成分と正準変量

村　上　　隆

1. 序

質問紙調査やテストを用いた多くの教育心理学的研究においては，概念的に区別される複数の変数集合から導かれる概念間の相関関係が，関心の中心になる。

例えば，高校生の性格特性，学科興味，学業成績の間に，何らかの関係があるか否かが研究される，という事態を考えてみよう。この場合，変数は，性格特性に関わる質問項目（に対する反応），学科興味に関する質問項目，各教科の成績，という３つのグループに大きく分けられる。このような変数のグループを集合と呼び，これから得られるデータを多集合データ（multiset data）と呼ぶ。

多集合データは，通常以下のような目的意識のもとに分析されることになる。

まず，研究者は，何らかの基準にもとづいて，事前に変数を適切な範囲に絞っているはずである。例えば，性格特性といっても，そのすべてを一回の調査で被い尽すことができるとは思えないから，学科興味や学業成績と意味のある関連が期待できるような，一部の特性だけが

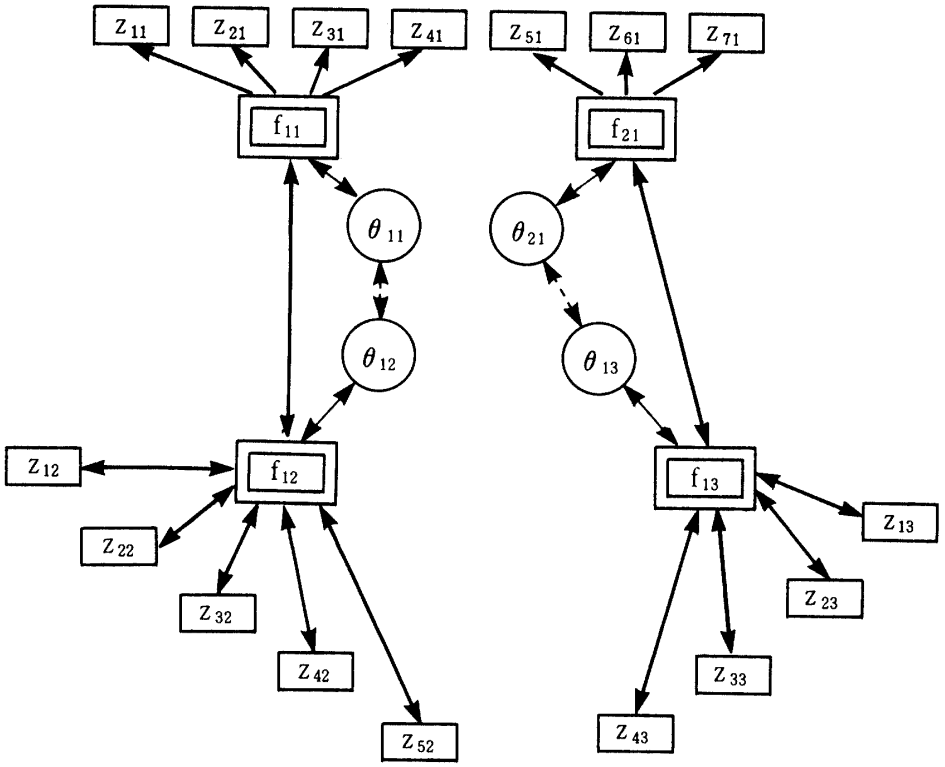

図1　個別変数（▭ 内），合成変量（▭ 内），属性（ ○ 内）の関係。
←→は，実際に計算可能な相関関係で，個別変数と合成変量の間の相関は集合内構造，合成変量間の相関は集合間構造である。←----→は，属性間に理論的に想定された相関関係であり，対応する合成変量間の相関と符号と大きさが一致することが期待される。←→は，合成変量と属性の間の相関で，合成変量の妥当性に対応する。

選ばれるであろう。

そのような限定がなされたとしても，変数を数えあげれば，相当の数にのぼることになる。しかし，研究者は，それらの変数間の関係を個別に調べあげることに関心をもっているのではなく，性格特性にせよ，学科興味にせよ，比較的少数の基本的次元が存在することを想定しているのが普通である。それらの次元は，通常，属性 (attribute)，または，構成概念 (construct) と呼ばれる，直接的には観測不可能な量である。実際の議論は，その属性に関してなされる。例えば，内向的性格は，人文科学に対する興味と関連している，といった場合，内向的性格も，人文科学に対する興味も，直接には観測不可能な属性である。

個々の変数は，それらの属性を一定の度合いで反映する指標であると見なされる。これらの変数は何らかの方法で，想定された属性に対応する比較的少数の変数，すなわち合成変量 (composite) へとまとめあげられる。その際，個々の変数はあらかじめ特定の属性に対応することが規定され，それらの変数の単純和として合成変量が定義される場合もある。あるいは，ここで何らかの因子分析的手法にもとづいて，各変数に重みをつけて加算するという方法がとられることもある。いずれにせよ，それらの合成変量間の相関係数が求められ，それが，問題となる属性間の関係と見なして解釈がなされる。

ここまでに述べてきた構想を図示すると，図1のようになる。ここで，□内に示したものは，個々の変数であり，これらは観測可能な水準にある。▣で示したのは，合成変量であって，個々の変数に対する重みの値が決まれば，観測可能（あるいは計算可能）な量となる。こちらは，いわば観測可能だが重みのとり方による不定性をもつ量ということになろう。○で示したのは観測不可能な属性である。それぞれの間をつなぐ矢印は，その間の相関係数を示す。個別変数と合成変量との間の相関を，ここでは，これを集合内構造 (intraset structure)と呼ぶことにしよう。これに対応して，合成変量間の相関を，集合間構造 (interset structure) と呼ぶ。このように考えると，データ分析の目標は，全変数間の相関行列Rを，集合内構造を示す個別項目と合成変量との相関行列である，構造行列 (structure matrix) Aと，集合間構造を示す全集合の合成変量間相関行列 $\boldsymbol{\Phi}$ とに要約することである，ということになる（図2）。

このような構想は，心理測定を用いた通常の研究において，常識的に仮定されていることがらである。それでは何故，このような複数の変数の合成という形で，属性が定義されなければならないのであろうか。それには，根本的に2つの理由があると考えられる。

まず，集合内構造，すなわち，個別変数と合成変量との相関の高さは，全体としてその合成変量の内的整合性の高さを示すが，これは通常合成変量の信頼性と同一視される。

この観点からすると，一つの属性に対応する測定値が，複数の変数の合成変量という形で定義される理由は，測定の信頼性を高めるためである。

一方，集合内構造の解釈は少々複雑である。とりあえず，ここにあらわれる相関係数の値は，想定された属性間の相関関係の反映とみなされる。この解釈が可能になるためには，図で点線で示した，合成変量と属性の間の相関が高いこと，すなわち，合成変量が当該の属性の測定値として高い妥当性をもつことが前提である。しかしもちろん，本質的に観測不可能な属性との相関は計算で

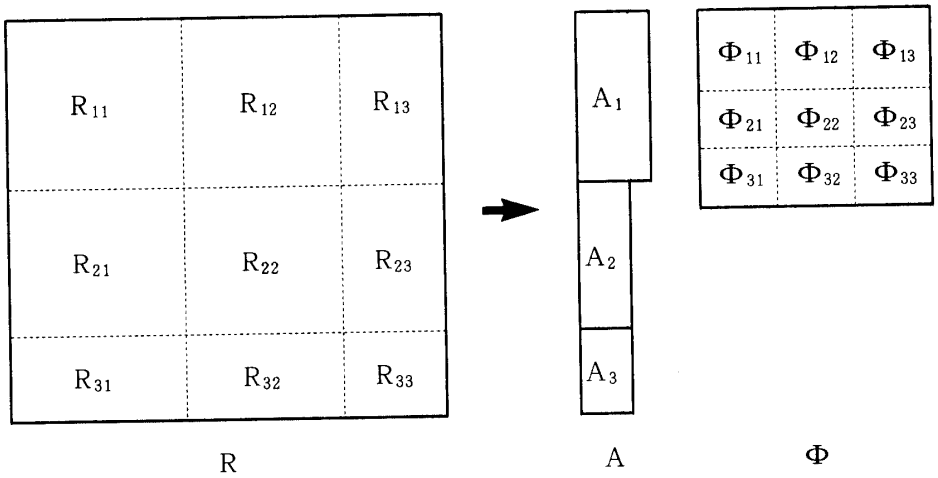

図2　すべての変数間の相関行列Rが，集合内構造を示す構造行列A（個別変数と合成変量との相関行列）と集合内構造を示す合成変量間相関行列$\boldsymbol{\Phi}$に要約される。

きないから，この前提をあらかじめ確かめることはできない。このような問題が生ずるのは，心理学的属性とその測定操作とが，もともと一対一に対応しないからである。この観点からすると，一つの属性に対して複数の変数の合成変量を対応させているのは，属性の測定操作をこのような一種の総合として定義せざるを得ないことからくると考えられる。

そこで，合成変量間に，対応する属性間に理論的（あるいは常識的）に想定される方向（符号）で，適切な大きさ（絶対値）の相関が認められることが，合成変量の妥当性の確認につながる，とも考えられる（構成概念妥当性）。

純粋にデータ分析の立場からしても，ここで2つの目的が同時に追求されなければならないことになる。1つは，合成変量の信頼性を高めるように各変数の重みを決めることであり，もう1つは，異なる集合の合成変量間相関を高めるように，各変数の重みを決めることである。このことは，Cronbach（1970）の bandwidth-fidelity dilemma に相当し，2つの目的をともに最大限に満足させることは不可能である。（このあたりの事情については，村上（1987 a）において，具体例にもとづいて，より詳細に論じた。）

村上（1986）は，多集合データの因子分析的取り扱いのためのモデルとして，階層的主成分分析 hierarchical principal component analysis を提案した。その際，集合内構造と，集合間構造のどちらを強調するかによって，極端に異なる2つの立場が生ずることを指摘した。すなわち，集合の数が2の場合に限定すれば，各集合の因子負荷の2乗和を最大化する，集合ごとの個別の主成分分析は，集合内構造を最も強調する立場であり，合成変量の集合間の相関の和を最大化する，正準相関分析は，集合間構造を強調する立場とみなされる。

階層的主成分分析は，この両極端の立場の中間に，分析の目的に適合した解を求める方法として定式化された。また，村上（1987 a）では，それぞれの集合ごとに合成変量（1次合成変量）を定義し，さらにそれらを全集合を通じて合成した2次合成変量を定義することによって，別の観点から階層的主成分分析を定義した。この方法では，2次合成変量の数を減らせば，集合間構造が強調され，2次合成変量の数を増やせば，集合内構造が強調されるように，1次合成変量の性質が変化することが示された。その際，最適化基準としては，2次合成変量のα係数の和の最大化が採用された。

しかし，この方法の集合間構造を強調する側の極限の解は，正準相関分析には一致しない。本研究では，集合ごとに導かれる1次合成変量の性質について，正準相関分析も含めた統一的な理解を目指す。4節において明らかになるように，主成分分析と正準相関分析を単一の方法に統合することは不可能であるが，本研究ではそれぞれの方法の一般化とみなすことができるような2つの方法を提案する。α係数の和の最大化基準は，2次合成変量を解釈の対象とする場合には便利であるが，通常の主成分分析と正準相関分析との関連性がやや希薄である。本研究では，2次合成変量の分散の和の最大化を最適化基準とした結果，集合ごとの個別の主成分分析と正準相関分析（およびその一般化）を特殊な場合として含む方法が定式化できることになった。

2. 方法の定式化

N人の被験者のn個の変数上における測定値が存在するものとする。ここで，変数は内容の区別等によって，p個の集合に分割されるものとする。個々の集合k（$k = 1,\ldots, p$）の変数の数をn_kとする。　$n = \sum_{k=1}^{p} n_k$

である。集合kのデータ行列を，$\boldsymbol{Z}_k$とする。$\boldsymbol{Z}_k$は$n_k \times N$の行列であって，各行の和が0，分散が1に基準化されているものとする。したがって，

$$\boldsymbol{R}_{kl} = \frac{1}{N}\boldsymbol{Z}_k\boldsymbol{Z}_l' \qquad k = 1,\ldots,p \quad l = 1,\ldots,p \qquad (2-1)$$

は，集合kと集合lの変数間の$n_k \times n_l$の相関行列である。

集合ごとに，データの1次結合としての合成変量，

$$\boldsymbol{F}_k = \boldsymbol{V}_k'\boldsymbol{Z}_k \qquad k = 1,\ldots,p \qquad (2-2)$$

を定義する。$\boldsymbol{V}_k$は$n_k \times q_k$の行列であり，q_kは，あらかじめ定められた合成変量の数で，$q_k \leqq n_k$とする。行列$\boldsymbol{F}_k$の要素を集合kの1次合成変量と呼ぶ。

1次合成変量について，

$$\frac{1}{N}\boldsymbol{F}_k\boldsymbol{F}_k' = \boldsymbol{V}_k'\boldsymbol{R}_{kk}\boldsymbol{V}_k = \boldsymbol{I} \qquad k = 1,\ldots,p \qquad (2-3)$$

なる制約条件を課す。すなわち，1次合成変量は，平均値が0（このことは，$\boldsymbol{Z}_k$の各行の和が0であることの必然的な帰結である），分散が1の標準得点で，集合内で相互に直交する。また，このことは，$\boldsymbol{V}_k$のランクがq_kであることも意味する。さらに，データ行列$\boldsymbol{Z}_k$のランクがq_kを越えていることを仮定したことにもなっている。

次に，1次合成変量の全集合にわたる線型結合として，2次合成変量，

$$\boldsymbol{G} = \sum_{k=1}^{p} \boldsymbol{W}_k' \boldsymbol{F}_k = \sum_{k=1}^{p} \boldsymbol{W}_k' \boldsymbol{V}_k' \boldsymbol{Z}_k \qquad (2-4)$$

を定義する。ここで，$\boldsymbol{W}_k$ は，$q_k \times Q$ の行列であり，Q はあらかじめ定められた2次合成変量の数であって，$\max q_k \leqq Q \leqq \sum_{k=1}^{p} q_k$ とする。また，$\boldsymbol{G}$ の各行は相互に1次独立であるとする。このことは，

$$\boldsymbol{W}' = [\boldsymbol{W}_1' \quad \boldsymbol{W}_2' \cdots \boldsymbol{W}_p'] \qquad (2-5)$$

によって定義される行列 $\boldsymbol{W}$ のランクが Q であることを意味するとともに，データ全体のランクが Q を越えることを仮定したことにもなっている。ここまでの，合成変量の定義及び制約条件は，村上（1987 a）のそれと実質的に同一である。

本研究では，2次合成変量の分散の和，すなわち，

$$S = \mathrm{tr}\frac{1}{N}\boldsymbol{G}\boldsymbol{G}' = \mathrm{tr}\sum_{k=1}^{p}\sum_{l=1}^{p} \boldsymbol{W}_k' \boldsymbol{V}_k' \boldsymbol{R}_{kl} \boldsymbol{V}_l \boldsymbol{W}_l \qquad (2-6)$$

を最大化する，という基準を満足する重み行列，$\boldsymbol{W}_k$，$\boldsymbol{V}_k$ を求めることを考える。

この定式化を，図1にもとづいて図示すると，図3のようになる。すなわち，全集合の合成変量の背後に ◇ で示すもう1つ高次の合成変量を仮定するわけである。後に示すように，この2次合成変量の数 Q を増減することによって，一次合成変量の質，すなわち，重み行列 $\boldsymbol{V}_k$ の内容が変化する。それが，bandwidth と fidelity の間のバランスを調整する働きをするわけである。

ここで，行列 $\boldsymbol{W}_k$ に対する制約条件として，次の2つの場合を考える。

（Ⅰ）階層的主成分分析

$$\sum_{k=1}^{p} \boldsymbol{W}_k' \boldsymbol{V}_k' \boldsymbol{V}_k \boldsymbol{W}_k = \boldsymbol{I} \qquad (2-7)$$

（Ⅱ）一般化された正準相関分析

$$\sum_{k=1}^{p} \boldsymbol{W}_k' \boldsymbol{W}_k = \boldsymbol{I} \qquad (2-8)$$

（Ⅰ）による解が，集合ごとの個別主成分分析を，（Ⅱ）による解が，正準相関分析（あるいは，Horst（1961）による，その3つ以上の集合への一般化）を，特殊な場合として含んでいることは5節で示される。

この種の合成変量タイプの分析の解釈にあたって，重み行列そのものの要素に主としてもとづくか，あるいは，合成変量と個々の変数との間の相関係数からなる構造行列（因子分析の伝統にしたがって，この要素は「負荷」と呼ばれることも多い）によるかは，しばしば問題になる。一般には，主成分分析の解釈にあたっては，負荷が解釈されるケースが多く，正準相関分析にあたっては，

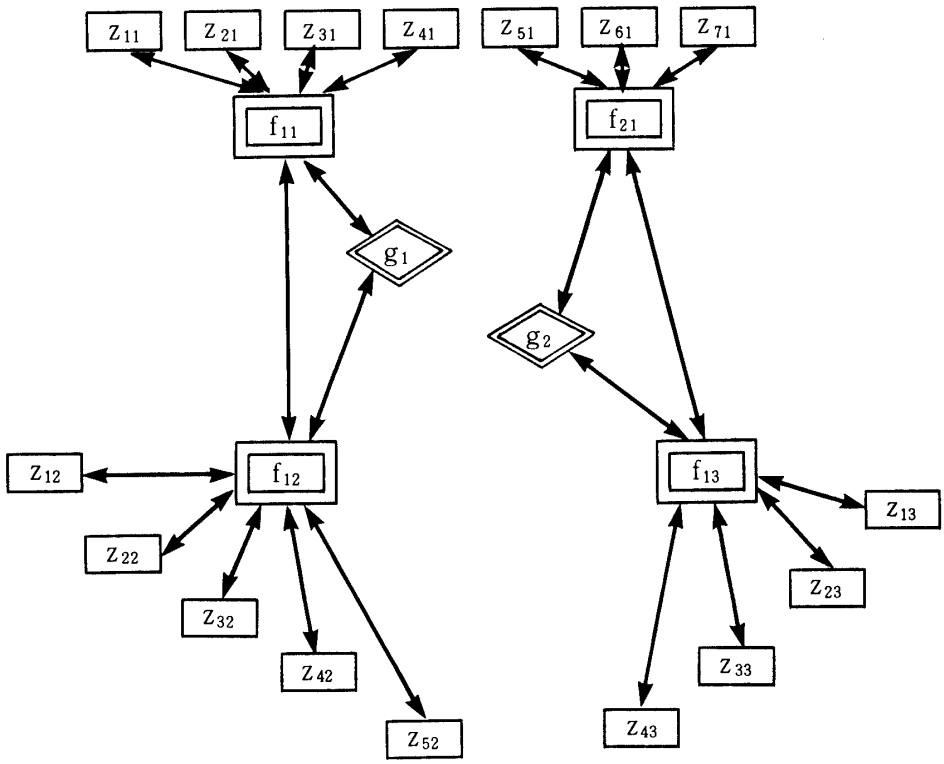

図3　個別変数，合成変量に加えて2次合成変量（ ◇ 内）を定義する。図1と同様，相関関係は適宜省略。属性が図から消えているのは，それを考慮するのをやめたからではなく，単にスペースの都合にすぎない。

重みそのものが解釈の対象とされることが多い。しかしながら，Levine（1977）のように，正準相関分析においても構造行列を解釈することを勧めるものもある。ここでは，構造行列を解釈する立場をとることにする。

集合 k の1次合成変量の構造行列は，

$$\boldsymbol{A}_k = \boldsymbol{R}_{kk}\boldsymbol{V}_k \qquad k = 1, \ldots, p \quad (2-9)$$

によって定義される。

また，この分析では，1次合成変量の集合間相関にも関心が寄せられることになろう。これは，

$$\boldsymbol{\Phi}_{kl} = \boldsymbol{V}_k'\boldsymbol{R}_{kl}\boldsymbol{V}_l \qquad k = 1, \ldots, p \quad l = 1, \ldots, p \quad (2-10)$$

によって与えられる。

2次合成変量の解釈については，このモデルではあまり考慮しないこととする。どちらかと言えば，1次合成変量の性質を調整するための手段という色彩が強いからである。

3. 階層的主成分分析のアルゴリズム

アルゴリズムの記述にあたって重み行列 $\boldsymbol{W}_k$ について若干の仮定を追加しておく。すなわち，$\boldsymbol{W}_k$ のランクは q_k であるとする。これは証明の簡略化のための条件であり，実用上強い制約となるようなものではない。実際，現実のデータへの適用にあたっては，どちらも問題なく満たされると考えられる。

式（2－6）で定義される S を，（2－3），（2－7）の制約条件の下で最大化するためには，

$$S^* = S - \mathrm{tr}\,\boldsymbol{L}^*\left(\sum_{k=1}^{p}\boldsymbol{W}_k'\boldsymbol{V}_k'\boldsymbol{V}_k\boldsymbol{W}_k - \boldsymbol{I}\right) - \mathrm{tr}\sum_{k=1}^{p}\boldsymbol{M}_k^*\left(\boldsymbol{V}_k'\boldsymbol{R}_{kk}\boldsymbol{V}_k - \boldsymbol{I}\right) \quad (3-1)$$

の停留条件を求めればよい。ここで，$\boldsymbol{L}^*$，$\boldsymbol{M}_k^*$ はラグランジュの定数行列である。

このために，S^* を，$\boldsymbol{V}_k$，$\boldsymbol{W}_k$ で偏微分して0とおくことにより，

$$\sum_{l=1}^{p}\boldsymbol{R}_{kl}\boldsymbol{V}_l\boldsymbol{W}_l\boldsymbol{W}_k' - \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{L}\boldsymbol{W}_k' - \boldsymbol{R}_{kk}\boldsymbol{V}_k\boldsymbol{M}_k = \boldsymbol{0} \qquad k = 1, \ldots, p \quad (3-2)$$

$$\sum_{l=1}^{p}\boldsymbol{V}_k'\boldsymbol{R}_{kl}\boldsymbol{V}_l\boldsymbol{W}_l - \boldsymbol{V}_k'\boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{L} = \boldsymbol{0} \qquad k = 1, \ldots, p \quad (3-3)$$

が得られる。ここで，

$$\boldsymbol{M}_k = \frac{\boldsymbol{M}_k^* + \boldsymbol{M}_k^{*\prime}}{2} \qquad k = 1, \ldots, p \quad (3-4)$$

$$\boldsymbol{L} = \frac{\boldsymbol{L}^* + \boldsymbol{L}^{*\prime}}{2} \quad (3-5)$$

である。

式（3－2）の左から $\boldsymbol{V}_k'$ をかけ，（2－3），（3－3）を用いて，

$$\boldsymbol{M}_k = \left(\sum_{l=1}^{p}\boldsymbol{V}_k'\boldsymbol{R}_{kl}\boldsymbol{V}_l\boldsymbol{W}_l - \boldsymbol{V}_k'\boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{L}\right)\boldsymbol{W}_k' = \boldsymbol{0} \qquad k = 1, \ldots, p \quad (3-6)$$

次に，行列 $\boldsymbol{V}_k$ のランクは q_k であると仮定したので，

$$\boldsymbol{\Xi} = (\boldsymbol{V}_k'\boldsymbol{V}_k)^{\frac{1}{2}} \qquad k = 1, \ldots, p \quad (3-7)$$

およびこの逆行列 $\boldsymbol{\Xi}^{-1}$ は必ず存在する。これを用いて，

$$\boldsymbol{\gamma}_k^* = \boldsymbol{V}_k\boldsymbol{\Xi}^{-1} \qquad k = 1, \ldots, p \quad (3-8)$$

$$\boldsymbol{\Gamma}_k^* = \boldsymbol{\Xi}\boldsymbol{W}_k \qquad k = 1, \ldots, p \quad (3-9)$$

を定義する。$\boldsymbol{\Xi}$ の定義および（2－7）により，$\boldsymbol{\gamma}_k^*$ と，$\boldsymbol{\Gamma}_k^*$ は，

$$\boldsymbol{\gamma}_k^{*\prime}\boldsymbol{\gamma}_k^* = \boldsymbol{I} \qquad k = 1, \ldots, p \quad (3-10)$$

$$\sum_{k=1}^{p}\boldsymbol{\Gamma}_k^{*\prime}\boldsymbol{\Gamma}_k^* = \boldsymbol{I} \quad (3-11)$$

の条件を満足する。

式（3－3）の左から $\boldsymbol{\Xi}^{-1}$ を掛けて整理することにより，

$$\sum_{l=1}^{p}\boldsymbol{\gamma}_k^{*\prime}\boldsymbol{R}_{kl}\boldsymbol{\gamma}_l^*\boldsymbol{\Gamma}_l^* = \boldsymbol{\Gamma}_k^*\boldsymbol{L} \qquad k = 1, \ldots, p \quad (3-12)$$

を得る。

ここで，（3－12）の左から $\boldsymbol{\Gamma}_k^{*\prime}$ を掛け，k について総和し，（3－11）を用いて，

$$\boldsymbol{L} = \sum_{k=1}^{p}\sum_{l=1}^{p}\boldsymbol{\Gamma}_k^{*\prime}\boldsymbol{\gamma}_k^{*\prime}\boldsymbol{R}_{kl}\boldsymbol{\gamma}_l^*\boldsymbol{\Gamma}_l^* \quad (3-13)$$

したがって，（3－8），（3－9），（2－6）により，

$$\boldsymbol{L} = \sum_{k=1}^{p}\sum_{l=1}^{p}\boldsymbol{W}_k'\boldsymbol{V}_k'\boldsymbol{R}_{kl}\boldsymbol{V}_l\boldsymbol{W}_l = \frac{1}{N}\boldsymbol{G}\boldsymbol{G}' \quad (3-14)$$

となる。行列 $\boldsymbol{W}$ に関する仮定により $Q \times Q$ の行列 $\boldsymbol{L}$ は正定符号である。したがって，適当な正定符号の対角行列 $\boldsymbol{\Theta}$ と直交行列 $\boldsymbol{U}$ を用いて，

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Theta}\boldsymbol{U}' \quad (3-15)$$

のように分解することができる。

次に，（3－2）の右からΞをかけ，（3－6）を代入して整理することにより，

$$\sum_{l=1}^{p} \boldsymbol{R}_{kl} \boldsymbol{\Gamma}_l^* \boldsymbol{\Gamma}_l^* \boldsymbol{\Gamma}_k^{*\prime} = \boldsymbol{\Gamma}_k^* \boldsymbol{\Gamma}_k^* \boldsymbol{L} \boldsymbol{\Gamma}_k^{*\prime} \qquad k = 1, \ldots, p \quad (3-16)$$

行列$\boldsymbol{W}_k$に関する仮定により，$\boldsymbol{\Gamma}_k^* \boldsymbol{L} \boldsymbol{\Gamma}_k^{*\prime}$のランクは$q_k$であり，したがって正定符号である。したがって，適当な正定符号の対角行列$\boldsymbol{\Lambda}_k$と直交行列$\boldsymbol{T}_k$をもちいて，

$$\boldsymbol{\Gamma}_k^* \boldsymbol{L} \boldsymbol{\Gamma}_k^{*\prime} = \boldsymbol{T}_k \boldsymbol{\Lambda}_k \boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (3-17)$$

と分解することができる。

そこで，

$$\boldsymbol{\gamma}_k = \boldsymbol{\gamma}_k^* \boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (3-18)$$

$$\boldsymbol{\Gamma}_k = \boldsymbol{T}_k \boldsymbol{\Gamma}_k^* \boldsymbol{U} \qquad k = 1, \ldots, p \quad (3-19)$$

と定義すると，（3－12）は左から$\boldsymbol{T}_k$を，右から$\boldsymbol{U}'$を掛け，（3－16）は右から$\boldsymbol{T}_k'$を掛けて整理することにより，それぞれ，

$$\sum_{l=1}^{p} \boldsymbol{\gamma}_k' \boldsymbol{R}_{kl} \boldsymbol{\gamma}_l \boldsymbol{\Gamma}_l = \boldsymbol{\Gamma}_k \boldsymbol{\Theta} \qquad k = 1, \ldots, p \quad (3-20)$$

$$\sum_{l=1}^{p} \boldsymbol{R}_{kl} \boldsymbol{\gamma}_l \boldsymbol{\Gamma}_l \boldsymbol{\Gamma}_k' = \boldsymbol{\gamma}_k \boldsymbol{\Lambda}_k \qquad k = 1, \ldots, p \quad (3-21)$$

のように書き換えられる。式（3－20）から，

$$\boldsymbol{\Gamma}_k = \sum_{l=1}^{p} \boldsymbol{\gamma}_k' \boldsymbol{R}_{kl} \boldsymbol{\gamma}_l \boldsymbol{\Gamma}_l \boldsymbol{\Theta}^{-1} \qquad k = 1, \ldots, p \quad (3-22)$$

だから，これを（3－21）に代入して，

$$\{(\sum_{l=1}^{p} \boldsymbol{R}_{kl} \boldsymbol{\gamma}_l \boldsymbol{\Gamma}_l) \boldsymbol{\Theta}^{-1} (\sum_{l=1}^{p} \boldsymbol{\Gamma}_l' \boldsymbol{\gamma}_l' \boldsymbol{R}_{lk})\} \boldsymbol{\gamma}_k = \boldsymbol{\gamma}_k \boldsymbol{\Lambda}_k \qquad k = 1, \ldots, p \quad (3-23)$$

となる。

$$\boldsymbol{\Psi}_{kl} = \boldsymbol{\gamma}_k' \boldsymbol{R}_{kl} \boldsymbol{\gamma}_l \qquad k = 1, \ldots, p \quad l = 1, \ldots, p \quad (3-24)$$

$$\boldsymbol{\Omega}_k = \sum_{l=1}^{p} \boldsymbol{R}_{kl} \boldsymbol{\gamma}_l \boldsymbol{\Gamma}_l \qquad k = 1, \ldots, p \quad (3-25)$$

と定義し，

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} & \cdots & \boldsymbol{\Psi}_{1p} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} & \cdots & \boldsymbol{\Psi}_{2p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Psi}_{p1} & \boldsymbol{\Psi}_{p2} & \cdots & \boldsymbol{\Psi}_{pp} \end{pmatrix} \quad (3-26)$$

$$\boldsymbol{\Gamma}' = \begin{bmatrix} \boldsymbol{\Gamma}_1' & \boldsymbol{\Gamma}_2' & \cdots & \boldsymbol{\Gamma}_p' \end{bmatrix} \quad (3-27)$$

とすると，（3－20）と（3－23）は，

$$\boldsymbol{\Psi} \boldsymbol{\Gamma} = \boldsymbol{\Gamma} \boldsymbol{\Theta} \quad (3-28)$$

$$\boldsymbol{\Omega}_k \boldsymbol{\Theta}^{-1} \boldsymbol{\Omega}_k' \boldsymbol{\gamma}_k = \boldsymbol{\gamma}_k \boldsymbol{\Lambda}_k \qquad k = 1, \ldots, p \quad (3-29)$$

となる。

かくして，$\boldsymbol{\Gamma}$は，$\sum_{k=1}^{p} q_k \times \sum_{k=1}^{p} q_k$の行列$\boldsymbol{\Psi}$の基準化された固有ベクトルとして，$\boldsymbol{\gamma}_k$は，$n_k \times n_k$の行列$\boldsymbol{\Omega}_k \boldsymbol{\Theta}^{-1} \boldsymbol{\Omega}_k'$の基準化された固有ベクトルとして，それぞれ得られることになる。

なお，（2－6），（3－14）から，

$$S = \mathrm{tr}\boldsymbol{L} = \mathrm{tr}\boldsymbol{\Theta} \quad (3-30)$$

であり，（3－23）の左から$\boldsymbol{\gamma}_k'$を掛けると，

$$(\sum_{l=1}^{p} \boldsymbol{\Psi}_{kl} \boldsymbol{\Gamma}_l) \boldsymbol{\Theta}^{-1} (\sum_{l=1}^{p} \boldsymbol{\Gamma}_l' \boldsymbol{\Psi}_{lk}) = \boldsymbol{\Gamma}_k \boldsymbol{\Theta} \boldsymbol{\Gamma}_k' = \boldsymbol{\Lambda}_k \qquad k = 1, \ldots, p$$

となるから，

$$\mathrm{tr} \sum_{k=1}^{p} \boldsymbol{\Lambda}_k = \mathrm{tr}\boldsymbol{\Theta} \sum_{k=1}^{p} \boldsymbol{\Gamma}_k' \boldsymbol{\Gamma}_k = \mathrm{tr}\boldsymbol{\Theta}$$

となる。これにより，

$$S = \mathrm{tr} \sum_{k=1}^{p} \boldsymbol{\Lambda}_k \quad (3-31)$$

これが，最大化すべきものであるから，固有ベクトルは，それぞれ大小順にQ番目，およびq_k番目までの固有値に対応するものをとればよい。

もちろん，$\boldsymbol{\gamma}_k$が既知でなければ（3－28）は解けないし，$\boldsymbol{\gamma}_k$と$\boldsymbol{\Gamma}_k$がともに既知でなければ（3－29）は解けないことになるが，$\boldsymbol{\gamma}_k$の適当な初期値を用いて，（3－28）から$\boldsymbol{\Gamma}_k$の近似値を求め，それらを用いて，（3－29）から$\boldsymbol{\gamma}_k$の近似値を求めて，（3－28）にもどる，というプロセスを収束するまで反復することによって，（3－28）と（3－29）を同時に満足する$\boldsymbol{\gamma}_k$と$\boldsymbol{\Gamma}_k$を得ることができるはずである。

なおこの解は結局のところ，

$$\Gamma_k'\Gamma_k = \boldsymbol{I} \qquad k = 1, \ldots, p \quad (3-32)$$

$$\sum_{k=1}^{p} \Gamma_k'\Gamma_k = \boldsymbol{I} \quad (3-33)$$

を制約条件として，

$$S = \mathrm{tr}\sum_{k=1}^{p}\sum_{l=1}^{p} \Gamma_k'\Gamma_k'\boldsymbol{R}_{kl}\Gamma_l\Gamma_l \quad (3-34)$$

を最大化する解をもとめていることになることに注意しよう。このことは，（3－26）で示される行列 Ψ の，大小順に Q 番目までの固有値の和を最大化する，ということであると解することもできる。Ψ は，合成変量 $\Gamma_k'\boldsymbol{Z}_k$ の共分散行列であり，Qが小さい場合には，その少数の固有値の和を最大化するためには，この非対角成分である異なる集合間の共分散を大きくするように，Γ_kを決めなければならない。Q が大となるにつれて，Ψ の対角成分を大きくすることが，固有値の和の最大化につながるから，今度は集合間共分散よりも，合成変量の分散の拡大の方に重みがかかることになる。こうして，Q の増減は，集合内構造と集合間構造の間のバランスを変化させる役割をはたすことが納得されよう。

さて，こうして，Γ_k と Γ_k が得られたとすると，求める重み行列，$\boldsymbol{V}_k$ と $\boldsymbol{W}_k$ は，

$$\boldsymbol{V}_k = \Gamma_k\boldsymbol{T}_k'\Xi_k \qquad k = 1, \ldots, p \quad (3\ \ 35)$$

$$\boldsymbol{W}_k = \Xi_k^{-1}\boldsymbol{T}_k\Gamma_k\boldsymbol{U}' \quad k = 1, \ldots, p \quad (3-36)$$

によって得られることになる。ここで，Ξ_k は，(2－3) により，

$$\boldsymbol{I} = \boldsymbol{V}_k'\boldsymbol{R}_{kk}\boldsymbol{V}_k = \Xi_k\boldsymbol{T}_k(\Gamma_k'\boldsymbol{R}_{kk}\Gamma_k)\boldsymbol{T}_k'\Xi_k$$

だから，

$$\Xi_k^{-2} = \boldsymbol{T}_k(\Gamma_k'\boldsymbol{R}_{kk}\Gamma_k)\boldsymbol{T}_k' = \boldsymbol{T}_k\Psi_{kk}\boldsymbol{T}_k'$$

したがって

$$\Xi_k = \boldsymbol{T}_k\Psi_{kk}^{-\frac{1}{2}}\boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (3-37)$$

によって得られることになる。ここで，$\Psi_{kk}^{-\frac{1}{2}}$ は，Λ_k^* を正定符号の対角行列，$\boldsymbol{X}_k$ を直交行列として，

$$\Psi_{kk} = \boldsymbol{X}_k\Lambda_k^*\boldsymbol{X}_k' \qquad k = 1, \ldots, p \quad (3-38)$$

のように分解したとき，

$$\Psi_{kk}^{-\frac{1}{2}} = \boldsymbol{X}_k\Lambda_k^{*-\frac{1}{2}}\boldsymbol{X}_k' \quad k = 1, \ldots, p \quad (3-39)$$

によって与えられるものである。

こうして，重みベクトルは，

$$\boldsymbol{V}_k = \Gamma_k\Psi_{kk}^{-\frac{1}{2}}\boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (3-40)$$

$$\boldsymbol{W}_k = \boldsymbol{T}_k\Psi_{kk}^{\frac{1}{2}}\Gamma_k\boldsymbol{U}' \quad k = 1, \ldots, p \quad (3-41)$$

となる。

最終的に2つの直交行列 $\boldsymbol{T}_k$ と $\boldsymbol{U}$ が不定のまま残された。このうち $\boldsymbol{T}_k$ は，（2－9）によって与えられる1次合成変量の構造行列，$\boldsymbol{A}_k$ を単純構造に近付けるような回転を行うことで決定する。$\boldsymbol{U}$ については，1次合成変量に影響を与えないので，とりあえず不定のままとしておこう。

以上のアルゴリズムをまとめると，

① 全変数間の $n \times n$ の相関行列，

$$\boldsymbol{R} = \frac{1}{N}\boldsymbol{Z}\boldsymbol{Z}' \quad (3-42)$$

を算出する。ここで，

$$\boldsymbol{Z}' = [\boldsymbol{Z}_1' \quad \boldsymbol{Z}_2' \quad \cdots \quad \boldsymbol{Z}_p'] \quad (3-43)$$

であり，

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} & \cdots & \boldsymbol{R}_{1p} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} & \cdots & \boldsymbol{R}_{2p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \boldsymbol{R}_{p1} & \boldsymbol{R}_{p2} & \cdots & \boldsymbol{R}_{pp} \end{bmatrix} \quad (3-44)$$

である。

② $\boldsymbol{R}_{kk}$ の大小順に q_k 番目までの固有値に対応する固有ベクトルからなる $n_k \times q_k$ の行列を Γ_k の0次近似(初期値）とする。$t = 0$ とおく。

③ 式（3－24）によって，（3－26）の行列 Ψ を求め，その最大 Q 番目までの固有値を要素とする対角行列を Θ の，対応する固有ベクトルを列とする $\sum q_k \times Q$ の行列を Γ のそれぞれ第 t 次近似とする。

④ Γ_k と Γ_k の第 t 次近似を用いて，（3－25）により行列 Ω_k を求める。これと Θ を用いて，行列 $\Omega_k\Theta^{-1}\Omega_k'$ の大小順に q_k 番目までの固有値に対応する固有ベクトルを列する行列を Γ_k の第 $t+1$ 次近似値とする。

⑤ 収束したら⑥へ。そうでなければ，t を $t+1$ におきかえて③へ。

⑥ Ψ_{kk} の全ての固有値と固有ベクトルを求め，(3－39) によって $\Psi_{kk}^{\frac{1}{2}}$ を求める。

$$\boldsymbol{V}_k^* = \Gamma_k\Psi_{kk}^{-\frac{1}{2}} \qquad k = 1, \ldots, p \quad (3-45)$$

$$\boldsymbol{W}_k^* = \Psi_{kk}^{\frac{1}{2}}\Gamma_k \qquad k = 1, \ldots, p \quad (3-46)$$

によって，回転前の重み行列を求める。

⑦ 次の式によって，回転前の負荷行列を求める。

$$\boldsymbol{A}_k^* = \boldsymbol{R}_{kk}\boldsymbol{V}_k^* \qquad k = 1, \ldots, p \quad (3-47)$$

⑧ $\boldsymbol{A}_k^*$をバリマックス回転して，最終的な負荷行列，

$$\boldsymbol{A}_k = \boldsymbol{A}_k^*\boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (3-48)$$

を求め，出力する。この回転行列 $\boldsymbol{T}_k$ は保存し，これによって，最終的な重み行列，

$$\boldsymbol{V}_k = \boldsymbol{V}_k^*\boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (3-49)$$

を求め出力する。

⑨ 式（2－10）によって，1次合成変量間の相関行列 $\boldsymbol{\Phi}_{kl}$ をもとめる。これらは，

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} & \cdots & \boldsymbol{\Phi}_{1p} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} & \cdots & \boldsymbol{\Phi}_{2p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Phi}_{p1} & \boldsymbol{\Phi}_{p2} & \cdots & \boldsymbol{\Phi}_{pp} \end{pmatrix} \quad (3-50)$$

の形の行列 $\boldsymbol{\Phi}$ として出力する。

⑩ 行列 $\boldsymbol{W}_k$（あるいは，（2－5）で定義される行列 $\boldsymbol{W}$）は，$\boldsymbol{W}_k^*$ のまま，あるいは，これをバリマックス回転する。これも一応出力しておく。

4. 一般化された正準相関分析のアルゴリズム

3節同様，2節の仮定に加えて，すべての $\boldsymbol{W}_k$ のランクが q_k であるという仮定を追加するが，この節では更にもう1つ仮定が加わる。すなわち，各集合内の変数間相関行列，$\boldsymbol{R}_{kk}$ のランクは，その集合の変数の数 n_k に等しいとする。この仮定は被験者数 N が n_k をかなり上回るという普通の条件ではまず満たされていると考えてよく，さほど無理なものではない。

式（2－6）で定義される S を，今度は（2－3），（2－8）の条件の下で最大化することが目標であるから，

$$S^* = S - \mathrm{tr}\boldsymbol{L}^*\left(\sum_{k=1}^{p}\boldsymbol{W}_k'\boldsymbol{W}_k' - \boldsymbol{I}\right) - \mathrm{tr}\sum_{k=1}^{p}\boldsymbol{M}_k^*\left(\boldsymbol{V}_k'\boldsymbol{R}_{kk}\boldsymbol{V}_k - \boldsymbol{I}\right) \quad (4-1)$$

の停留条件を求めればよい。$\boldsymbol{L}^*$，$\boldsymbol{M}_k^*$ は，ともにラグランジュの定数行列である。S^* を $\boldsymbol{V}_k$，$\boldsymbol{W}_k$ で偏微分して0とおくことにより，

$$\sum_{l=1}^{p}\boldsymbol{R}_{kl}\boldsymbol{V}_l\boldsymbol{W}_l\boldsymbol{W}_k' - \boldsymbol{R}_{kk}\boldsymbol{V}_k\boldsymbol{M}_k = \boldsymbol{0} \qquad k = 1, \ldots, p \quad (4-2)$$

$$\sum_{l=1}^{p}\boldsymbol{V}_k'\boldsymbol{R}_{kl}\boldsymbol{V}_l\boldsymbol{W}_l - \boldsymbol{W}_k\boldsymbol{L} = \boldsymbol{0} \qquad k = 1, \ldots, p \quad (4-3)$$

が得られる。ここで，

$$\boldsymbol{M}_k = \frac{\boldsymbol{M}_k^* + \boldsymbol{M}_k^{*\prime}}{2} \qquad k = 1, \ldots, p \quad (4-4)$$

$$\boldsymbol{L} = \frac{\boldsymbol{L}^* + \boldsymbol{L}^{*\prime}}{2} \quad (4-5)$$

である。

次に，

$$\boldsymbol{P}_{kl} = \boldsymbol{R}_{kk}^{-\frac{1}{2}}\boldsymbol{R}_{kl}\boldsymbol{R}_{ll}^{-\frac{1}{2}} \qquad k = 1, \ldots, p \quad l = 1, \ldots, p \quad (4-6)$$

$$\boldsymbol{r}_k^* = \boldsymbol{R}_{kk}^{\frac{1}{2}}\boldsymbol{V}_k \qquad k = 1, \ldots, p \quad (4-7)$$

とおく。ここで，（2－3）により，

$$\boldsymbol{r}_k^{*\prime}\boldsymbol{r}_k^* = \boldsymbol{I} \qquad k = 1, \ldots, p \quad (4-8)$$

である。そこで，（3－2）の左から $\boldsymbol{R}_{kk}^{-\frac{1}{2}}$ を掛けると，

$$\sum_{l=1}^{p}\boldsymbol{P}_{kl}\boldsymbol{r}_l^*\boldsymbol{W}_l\boldsymbol{W}_k' - \boldsymbol{r}_k^*\boldsymbol{M}_k = \boldsymbol{0} \qquad k = 1, \ldots, p \quad (4-9)$$

が得られ，（4－3）は，

$$\sum_{l=1}^{p}\boldsymbol{r}_k^{*\prime}\boldsymbol{P}_{kl}\boldsymbol{r}_l^*\boldsymbol{W}_l - \boldsymbol{W}_k\boldsymbol{L} = \boldsymbol{0} \qquad k = 1, \ldots, p \quad (4-10)$$

と書き換えられる。

式（4－9）の左から $\boldsymbol{W}_k'$ を掛け，k について総和すると，（2－8）により，

$$\boldsymbol{L} = \sum_{k=1}^{p}\sum_{l=1}^{p}\boldsymbol{W}_k'\boldsymbol{r}_k^{*\prime}\boldsymbol{P}_{kl}\boldsymbol{r}_l^*\boldsymbol{W}_l = \sum_{k=1}^{p}\sum_{l=1}^{p}\boldsymbol{W}_k'\boldsymbol{V}_k'\boldsymbol{R}_{kl}\boldsymbol{V}_l\boldsymbol{W}_l \quad (4-11)$$

これは，2次合成変量間の共分散行列であるから，仮定により，$\boldsymbol{L}$ は正定符号の行列で，適当な正定符号の対角行列 $\boldsymbol{\Theta}$ と直交行列 $\boldsymbol{T}^*$ を用いて，

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Theta}\boldsymbol{U}' \quad (4-12)$$

と分解できる。

次に，（4－9）の左から $\boldsymbol{r}_k^{*\prime}$ を掛け，（4－8），（4－10）を用いて，

$$\boldsymbol{M}_k = \boldsymbol{W}_k \boldsymbol{L} \boldsymbol{W}_k' \qquad k = 1, \ldots, p \quad (4-13)$$

$\boldsymbol{L}$が，正定符号の行列であることと，$\boldsymbol{W}_k$のランクに関する仮定により，$\boldsymbol{M}_k$は正定符号行列で，適当な正定符号の対角行列Λ_kと，直交行列$\boldsymbol{T}_k$を用いて，

$$\boldsymbol{M}_k = \boldsymbol{T}_k \Lambda_k \boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (4-14)$$

と分解できる。

そこで，

$$\Gamma_k = \Gamma_k^* \boldsymbol{T}_k \qquad k = 1, \ldots, p \quad (4-15)$$

$$\Gamma_k = \boldsymbol{T}_k' \boldsymbol{W}_k \boldsymbol{U} \qquad k = 1, \ldots, p \quad (4-16)$$

と定義すると，(4－10) の左から$\boldsymbol{T}_k'$を，右から$\boldsymbol{U}$を掛けて，

$$\sum_{l=1}^{p} \Gamma_k' \boldsymbol{P}_{kl} \Gamma_l \Gamma_l = \Gamma_k \Theta \qquad k = 1, \ldots, p \quad (4-17)$$

が得られ，また，(4－9) の右から$\boldsymbol{T}_k$を掛けて，

$$\sum_{l=1}^{p} \boldsymbol{P}_{kl} \Gamma_l \Gamma_l \Gamma_k' = \Gamma_k \Lambda_k \qquad k = 1, \ldots, p \quad (4-18)$$

が得られる。

これらは，$\boldsymbol{R}_{kl}$が$\boldsymbol{P}_{kl}$に変わる以外，階層的主成分分析における，(3－20)，(3－21) 式と，まったく同じであり，以下，(4－18) を，

$$\{(\sum_{l=1}^{p} \boldsymbol{P}_{kl} \Gamma_l \Gamma_l) \Theta^{-1} (\sum_{l=1}^{p} \Gamma_l' \Gamma_l' \boldsymbol{P}_{lk})\} \Gamma_k = \Gamma_k \Lambda_k \qquad k = 1, \ldots, p \quad (4-19)$$

と書き換えた上で，

$$\Psi_{kl} = \Gamma_k' \boldsymbol{P}_{kl} \Gamma_l \qquad k = 1, \ldots, p \quad l = 1, \ldots, p \quad (4-20)$$

$$\Omega_k = \sum_{l=1}^{p} \boldsymbol{P}_{kl} \Gamma_l \Gamma_l \qquad k = 1, \ldots, p \quad (4-21)$$

と定義し，(3－26)，(3－27) でΨ，Ωを定義して，(3－28)，(3－29) を同時に満足するΓ_k，Γ_kを求めるところまでは，完全な平行関係にある。

この場合，Ψの対角要素はすべて1であり，Ψは相関行列である。3節の階層的主成分分析における解釈と同様，この相関行列のQ番目までの固有値の和を最大化するのがこのアルゴリズムである。ただこの場合は，$Q = \sum q_k$としたときには，固有値の和は$\sum q_k$であることが明らかであり，解は完全に不定となる。このことから類推すれば，この方法ではQを$\max q_k$よりあまり大きくすることは，解釈上もあまり有効性のない解を生み出すことになりそうである。

さて，求める重み行列，$\boldsymbol{V}_k$と$\boldsymbol{W}_k$は，

$$\boldsymbol{V}_k = \boldsymbol{R}_{kk}^{-\frac{1}{2}} \Gamma_k \boldsymbol{T}_k' \qquad k = 1, \ldots, p \quad (4-22)$$

$$\boldsymbol{W}_k = \boldsymbol{T}_k \Gamma_k \boldsymbol{U}' \qquad k = 1, \ldots, p \quad (4-23)$$

によって求められるが，この段階で不定なのは，2つの直交行列だけであり，$\boldsymbol{T}_k$は負荷行列$\boldsymbol{A}_k$を単純構造化する回転によって定めることができる。

以上のアルゴリズムをまとめておこう。

① 3節①と同じ。

② $\boldsymbol{R}_{kk}$の固有値のすべての固有値を対角要素とする行列Λ_k^*と対応する固有ベクトルを各列にもつ行列$\boldsymbol{X}_k$を用いて，

$$\boldsymbol{R}_{kk}^{-\frac{1}{2}} = \boldsymbol{X}_k \Lambda_k^{*-\frac{1}{2}} \boldsymbol{X}_k' \quad k = 1, \ldots, p \quad (4-24)$$

を求める。

③ 式（4－6）にもとづいて，行列$\boldsymbol{P}_{kl}$を求める。これらは，

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} & \cdots & \boldsymbol{P}_{1p} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} & \cdots & \boldsymbol{P}_{2p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \boldsymbol{P}_{p1} & \boldsymbol{P}_{p2} & \cdots & \boldsymbol{P}_{pp} \end{pmatrix} \quad (4-25)$$

のように配列される。

④ 行列$\boldsymbol{P}$の大小順に$\max q_k$個の固有値に対応する固有ベクトルの行列，$\boldsymbol{Q}$を求める。この行列は，

$$\boldsymbol{Q}' = [\boldsymbol{Q}_1' \quad \boldsymbol{Q}_2' \quad \cdots \quad \boldsymbol{Q}_p'] \quad (4-26)$$

のように分割できる。$n_k \times \max p_k$の行列$\boldsymbol{Q}_k$のp_k列までを，Γ_kの0次近似とする。（このことの根拠は5節で示す。）$t = 0$とおく。

⑤ 行列Ψの要素行列をもとめるための式が(3－24)から（4－20）に変わること以外は，3節の③と同じ。

⑥ 行列Ω_kを求めるための式が(3－25)から(4－21)に変わる以外3節の④と同じ。

⑦ 収束したら⑧へ，そうでなければ，tを$t+1$に置き換えて⑤へ。

⑧ 次の式で回転前の重み行列$\boldsymbol{V}_k^*$を求める。Γ_kはそのまま$\boldsymbol{W}_k^*$となる。

$$\boldsymbol{V}_k^* = \boldsymbol{R}_{kk}^{-\frac{1}{2}} \boldsymbol{r}_k \qquad k = 1, \ldots, p \quad (4-27)$$

⑨～⑫　3節の⑦～⑩と同じ。

なお，正準相関分析では，構造行列を回転することはあまり行われていないが，Cliff & Krus (1976)が，これに関する一応の論理的根拠を与えている。

もう一つ，この方法には次のような解釈もあり得ることをつけ加えておこう。

一般に，$n \times N$の標準化されたデータ行列を$\boldsymbol{Z}$とし，その一次変換，

$$\boldsymbol{Z}^* = \boldsymbol{U}'\boldsymbol{Z} \quad (4-28)$$

を考える。ただし，$\boldsymbol{Z}$のランクはn（したがって，$n \leqq N$），$\boldsymbol{U}$は$n \times n$の行列，したがって，$\boldsymbol{Z}^*$もまた$n \times N$の行列である。ここで，$\boldsymbol{Z}^*$の各行は分散1で相互に直交する変量をあらわすと仮定する。すなわち，

$$\frac{1}{N}\boldsymbol{Z}^*\boldsymbol{Z}^{*\prime} = \boldsymbol{I} \quad (4-29)$$

したがって，

$$\boldsymbol{R} = \frac{1}{N}\boldsymbol{Z}\boldsymbol{Z}' \quad (4-30)$$

とする（もちろん，$\boldsymbol{R}$は変数間相関行列である）と，

$$\boldsymbol{U}'\boldsymbol{R}\boldsymbol{U} = \boldsymbol{I} \quad (4-31)$$

である。このような変換は無数にありうるが，このうち，もとのデータ行列に最小2乗法の意味で最も近いもの，すなわち，

$$S = \mathrm{tr}\frac{1}{N}(\boldsymbol{Z}-\boldsymbol{Z}^*)(\boldsymbol{Z}-\boldsymbol{Z}^*)' \quad (4-32)$$

が最小になるようなものを考えよう。このとき，$\boldsymbol{U}$はどのようにとったらよいだろうか。

この問題は$\boldsymbol{L}^*$をラグランジュの定数行列として，

$$S^* = S - \mathrm{tr}\boldsymbol{L}^*(\boldsymbol{U}'\boldsymbol{R}\boldsymbol{U}-\boldsymbol{I}) \quad (4-33)$$

の停留条件をもとめることによって解かれる。

S^*を$\boldsymbol{U}$で偏微分して0とおくことにより，

$$\frac{1}{N}\boldsymbol{Z}(\boldsymbol{Z}-\boldsymbol{U}'\boldsymbol{Z})' - \boldsymbol{R}\boldsymbol{U}\boldsymbol{L} = \boldsymbol{0} \quad (4-34)$$

が得られる。ここで，

$$\boldsymbol{L} = \frac{\boldsymbol{L}^*+\boldsymbol{L}^{*\prime}}{2} \quad (4-35)$$

である。式（4—34）は，

$$\boldsymbol{R} - \boldsymbol{R}\boldsymbol{U} - \boldsymbol{R}\boldsymbol{U}\boldsymbol{L} = \boldsymbol{0} \quad (4-36)$$

これに$\boldsymbol{R}^{-1}$を掛けて，

$$\boldsymbol{U} = (\boldsymbol{I}+\boldsymbol{L})^{-1} \quad (4-37)$$

ここで，$\boldsymbol{L}$は，(4—35)により対称行列だから，$\boldsymbol{I}+\boldsymbol{L}$も対称行列，その逆行列である，$\boldsymbol{U}$も対称行列である。したがって，(4—31)は，

$$\boldsymbol{U}\boldsymbol{R}\boldsymbol{U} = \boldsymbol{I} \quad (4-38)$$

これは，

$$\boldsymbol{R}^{-1} = \boldsymbol{U}^2 \quad (4-39)$$

これを満たす行列$\boldsymbol{U}$は無数にあるが，Sを最小化するという条件を満足するのは，

$$\boldsymbol{U} = \boldsymbol{R}^{-\frac{1}{2}} \quad (4-40)$$

に限られる（Johnson, 1966）。*) こうして，行列，

$$\boldsymbol{Z}^* = \boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{Z} \quad (4-41)$$

は，もとのデータに最も近い正規直交な行列であることがわかった。

こうしてみると，この節で説明した一般化正準相関分析は，(4—41)にしたがって変換したデータに対して階層的主成分分析をほどこしている，という解釈がなりたつことになる。

5. 他の方法との関係

この節では，本研究で提案した2つのモデルが，従来のモデルとどのような関係にあるかを検討する。

まず，階層的主成分分析において，2次合成変量の数を1次合成変量の数の和に等しくとった場合，すなわち，

$$Q = \sum_{k=1}^{p} q_k \quad (5-1)$$

の場合から始めよう。この場合，階層的主成分分析の1次合成変量は，各集合ごとの個別の主成分分析によって得られる合成変量と同じになる。

このことを直接式の上で示すことは難しいようであるが，次のように考えれば納得されよう。3節において，階層的主成分分析のアルゴリズムが，行列ΨのQ個の固有値の和の最大化に等しいことを示した。$Q = \sum q_k$の場合，このことは行列Ψの全固有値の和，すなわち，行列Ψのトレースを最大化することに等しい。これは，$\sum_{k=1}^{p} \mathrm{tr}\Psi_{kk}$　の最大化を意味する。ここで，Ψ_{kk}の対角要素は各集合の変数の1次結合の分散を示している。ノルム1で直交する重みによって，これらが求められるということは，$\boldsymbol{M}_k^*$をラグランジュの定数行列として，

*) 大学入試センターの柳井晴夫教授のご教示に深く感謝する。

$$S_k = \mathrm{tr}\,\Gamma_k^{*\prime} R_{kk} \Gamma_k^* - \mathrm{tr}\,M_k^* (\Gamma_k^{*\prime}\Gamma_k^* - I) \quad (5-2)$$

の停留条件を求めることにつながるが，これは個別の主成分分析に他ならない。

次に，すべての集合の1次合成変量の数が2次合成変量の数に等しい，すなわち，

$$q_1 = q_2 = \cdots = q_p = Q \quad (5-3)$$

の場合を考えよう。この場合，階層的主成分分析の2次合成変量は，全変数間の相関行列，すなわち，(3—44) によって与えられる R の主成分分析と同等になる。

このことは，次のようにして示される。行列 $\Gamma_k^{*\prime}$ は (5—3) の条件の下では $Q \times Q$ の正方行列であり，かつランクに関する仮定により非特異である。したがって逆行列 $\Gamma_k^{*\prime -1}$ が存在するから，これを (3—16) の右から掛けると，

$$\sum_{l=1}^{p} R_{kl} \Gamma_l^* \Gamma_l^* = \Gamma_k^* \Gamma_k^* L \qquad k = 1, \ldots, p \quad (5-4)$$

ここで，(3—15) を用いると，

$$\sum_{l=1}^{p} R_{kl} \Gamma_l^* \Gamma_l^* T^* = \Gamma_k^* \Gamma_k^* T^* \Theta$$

したがって，(3—18)，(3—19) により，

$$\sum_{l=1}^{p} R_{kl} \Gamma_l \Gamma_l = \Gamma_k \Gamma_k \Theta \qquad k = 1, \ldots, p \quad (5-5)$$

このことは，$[\Gamma_1'\Gamma_1' \quad \Gamma_2'\Gamma_2' \quad \cdots \quad \Gamma_p'\Gamma_p']'$ が，行列 R の固有ベクトルを各列とする行列であることを意味している。

ただもちろん，R の固有ベクトルを列とする行列，すなわち，

$$RQ = Q\Theta \quad (5-6)$$

を満足する行列 Q は，式 (4—26) の形をとり，

$$Q_k = \Gamma_k \Gamma_k \quad (5-7)$$

という分解を行う必要はある。この分解は，直交変換に関して不定性をもつが，(3—15)，(3—17)，(3—19) から，

$$\Gamma_k \Theta \Gamma_k' = \Lambda_k \quad (\Lambda_k \text{は対角行列}) \quad (5-8)$$

を条件として，一意の解を求めることができる。(ただし，回転を前提とすれば，あえてこの「回転前」の解を求める必要もないであろう。)

こうして，($q_1 = q_2 = \cdots = q_p$ の条件の下で，) 最も内的整合性を重視した解法と最も集合間相関を重視した解法は，ともに比較的よく用いられる方法と一致していたことがわかった。(したがって，これらの条件の下では，面倒な反復計算は実は不要であった。)

次に，一般化された正準相関分析に関して，(5—3) の条件の場合について調べてみよう。このとき，同様の過程を経て，そのアルゴリズムは，(4—25) の行列 P の固有値と固有ベクトルに帰着することが示される。すなわち，(4—26) の Q そのものが，最終的な解となる。(これが，4節の④のステップで，Q を初期値とする理由である。) この場合も，(5—7) の分解が必要になる。

なおこれは，Horst (1961) によって定式化され，Kettenring (1971) によって，MAXVAR と呼ばれた方法と一致する。またこれが，集合の数が2の場合に通常の正準相関分析と一致することは，Kettenring (1971) 参照。

また，4節の最後で述べた解釈にしたがえば，$Z_k^* = R_{kk}^{-\frac{1}{2}} Z_k$ を素データとみなして，全変数を主成分分析していることにもなる。

本研究では，集合ごとに正準変量の数 q_k が変化するように定式化されているが，このことは集合の数 p が3以上の場合には意味があると考えられる。実際，図1に示したようなケースにおいて，集合1の第1の正準変量は集合2の正準変量と，集合1の第2の正準変量は集合3の正準変量と高い相関をもつ，といった構造の場合には，集合2と3で正準変量を2つとる必要はない，と考えられるからである。

なお，一般化正準相関分析において $Q = \sum q_k$ の場合は，前述のように解は完全に不定であって意味がない。

また，集合の数が1の場合も，一般化正準相関分析には意味がないが，階層的主成分分析は，通常の主成分分析と一致する。更に，集合の数は複数で，各集合に含まれる変数の数がすべて1の場合，ここで述べた2つの方法はともに，主成分分析 (ただし，重みのノルムを1に基準化した定式化) となる。

次に，村上 (1986) の階層的主成分分析との関係について述べよう。そこでは，データ行列 Z_k を，

$$Z_k = A_k C_k G + E_k \qquad k = 1, \ldots, p \quad (5-9)$$

という形に分解することが目指されていた (ただし，原論文では，この式の G は F と書かれている)。最適化基準は，

$$S = \mathrm{tr} \sum_{k=1}^{p} \frac{1}{N} E_k E_k' \quad (5-10)$$

の最小化であり，制約条件は，

$$C_k C_k' = I \qquad k = 1, \ldots, p \quad (5-11)$$

$$\frac{1}{N}\boldsymbol{G}\boldsymbol{G}' = \boldsymbol{I} \qquad (5-12)$$

であった。

この方法のアルゴリズムは，3節において求められた Γ_k と Γ_k により，

$$\boldsymbol{A}_k = \Gamma_k \Lambda_k^{\frac{1}{2}} \boldsymbol{T}_k' \qquad k = 1, \ldots, p \ (5-13)$$

$$\boldsymbol{C}_k = \boldsymbol{T}_k \Lambda_k^{-\frac{1}{2}} \Gamma_k \Theta^{\frac{1}{2}} \boldsymbol{U} \quad k = 1, \ldots, p \ (5-14)$$

$$\boldsymbol{G} = \boldsymbol{U}' \Theta^{-\frac{1}{2}} \sum_{k=1}^{p} \Gamma_k' \Gamma_k' \boldsymbol{Z}_k \qquad k = 1, \ldots, p \ (5-15)$$

となる。すなわち，アルゴリズムの実質的な部分は，本研究の階層的主成分分析と同一である。村上（1986）は通常の因子分析モデルからの類推により，集合ごとの主成分を，

$$\boldsymbol{F}_k = \boldsymbol{C}_k \boldsymbol{G} \qquad (5-16)$$

と定義した。このとき，（5—13）による行列 $\boldsymbol{A}_k$ は，

$$\boldsymbol{A}_k = \frac{1}{N} \boldsymbol{Z}_k \boldsymbol{F}_k' \qquad k = 1, \ldots, p \ (5-17)$$

という意味をもつ。また，

$$\boldsymbol{C}_k \boldsymbol{C}_l' = \frac{1}{N} \boldsymbol{F}_k \boldsymbol{F}_l' \qquad k = 1, \ldots, p \quad l = 1, \ldots, p \ (5-18)$$

は，集合間の主成分間相関行列の意味をもつ。このように，このモデルにおけるパラメータは，本研究における集合内構造と集合間構造に対応する。ただ，（5—15）に見られるように，$\boldsymbol{F}_k$ の元になる $\boldsymbol{G}$ が，全集合の変数の一次結合として求められているために，（5—18）の相関係数は過大になりがちであることが，1節で述べたこのモデルの問題点である。

最後に，村上（1987 a,b）の「α 係数の和を最大化する階層的主成分分析」と本研究の方法との違いについて述べておこう。それは，制約条件（2—7）を，

$$\frac{1}{N}\boldsymbol{G}\boldsymbol{G}' = \sum_{k=1}^{p}\sum_{l=1}^{p} \boldsymbol{W}_k' \boldsymbol{V}_k' \boldsymbol{R}_{kl} \boldsymbol{V}_l \boldsymbol{W}_l = \boldsymbol{I} \qquad (5-19)$$

に変え，最適化基準を，

$$S = \mathrm{tr} \sum_{k=1}^{p} \boldsymbol{V}_k' \boldsymbol{W}_k' \boldsymbol{W}_k \boldsymbol{V}_k \qquad (5-20)$$

の最小化に変えたものである（ちょうど，最適化基準と制約条件が入れ代わった形になっている。この，一種の双対性の意味については，検討する価値があろう）。この方法のアルゴリズムは，本研究の方法とは，実質的に異なる。その詳細は村上（1987，b）を参照されたい。

1節においても述べたように，この方法は，（5—19）の基準化により，2次合成変量を解釈するためには都合がよいが，$Q = \sum_{k=1}^{p} q_k$ の場合が個別主成分分析と一致せず，ここでの定式化より他の方法との関連が希薄である。

6. 適用例

斎藤・村上・若林（1986）における，愛知県下の大学一年生592名のデータを分析する。ここでの変数は，

1）職業志向性に関する10の質問項目（5段階評定）

2）ライフ・スタイルに関する10の質問項目（5段階評定）

3）専攻（理科系を1，文科系を2とコーディングした2値変数）

4）性別（男子を1，女子を2とコーディング）

5）東海地方への将来の居住希望の程度（5段階評定）

表1　各条件における固有値の和

	HCA			GCAN	
	$Q=2$	$Q=4$	$Q=9$	$Q=2$	$Q=4$
1	4.57	4.55	4.51	1.98	1.93
2	**7.86**	7.81	7.61	**3.54**	3.43
3	9.80	9.93	9.77	4.71	4.76
4	11.40	**11.74**	11.58	5.74	**5.88**
5	12.75	13.18	13.19	6.64	6.70
6	13.75	14.17	14.38	7.41	7.50
7	14.69	15.12	15.36	8.03	8.05
8	15.44	15.87	16.29	8.56	8.57
9	16.00	16.49	**16.91**	9.00	9.00

6）8つの企業の就職先としての魅力度（5段階評定）という6つの集合からなっている。したがって，$p=6$，$n_1=n_2=10$，$n_3=n_4=n_5=1$，$n_6=8$ となる。各集合から導かれる1次合成変量の数は，$q_1=q_2$ $q_6=2$ とし，2次合成変量の数 Q をさまざまに変化させた（$q_3=q_4=q_5=1$ とする。集合3～5は単一の項目からなるから，これ以外にはあり得ない）。ここではこのうちから，階層的主成分分析（以下 HCA と略記する）の $Q=2$，4，9の場合，および一般化された正準相関分析（以下 $GCAN$ と略記）の $Q=2$，4の場合の結果を示す。

表1は，各条件における，行列 Ψ の固有値を大小順にならべ，累積した和を示したものである。ゴチックで示した，当該の Q の値のところが，最大化の対象となった値（Θ のトレース）であるが，当然他の条件より大きくなっている。

表2　職業志向性（集合1）の構造行列
（$q_1=q_2=q_6=2$ の場合）

項　　目	HCA $Q=2$		HCA $Q=4$		HCA $Q=9$		$GCAN$ $Q=2$		$GCAN$ $Q=4$	
	I	II	I	II	I	II	I	II	I	II
1. 安定した会社や勤め先	02	**70**	02	**68**	01	**64**	01	**83**	-30	**70**
2. 困難な仕事に挑戦する機会	**75**	-23	**74**	-23	**75**	-22	**67**	-29	**69**	-18
3. 高い給与やボーナス	09	**78**	09	**77**	09	**75**	06	**79**	14	**91**
4. 仕事上の責任の重さ	**74**	-17	**74**	-18	**74**	-19	**47**	-03	**40**	-03
5. 休日多い・勤務時間短い	01	**74**	01	**75**	01	**77**	-18	37	21	**62**
6. 能力が試される機会	**81**	09	**81**	09	**81**	10	**72**	02	**79**	12
7. 仕事の気楽さ	-12	**62**	-12	**64**	-12	**68**	**52**	14	-04	**42**
8. 仕事を通じ勉強・成長	**71**	12	**71**	13	**71**	13	30	01	**45**	09
9. 通勤の便利さ	01	**64**	01	**64**	01	**66**	-06	**47**	04	**57**
10. 自力で成し遂げる機会	**84**	14	**84**	15	**83**	15	**73**	01	**72**	08
2　乗　和	298	256	298	257	298	258	211	178	213	226
α　係　数	74	67	74	68	74	68	17	24	18	46

（小数点省略）

表3　ライフ・スタイル（集合2）の構造行列
（$q_1=q_2=q_6=2$ の場合）

項　　目	HCA $Q=2$		HCA $Q=4$		HCA $Q=9$		$GCAN$ $Q=2$		$GCAN$ $Q=4$	
	I	II	I	II	I	II	I	II	I	II
1. お金持ちになること	**73**	04	**75**	02	**88**	-10	**68**	19	**65**	**44**
2. 仕事で成功すること	37	**57**	38	**56**	**66**	22	17	**72**	29	**67**
3. 自分にあった仕事	12	**67**	12	**66**	30	**41**	-02	**48**	02	**54**
4. 親友をもつこと	25	**64**	24	**66**	12	**76**	19	**64**	32	**40**
5. 安定した仕事につく	**77**	18	**77**	19	**54**	**50**	**81**	30	**85**	-08
6. 大きな組織で高い地位	**75**	10	**76**	10	**75**	15	**52**	**44**	**59**	32
7. 好きなことをする暇	**41**	18	**41**	18	**40**	18	32	-03	28	**49**
8. よい結婚相手・幸福な家庭	**47**	**43**	**46**	**45**	25	**67**	35	**60**	**49**	29
9. 世の中の不平等なくす	00	**64**	-01	**64**	-09	**69**	12	16	05	26
10. 両親や親類のそばに住む	35	23	35	24	18	**45**	34	19	38	11
2　乗　和	244	190	242	193	241	223	181	190	215	160
α　係　数	48	28	48	30	57	52	-08	-02	21	-11

（小数点省略）

表4 就職先としての魅力度（集合6）の構造行列
（$q_1 = q_2 = q_6 = 2$の場合）

項　　目	*HCA* $Q=2$		*HCA* $Q=4$		*HCA* $Q=9$		*GCAN* $Q=2$		*GCAN* $Q=4$	
	I	II	I	II	I	II	I	II	I	II
1. 愛知県庁	−05	**81**	−03	**69**	−03	**68**	09	**75**	−01	**66**
2. 中部電力	**57**	−13	**50**	28	**48**	31	**45**	00	06	−11
3. ＣＢＣ	**43**	−17	16	**43**	15	**45**	**40**	07	**55**	06
4. 東海銀行	18	**64**	00	**74**	−01	**74**	26	**58**	07	**50**
5. トヨタ	**64**	15	**71**	07	**72**	07	**77**	14	**−53**	−08
6. 日本電装	**68**	−12	**83**	−11	**83**	−10	**47**	−30	−24	**−56**
7. ブラザー	**69**	−06	**69**	11	**70**	10	**52**	05	33	−10
8. 松坂屋	12	**65**	12	**67**	12	**65**	−13	**62**	−08	**55**
2 乗 和	190	157	196	175	195	175	154	139	76	134
α 係 数	45	33	54	47	54	47	07	18	−43	16

（小数点省略）

次に，各集合の集合内構造について見てみよう。表2～4は，複数の変数を含む，集合1（職業志向性），集合2（ライフ・スタイル），集合6（組織の就職先としての魅力度）の構造行列を，それぞれ示したものである。

集合1では，どの条件でもほとんど同一の構造が現れている。すなわち，偶数番号の項目は第1の合成変量と，奇数番号の項目は第2の合成変量と高い相関をもっている。それぞれの合成変量は，斎藤らの言う，“挑戦志向”と“安定志向”と解される。

集合2では，条件によって結果が異なる。*HCA*における，$Q=2$，4の場合，および*GCAN*における$Q=2$の場合は，定性的にみれば同一の結果であって，項目1，5，6，7，8等が第1の合成変量と，項目2，3，4，8，9等が第2の合成変量と高い相関をもっている。それぞれの合成変量は，斎藤らの“体制型”と“自己実現型”と解釈される。*GCAN*で$Q=4$の場合もほぼこれに対応する。一方，*HCA*の$Q=9$の場合（前述のように，集合ごとの個別主成分分析に相当）は，項目1，2，5，6，7が第1の合成変量と，項目3，4，5，8，9が第2の合成変量と高く相関し，斎藤らの“会社・出世型”と“私的生活享楽型”に対応する。斎藤らはこのデータを村上（1985）の多集合因子分析で被験者の学年別に分析し，1，2年生のデータでは，“体制型”と“自己実現型”の因子を，3，4年生のデータでは，“会社・出世型”と“私的生活享楽型”の因子を見出している。ここでは，2次合成変量の数を変えることによって，同一のデータから，2通りの結果が得られたことになる。

以上2つの集合の項目は，その素性が異なっている。集合1は，若林・中村・斉藤（1986）において，31の項目の因子分析の結果にもとづき，その2つの因子に高く負荷する項目を5つずつとって作られた。一方，集合2に含まれる項目は，元来，個別に日米の大学生の意識を比較するために作られたものである。この相違が，集合1では条件によらず安定した集合内構造を，集合2では条件ごとに異なる構造を，それぞれ得ることになった理由であろう。

集合6では，*HCA*の結果はQの値にかかわりなくほぼ同一であり，項目2，5，6，7が第1の合成変量と，項目1，4，8が第2の合成変量と高く相関している。ただし，$Q=2$の場合，項目3が第1の合成変量と相関が高いのに対して，$Q=4$，9の場合には第2の合成変量と相関が高くなる，という違いはある。第1の合成変量は，斎藤らの“工業技術性”に，第2の合成変量は，“お役所性”と“消費生活性”の混合であると解される。*GCAN*の$Q=2$では，*HCA*の$Q=2$の場合と定性的にはほぼ一致した結果が得られる。しかし，$Q=4$の場合，第2の合成変量は，ほぼ$Q=2$の場合と同一であるものの，第1の合成変量は，項目3と正の項目5と負の相関をもつ以外，明確な構造を示していない。

次に，集合間構造について見ることにしよう。表5～9に，各条件における1次合成変量間の相関を示した。幾つかの角度からこの5つの表を比較して見よう。

まず，集合1と集合2のそれぞれ2つずつの合成変量間の関係について見てみよう。どの条件でも，集合1の第2と集合1の第1の合成変量間の相関は高い。これは，それぞれの合成変量と高い相関をもつ集合1の項目1と

表5　1次合成変量間相関行列（HCA，$Q=2$）

		1	2	3	4	5	6	7	8	9
1. 職業志向性	Ⅰ	100	00	−18	33	−09	09	09	07	−13
2.	Ⅱ	00	100	49	16	16	−01	04	04	18
3. ライフ・	Ⅰ	−18	49	100	00	14	−06	13	11	23
スタイル	Ⅱ	33	16	00	100	02	−03	02	08	−04
5. 理系(1)—文系(2)		−09	16	14	02	100	13	−05	−18	40
6. 男(1)—女(2)		09	−01	−06	−03	13	100	01	03	00
7. 東海地方居住希望		−09	04	13	02	−05	01	100	10	07
8. 就職先として	Ⅰ	07	04	11	08	−18	03	10	100	00
9. の魅力度	Ⅱ	−13	18	23	−04	40	00	07	00	100

（小数点省略）

表6　1次合成変量間相関行列（HCA，$Q=4$）

		1	2	3	4	5	6	7	8	9
1. 職業志向性	Ⅰ	100	00	−18	33	−09	09	−09	04	−07
2.	Ⅱ	00	100	49	16	16	−01	04	02	16
3. ライフ・	Ⅰ	−18	49	100	00	14	−06	12	10	20
4. スタイル	Ⅱ	33	16	00	100	02	−03	02	07	−02
5. 理系(1)—文系(2)		−09	16	14	02	100	13	−05	−23	38
6. 男(1)—女(2)		09	−01	−06	−03	13	100	01	−02	05
7. 東海地方居住希望		−09	04	12	02	−05	01	100	10	08
8. 就職先として	Ⅰ	04	02	10	07	−23	−02	10	100	00
9. の魅力度	Ⅱ	−07	16	20	−02	38	05	08	00	100

（小数点省略）

表7　1次合成変量間相関行列（HCA，$Q=9$）

		1	2	3	4	5	6	7	8	9
1. 職業志向性	Ⅰ	100	00	−05	14	−09	09	−09	03	−07
2.	Ⅱ	00	100	46	18	16	−02	03	01	15
3. ライフ・	Ⅰ	−05	46	100	00	13	−06	03	07	14
4. スタイル	Ⅱ	14	18	00	100	04	−04	13	10	06
5. 理系(1)—文系(2)		−09	16	13	04	100	13	−05	−23	37
6. 男(1)—女(2)		09	−02	−06	−04	13	100	01	−02	05
7. 東海地方居住希望		−09	03	03	13	−05	01	100	10	08
8. 就職先として	Ⅰ	03	01	07	10	−23	−02	10	100	00
9. の魅力度	Ⅱ	−07	15	14	06	37	05	08	00	100

（小数点省略）

表8　1次合成変量間相関行列（$GCAN$，$Q=2$）

		1	2	3	4	5	6	7	8	9
1. 職業志向性	Ⅰ	100	00	−27	26	−14	04	−09	11	−17
2.	Ⅱ	00	100	45	31	16	01	11	17	23
3. ライフ・	Ⅰ	−27	45	100	00	18	−01	16	12	26
4. スタイル	Ⅱ	26	31	00	100	−04	−11	03	15	−00
5. 理系(1)—文系(2)		−14	16	18	−04	100	13	−05	−14	43
6. 男(1)—女(2)		04	01	−01	−11	13	100	01	02	00
7. 東海地方居住希望		−09	11	16	03	−05	01	100	06	03
8. 就職先として	Ⅰ	11	17	12	15	−14	02	06	100	00
9. の魅力度	Ⅱ	−17	23	26	−00	43	00	03	00	100

（小数点省略）

表9　1次合成変量間相関行列（$GCAN$，$Q=4$）

		1	2	3	4	5	6	7	8	9
1. 職業志向性	Ⅰ	100	00	−28	34	−10	−00	−17	04	−18
2.	Ⅱ	00	100	48	30	17	−02	06	−02	17
3. ライフ・	Ⅰ	−28	48	100	00	15	−05	17	−11	19
4. スタイル	Ⅱ	34	30	00	100	03	−02	−06	10	−05
5. 理系(1)—文系(2)		−10	17	15	03	100	13	−05	05	46
6. 男(1)—女(2)		−00	−02	−05	−02	13	100	01	19	02
7. 東海地方居住希望		−17	06	17	−06	−05	01	100	−00	−01
8. 就職先として	Ⅰ	04	−02	−11	10	05	19	−00	100	00
9. の魅力度	Ⅱ	−18	17	19	−05	46	02	−01	00	100

（小数点省略）

集合2の項目5が，ほとんど同一のものであることから，ある意味で当然であると見なされよう。（この場合は，まったく異なる素性をもつ異なる集合に，たまたま同義の項目が紛れ込んでいたわけである。これは，本来想定された心理学的属性の定義に問題があると考えられよう。しかし，多くの多集合データは，このように別のソースからの項目群から作られることが多く，このような重複の発見のために，むしろ多集合データの同時的分析が必要であると言えるかもしれない。）一方，集合1の第1の合成変量と，集合2の第1の合成変量の間の相関は，HCAではQの小さい場合に，また$GCAN$では常に，比較的高いのに対し，HCAで$Q=9$の場合にはほぼ無相関である。すなわち，“挑戦志向”は“自己実現”とはある程度相関するが，“私的生活享楽”的なものとは相関がない，ということである。このように表現すると，まったく当たり前にみえるが，Qが小さい場合と$GCAN$では，集合間構造が強調される結果として，1次合成変量の質が変化したわけで，興味深い結果と言えよう。それぞれの第1の合成変量どうしは，どの条件でも逆相関であるが，これも，HCAの$Q=9$の場合は無相関と言ってよく，$MAXVAR$では相当に目立つ。合成変量の命名にもとづいて表現すれば，“挑戦志向”は“体制型”であることとは逆相関だが，“会社・出世型”であることとは無相関ということになる。第2の合成変量どうしの相関は，$GCAN$においてのみ高い。これは，集合1で項目7（「仕事の気楽さ」）の相関が低下し，集合2では項目9（「世の中の不平等なくす」）のかわりに項目6（「大きな組織で高い地位」）の相関が高くなっており，どちらの集合でも“経済性”の要素が強まっているためであろう。

次に，集合6と他の集合との関係を見てみよう。集合1の第2，集合2の第1の合成変量が，集合6の第2の合成変量と比較的目だった相関を示しているが，“安定性”と相関するのは，集合6の第2合成変量の“お役所性”の成分の相関であろう。この結果はすべての条件で一貫してみられるが，相関の値は *HCA* で *Q* が小さい場合と，*GCAN* の場合に高い。*HCA* では，集合1と2の合成変量で，集合6の第1の合成変量と高い相関を示すものはないが，*GCAN* で $Q=2$ の場合には，それぞれの第2の合成変量がいくらか高い相関を示している。ここでとりあげられた企業はいずれも大企業であり，*GCAN* の集合1と集合2の第2の合成変量がともに“経済性”の要素をもつことから，この結果は理解しうるものである。

集合3の専攻は，どの条件でも就職先としての魅力度を予測する最強の要因である。集合4の性別は，ほとんど無相関だが，*GCAN* の $Q=4$ の場合のみ，第1の合成変量と目だった相関がある。この条件における第1の合成変量は単独で見ると解釈が困難であったが，男女の魅力度の差を反映するものであったと考えると理解できる。前述のように，*GCAN* で *Q* を増加しすぎると，解釈が困難になることが多いと思われるが，この条件ではたまたまこの部分が強調されることになった。

以上の結果は，データの中の異なった側面を見せている点で，それぞれ固有の特徴を示しているが，ここでは2つの点に特に注目しておきたい。第一に，*HCA* と *GCAN* は別系統の方法であるにもかかわらず，このデータでは，*HCA* において *Q* を小さくとった場合の結果が，*Q* を最大にした場合（個別主成分分析）よりも *GCAN* に近くなっていることである。このことは，*HCA* における *Q* の値の増減の効果がかなり強力であることを物語るものであると言えよう。第2に，このデータに関する限り，基準の著しい変化にもかかわらず，結果は全体にはかなりよく似たものとなっている点である。これは恐らく各集合の変数の数が比較的少ないためであろう。

文　　献

Cliff, N. & Krus, D. J. 1976 Interpretation of canonical analysis : Rotated vs. unrotated solutions. *Psychometrika,* **41**, 35-42.

Cronbach, L. J. 1970 *Essentials of Psychological Testing. 3rd ed.* Harper and Row.

Horst, P. 1961 Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology,* **14**, 331-347.

Johnson, R. M. 1966 The minimal transformation to orthonormality. *Psychometrika,* **31**, 61-66.

Kettenring, J. R. 1971 Canonical analysis of several sets of variables. *Biometrika,* **58**, 433-451.

村上　隆　1985　多集合データのための探索的因子分析　名古屋大学教育学部紀要――教育心理学科――，**32**，75-93.

村上　隆　1986　多集合データのための階層的主成分分析　名古屋大学教育学部紀要――教育心理学科――，**33**，75-93.

村上　隆　1987 a　2次合成変量の α 係数の和を最大化する多集合データの階層的主成分分析　I　経営行動科学，**2**，37-47.

村上　隆　1987 b　2次合成変量の α 係数の和を最大化する多集合データの階層的主成分分析　II　経営行動科学，**2**，77-88.

斎藤和志・村上　隆・若林　満　1986　組織の就職先としての魅力度と職業志向性及びライフ・スタイルとの関連　経営行動科学，**1**，101-113.

若林　満・中村雅彦・斎藤和志　1986　就職先としての組織の魅力と現代学生の職業志向　経営行動科学，**1**，11-25.

（1987年8月31日　受稿）

ABSTRACT

PRINCIPAL COMPONENTS AND CANONICAL VARIATES OF SEVERAL SETS OF VARIABLES

Takashi MURAKAMI

Many of psychological correlational studies using questionnaires or tests generally consist of two steps. In the first step, some composite variates as linear combinations of a single set of variables are constructed by means of factor analytical procedures. Each composite is interpreted as a realization of an unobservable psychological construct. In the second step, correlations between composites from different variable sets are obtained. From these correlations, one wishes to establish the relationships between several psychological constructs under consideration. However, the factor analytical methods applied to a single variable set produce internally consistent composites, but they do not account for the relationships of the composites with variables belonging to other variable sets at all.

The situation which we are confronted with is characterized by the bandwidth-fidelity dilemma in Cronbach's sense. Factor analysis of a single variable set overemphasizes the fidelity. The canonical correlation analysis, on the other hand, produces the composites with the highest correlations between sets, but they are often very unreliable, and are inappropriate measures of constructs. The canonical method overemphasizes the bandwidth. Because main objective of correlational studies may be finding interesting relationships between reliable measures, it is desirable to develop procedures through which one can balance the fidelity of the composites with the bandwidth of them.

In this paper two methods which analyze several sets of variables simultaneously are proposed. They can be considered to be a natural extension of the principal component analysis and the canonical correlation analysis, respectively. Both of them define the first-order composites as linear combinations of variables of each set, and the second order composites as linear combinations of all the first order composites. Let Z_k be a N by n_k row-wise standardized data matrix for k-th set, then the matrix of first-order composites for k-th set is defined as

$$F_k = V_k' Z_k, \tag{1}$$

and that of second order composites is defined as

$$G = \sum_k W_k' F_k, \tag{2}$$

where V_k and W_k are n_k by q_k and q_k by Q weight matrices, respectively. Both methods share the optimization criterion which is the maximization of tr GG'/N under the constraint,

$$F_k F_k'/N = I. \tag{3}$$

An additional constriant differs in two methods. That is,

$$\sum_k W_k' V_k' V_k W_k = I \tag{4}$$

is imposed for the first one, *hierarchical principal component analysis,* and,

$$\sum_k W_k' W_k = I \tag{5}$$

for the second, *generalized canonical correlation analysis.* Alternating algorithms for two methods are formulated.

It is expected that the property of the first order composites is changed through altering the number of second order composites. More concretely, the most internally consistent solution can be obtained when Q is set to be equal to Σq_k in hierarchical component analysis, which is equal to that of the separate principal component analysis for each set, and the most highly correlated solution is given by generalized canonical analysis when $Q = \max q_k$. Generally speaking, as Q is set to be large, correlations between composites of different sets are expected to be increased, and the fidelity of first order composites is improved as Q is smaller. Therefore, through changing the Q, one can balance the fidelity and the bandwidth of first order composites.

Application to the real data revealed that the balancing mechanism of these methods works as was expected.