ON A COMPUTATION METHOD FOR EIGENVALUE PROBLEMS AND ITS APPLICATION TO STATISTICAL ANALYSIS

By Hirotugu Akaike

(Received Aug. 20, 1958)

0. Introduction

In this paper we treat the numerical method for solving the N-dimensional eigenvalue problem $Ax=\lambda Bx$ where B is an $N\times N$ positive definite real symmetric matrix, A is an $N\times N$ real symmetric matrix and x is an N-dimensional column vector. Our main object is to propose a definite computation scheme which will practically allow us to get a complete solution of the problem just stated above even for large N.

When we adopt the linear discriminant function method for the statistical classification of objects into three or more groups or more generally when we apply the canonical correlation method to some multidimensional statistical phenomena we must solve numerically the equation of the type just mentioned above. Nevertheless we have been unable to get complete description of the numerical method for solving practically these equations with large N [4] [5]. Some computation method which does not necessarily use B^{-1} will be most desirable [6].

C. Hayashi at our institute has long been trying to apply the linear discriminant function method to the so-called quantification problem aiming at objectifing the scoring procedure of objects for their qualitative characters. In this case the problem is formulated in such a way that we seek the vector x which gives the largest value to the ratio $(x, Ax)/(x, Bx)$ [1]. This formulation of the problem and the fact that we usually encountered with the case where N is rather big invited the author in 1952 heuristically to a successive approximation method and the method was used by Hayashi and others in many practical applications of the "quantification" procedures with satisfactory results [2] [3]. Nevertheless, we were not able to give the complete theoretical description to the convergence properties of our solution.

In 1957 S. Huzino of Kyushu University announced some results on the same type problem at the annual meeting of Japan Mathematical Society. He has presented his results to the recent paper [7], but it seems that in his results there remains some difficulty to apply them to practical computation.

In this paper we start from general discussion of a type of successive approximation method for $Ax=\lambda Bx$ in §2 and then give some specific procedures in §§3 and 4. In §5 we present a definite computation scheme. We can observe in our computation procedure some interesting interplay between theoretical model and its finite digital representation and we discuss an acceleration scheme in §6. Also in §6 our approximation procedure is discussed about its convergence property against rounding off errors. In §7 convergence to the eigenvector with maximum eigenvalue is discussed. In practical applications it is sometimes necessary to get only such eigenvector. In §8 some generalization of our process is discussed and a computation scheme is proposed which will be suited for high speed computers. In §9 numerical examples for our procedure are given. Most of these results are obtained by a FACOM-128 automatic relay computer at our institute. Though their dimensions are too low to show the feature of our method these examples will serve to explain how our method works. Technical details of the computation scheme on FACOM-128 automatic relay computer will be presented to the future issue of the Proceedings of our institute. Statistical applications are also discussed very briefly in §10 which are mainly concerned with the canonical correlation analysis. We hope these discussions will be of some help to those who want to apply the canonical correlation method to practical problems and at the same time will clarify the importance in statistical analysis of the numerical method for the solution of $Ax=\lambda Bx$.

1. Notations

By $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_N$ and $x_1, x_2, \cdots, x_N$ we represent the eigenvalues and their corresponding eigenvectors. It is well known that in our problem every eigenvalue is real. We can further assume that x_i's are real and B-orthonormal i.e. $(x_i, Bx_j)=\delta_{ij}$ where $\delta_{ii}=1$, $\delta_{ij}=0$ $(i\neq j)$.

We shall use the notation $||x||_M^2$ for (x, Mx) and $(x, y)_M$ for (x, My)

where M is a real matrix and x, y are vectors. We omit the suffix M when M is equal to identity matrix. Of course (x, y) means the ordinary inner product of vectors x and y. We use the symbol $'$ to show the transposed matrix and $|M|$ to show the determinant of matrix M.

2. General approximation procedure

In this section we shall discuss the approximation procedure defined as follows:

0. We define how to construct a set of vectors $(\xi_1(x), \xi_2(x), \cdots, \xi_k(x))$ where $\xi_i(x)$'s are continuous functions of vector x and the vectors $x, \xi_1(x), \xi_2(x), \cdots, \xi_k(x)$ are supposed to be linearly independent N-dimensional vectors satisfying the following inequality except the case where x coincides with one of the eigenvectors of $Ax=\lambda Bx$:

$$\operatorname*{Max}_{(\alpha,\alpha_1\cdots\alpha_k)} \frac{\|\alpha x+\sum_{i=1}^{k}\alpha_i\xi_i(x)\|_A^2}{\|\alpha x+\sum_{i=1}^{k}\alpha_i\xi_i(x)\|_B^2} > \frac{\|x\|_A^2}{\|x\|_B^2} .$$

When x is an eigenvector above $>$ must be replaced by $\geqq$. We shall represent by $\alpha(x), \alpha_1(x), \cdots, \alpha_k(x)$ the coefficients which give the maximum value to the above stated ratio and satisfy the equation $\|\alpha x+\sum_{i=1}^{k}\alpha_i\xi_i(x)\|_B^2=1$.

1. Given a $(n-1)$-th approximation vector $x^{(n-1)}$ we compute $(\xi_1(x^{(n-1)}), \xi_2(x^{(n-1)}), \cdots, \xi_k(x^{(n-1)}))$.

2. We find the coefficients $(\alpha(x^{(n-1)}), \alpha_1(x^{(n-1)}), \cdots, \alpha_k(x^{(n-1)})$ and put $x^{(n)}=\alpha(x^{(n-1)})x^{(n-1)}+\sum_{i=1}^{k}\alpha_i(x^{(n-1)})\xi_i(x^{(n-1)})$.

3. Starting from some $x^{(0)}$, we repeat the steps 1. and 2. cyclically to get the sequence $\{x^{(n)}\}$.

It is easy to see that the above stated approximation procedure gives a sequence $\{x^{(n)}\}$ convergent to one of the eigenvectors when every eigenvalue has multiplicity unity and gives generally a sequence $\{x^{(n)}\}$ "convergent" to a space spanned by eigenvectors which correspond to one and the same eigenvalue.

The condition $\|x^{(n)}\|_B=1$ is not always necessary in practical computation.

To make the above stated procedure practicable it is necessary and sufficient to give the proper difinition for $\xi_i(x)$'s and a computing scheme for the coefficients $\alpha(x), \alpha_1(x), \cdots, \alpha_k(x)$. This set of coefficients is given as an eigenvector, which corresponds to the maximum eigenvalue $\eta_{\max}$, of the equation $A(x)a=\eta B(x)a$ where

$$a=\begin{pmatrix}\alpha\\ \alpha_1\\ \vdots\\ \alpha_k\end{pmatrix} \qquad A(x)=\begin{pmatrix}(x,x)_A & (x,\xi_1(x))_A & \cdots & (x,\xi_k(x))_A\\ (\xi_1(x),x)_A & (\xi_1(x),\xi_1(x))_A & \cdots & (\xi_1(x),\xi_k(x))_A\\ \vdots & \vdots & & \vdots\\ (\xi_k(x),x)_A & (\xi_k(x),\xi_1(x))_A & \cdots & (\xi_k(x),\xi_k(x))_A\end{pmatrix}$$

$$B(x)=\begin{pmatrix}(x,x)_B & (x,\xi_1(x))_B & \cdots & (x,\xi_k(x))_B\\ (\xi_1(x),x)_B & (\xi_1(x),\xi_1(x))_B & \cdots & (\xi_1(x),\xi_k(x))_B\\ \vdots & \vdots & & \vdots\\ (\xi_k(x),x)_B & (\xi_k(x),\xi_1(x))_B & \cdots & (\xi_k(x),\xi_k(x))_B\end{pmatrix}.$$

Thus to find the coefficients $(\alpha(x), \alpha_1(x) \cdots \alpha_k(x))$ seems to be essentially equivalent to solve the problem of the type for which we are just trying to provide a computation method. But our main concern is with the problem where dimension N is large and when we adopt a scheme with small k we are able to get the desired result by some simple computation, for example, when A is non-negative definite by taking the inverse of $B(x)$ and applying the power metheod to $B(x)^{-1}A(x)$*). When $k=1$ we can get an explicit representation for the coefficients, and we shall treat this case precisely in the following section. It should be noticed that in the formulation of our procedure we may take $\alpha\ \alpha_1 \cdots \alpha_k$ which minimize the ratio $\|\alpha x+\sum_{i=1}^{k}\alpha_i\xi_i(x)\|_A^2/\|\alpha x+\sum_{i=1}^{k}\alpha_i\xi_i(x)\|_B^2$. Such formulation is sometimes more desirable and we have only to change the sign before the square root in the following sections to get it.

3. Approximation procedure with $k=1$

In this section we shall treat the approximation procedure of the former section with $k=1$. We shall use the notation ξ for $\xi_1(x)$. We have the following results:

$$A(x)=\begin{pmatrix}\|x\|_A^2 & (x,\xi)_A\\ (\xi,x)_A & \|\xi\|_A^2\end{pmatrix}, \qquad B(x)=\begin{pmatrix}\|x\|_B^2 & (x,\xi)_B\\ (\xi,x)_B & \|\xi\|_B^2\end{pmatrix}$$

*) We take $(B(x)^{-1}A(x))^L a_0$ as a set of coefficients $(\alpha(x),\alpha^1(x)\cdots\alpha^{(k)}(x))'$ where a_0 is a properly chosen initial vector and L is some integer which assures practically sufficient accuracy as an approximation to the set of coefficients.

$$B(x)^{-1}A(x)=\frac{1}{\|x\|_B^2\|\xi\|_B^2\left(1-\left(\frac{(x,\xi)_B}{\|x\|_B\|\xi\|_B}\right)^2\right)}$$
$$\times\begin{pmatrix}\|\xi\|_B^2\|x\|_A^2-(x,\xi)_B(x,\xi)_A & \|\xi\|_B^2(x,\xi)_A-\|\xi\|_A^2(x,\xi)_B \\ \|x\|_B^2(x,\xi)_A-\|x\|_A^2(x,\xi)_B & \|x\|_B^2\|\xi\|_A^2-(x,\xi)_B(x,\xi)_A\end{pmatrix}.$$

Thus the maximum eigenvalue $\eta_{\max}$ for the equation $A(x)a=\eta B(x)a$ is given by solving the equation

$$|B(x)^{-1}A(x)-\eta I|=0\ ,$$

and we have

$$\eta_{\max}=\frac{\|\xi\|_B^2\|x\|_A^2+\|\xi\|_A^2\|x\|_B^2-2(x,\xi)_A(x,\xi)_B+\sqrt{D}}{2\|x\|_B^2\|\xi\|_B^2\left(1-\left(\frac{(x,\xi)_B}{\|x\|_B\|\xi\|_B}\right)^2\right)}$$

where

$$\begin{aligned}D=&(\|\xi\|_B^2\|x\|_A^2+\|\xi\|_A^2\|x\|_B^2-2(x,\xi)_A(x,\xi)_B)^2\\&-4(\|\xi\|_B^2\|x\|_A^2-(x,\xi)_B(x,\xi)_A)(\|x\|_B^2\|\xi\|_A^2-(x,\xi)_B(x,\xi)_A)\\&+4(\|x\|_B^2(x,\xi)_A-\|x\|_A^2(x,\xi)_B)(\|\xi\|_B^2(x,\xi)_A-\|\xi\|_A^2(x,\xi)_B)\\=&(\|\xi\|_B^2\|x\|_A^2-\|\xi\|_A^2\|x\|_B^2)^2\\&+4(\|x\|_B^2(x,\xi)_A-\|x\|_A^2(x,\xi)_B)(\|\xi\|_B^2(x,\xi)_A-\|\xi\|_A^2(x,\xi)_B)\ .\end{aligned}$$

Eigenvector corresponding to $\eta_{\max}$ must satisfy the equations

$$\begin{aligned}&(\|\xi\|_B^2\|x\|_A^2-(x,\xi)_B(x,\xi)_A-(\|x\|_B^2\|\xi\|_B^2-(x,\xi)_B^2)\eta_{\max})\alpha(x)\\&\quad+(\|\xi\|_B^2(x,\xi)_A-\|\xi\|_A^2(x,\xi)_B)\alpha_1(x)=0\\&(\|x\|_B^2(x,\xi)_A-\|x\|_A^2(x,\xi)_B)\alpha(x)\\&+(\|x\|_B^2\|\xi\|_A^2-(x,\xi)_B(x,\xi)_A-(\|x\|_B^2\|\xi\|_B^2-(x,\xi)_B^2)\eta_{\max})\alpha_1(x)=0\ .\end{aligned}$$

From these equations or more directly from the relation

$$\begin{aligned}\frac{\partial}{\partial\lambda}\mu(x+\lambda\xi)\Big|_{\lambda=0}&=\frac{2}{\|x\|_B^4}(\|x\|_B^2(x,\xi)_A-\|x\|_A^2(x,\xi)_B)\\&=\frac{2}{\|x\|_B^2}(Ax-\mu(x)Bx,\xi)\ ,^{*)}\end{aligned}$$

we can see that if ξ satisfies at least one of the two conditions $\mu(\xi)>\mu(x)$ and $(Ax-\mu(x)B,\xi)\neq 0$ we can expect $\eta_{\max}>\mu(x)$.

When $(Ax-\mu(x)Bx,\xi)=0$ and $\mu(\xi)\leqq\mu(x)$ hold we have by a simple calculation $\eta_{\max}=\mu(x)$, and in this case we can not expect the increase of μ by taking a linear combination of x and ξ.

---

*) We shall use the notation $\mu(x)$ for the ratio $\dfrac{\|x\|_A^2}{\|x\|_B^2}$.

Thus to assure $\eta_{\max} > \mu(x)$ it is necessary and sufficient for ξ to satisfy at least one of the two conditions $\mu(\xi) > \mu(x)$ and $(Ax - \mu(x)Bx, \xi) \neq 0$. It is also obvious that any such ξ is linearly independent of x. To find directly a ξ which satisfies the condition $\mu(\xi) > \mu(x)$ is rather difficult, but it is easy to get a ξ which satisfies the condition $(Ax - \mu(x)Bx, \xi) \neq 0$, except for x which is identical with some of the eigenvectors.

When x is an eigenvector we have $Ax - \mu(x)Bx = 0$ and $(Ax - \mu(x)Bx, \xi) = 0$ for any ξ and it is difficult in this case to get a ξ which assures $\eta_{\max} > \mu(x)$ for $\mu(x)$ which is different from the minimum eigenvalue.

From these observations it is now easy to give some specific examples of ξ satisfying the postulates for the correction factor given in §2. We shall give some examples of ξ in the next section.

The following relations should be noticed: for x in a subspace spanned by some subset of eigenvectors $(x_{\nu_1}, x_{\nu_2}, \cdots, x_{\nu_m})$ we have $(Ax, \xi) = (Ax, \xi')$, $(Bx, \xi) = (Bx, \xi')$ and $(Ax - \mu(x)Bx, \xi) = (Ax - \mu(x)Bx, \xi')$ where

$$\xi' = \xi - \sum_{i=1}^{m} (\xi, x_{\nu_i})_B x_{\nu_i}.$$

From these relations it becomes clear that we have only to define ξ as a function of x satisfying the inequality $(Ax - \mu(x)Bx, \xi) \neq 0$ and to use ξ' as a correction factor when we restrict our attension to some subspace as described above.

4. Some specific ξ's

1. $\xi = P^{-1}(Ax - \mu(x)Bx)$

P is a positive definite matrix. This ξ is characterized by the property that it maximizes $\left.\frac{\partial}{\partial \lambda}\mu(x + \lambda\xi)\right|_{\lambda=0}$ under the restriction $\|\xi\|_P^2 = 1$. We have $(Ax - \mu(x)Bx, \xi) = \|Ax - \mu(x)Bx\|_{P^{-1}}^2$. It is obvious that this ξ satisfies all the postulates for the correction factor.

2. $\xi = Ax - m(x)Bx$

$m(x)$ is defined as the ratio $(Ax, Bx)/(Bx, Bx)$. This ξ is characterized as a least square residual and satisfies the relations $(\xi, Bx) = (\xi, x)_B = 0$ $(Ax - \mu(x)Bx, \xi) = (Ax, \xi) = \|Ax\|^2\left(1 - \frac{(Ax, Bx)^2}{\|Ax\|^2\|Bx\|^2}\right)$. From the last equation it is obvious that except for x which satisfys the relation $Ax = \lambda Bx$ the inequality $(Ax - \mu(x)Bx, \xi) > 0$ holds. Continuity of ξ as a function of x is also obvious.

5. Computation scheme

By the observations made in §3 it is obvious that we can use the following computation scheme with ξ defined in §4. Here we suppose that some set of eigenvectors $(x_{\nu_1}, x_{\nu_2}, \cdots, x_{\nu_k})$ is already obtained.

1. Take a proper vector x and delete from it its components in $x_{\nu_1}, x_{\nu_2}, \cdots, x_{\nu_k}$ to get non zero $x^{(0)}$:

$$x^{(0)}=x-\sum_{i=1}^{k}\frac{(x, x_{\nu_i})_B}{\|x_{\nu_i}\|_B^2}x_{\nu_i} .$$

2. Calculate $\xi^{(0)}$*).
3. Delete from $\xi^{(0)}$ its components in $x_{\nu_1}, x_{\nu_2}, \cdots, x_{\nu_k}$ to get $\xi'^{(0)}$:

$$\xi'^{(0)}=\xi^{(0)}-\sum_{i=1}^{k}\frac{(\xi^{(0)}, x_{\nu_i})_B}{\|x_{\nu_i}\|_B^2}x_{\nu_i} .$$

4. Compute the ratio

$$Q_0=\frac{\{(x^{(0)}, \xi'^{(0)})_A-\mu(x^{(0)})(x^{(0)}, \xi'^{(0)})_B\}\{(x^{(0)}, \xi'^{(0)})_A-\mu(\xi'^{(0)})(x^{(0)}, \xi'^{(0)})_B\}}{(\mu(x^{(0)})-\mu(\xi'^{(0)}))^2\|x^{(0)}\|_B^2\|\xi'^{(0)}\|_B^2} .$$

5. When $|Q_0|$ is less than some preassigned positive small fixed value $\delta(\ll 1)$ we put

$$\lambda_{(0)}=\frac{(x^{(0)}, \xi'^{(0)})_A-\mu(x^{(0)})(x^{(0)}, \xi'^{(0)})_B}{|\mu(x^{(0)})-\mu(\xi'^{(0)})|\,\|\xi'^{(0)}\|_B^2} .$$

When $|Q_0|\geq\delta$ we put

$$\lambda_{(0)}=\frac{-(\mu(x^{(0)})-\mu(\xi'^{(0)}))\|x^{(0)}\|_B^2\|\xi'^{(0)}\|_B^2+\sqrt{D}}{2\|\xi'^{(0)}\|_B^2\{(x^{(0)}, \xi'^{(0)})_A-\mu(\xi'^{(0)})(x^{(0)}, \xi'^{(0)})_B\}}$$

where

$$D=(\mu(x^{(0)})-\mu(\xi'^{(0)}))^2\|x^{(0)}\|_B^4\|\xi'^{(0)}\|_B^4(1+4Q_0) .$$

6. Calculate $x^{(0)}+\lambda_{(0)}\xi'^{(0)}$ to get $x^{(1)}$.
7. Repeat the cycle after 2 taking $x^{(1)}$ instead of $x^{(0)}$ to get $x^{(2)}$, $x^{(2)}$ instead of $x^{(1)}$ to get $x^{(3)} \cdots$.
8 Make checks of rounding off errors by inserting step 1.
9. Continue the whole process untill $\frac{\|\xi^{(n)}\|^2}{\|x^{(n)}\|^2}$ becomes smaller than some preassigned small positive constant.

In actual computation we have to use finite digital representation of each quantity and as is shown in the next section our problem sometimes reduces to a linear one after some steps of approximation. This

*) In this section we shall use the abbreviated notation $\xi^{(n)}$ for $\xi(x^{(n)})$. Thus $\xi^{(0)}$ stands for $\xi(x^{(0)})$.

will afford us an interesting example which shows explicitly the fact that the theoretical expression of the problem is sometimes only a crude image of the actual computation procedure and when we can get at some stage of computation another theoretical model with difference less than some preassigned quantity (usually of the order of errors due to roundings which is inevitable in the computation procedure) from the original one it would be better to use the model which is more suited for the actual computation. We shall discuss this point in the next section.

6. Acceleration scheme

In the course of computations we have often observed the fact that when the approximate eigenvectors $x^{(n)}$ become steady up to their m-th digits the corresponding $\mu(x^{(n)})$'s become steady up to nearly $2m$-th digits. The following relation will explain the above stated fact: for vector $x=x_i+\varepsilon$ where $(x_i, \varepsilon)_B=0$

$$\mu(x_i)-\mu(x)=(\mu(x_i)-\mu(\varepsilon))\frac{\|\varepsilon\|_B^2}{\|x_i\|_B^2}\left(1+\frac{\|\varepsilon\|_B^2}{\|x_i\|_B^2}\right)^{-1} \text{ holds,}$$

and for ε with $\mu(\varepsilon)\leqq\mu(x_i)$ we have

$$\frac{\mu(x_i)-\mu(x)}{\mu(x_i)}<\frac{\|\varepsilon\|_B^2}{\|x_i\|_B^2}\,.$$

Thus it may sometimes happen that at x which has significant deviation even in its finite digital representation from x_i $\mu(x)$ takes a value whose difference from $\mu(x_i)$ is less than the order of error in the computation of $\mu(x)$ due to the roundings of x. In this way, when we use the same number of digits for all quantities, the following observations will explain that stage of computation where the degree of approximation is fairly high and $\mu(x^{(n)})$ stops its steady increase.

Hereafter we shall restrict ourselves to the case where the vector x is so near to x_1 that the difference $\lambda_1-\mu(x)$ is less than the order of errors due to roundings in the computation of $\mu(x)$ and $\lambda_1>\lambda_2$ holds. In this case if we use the representation $x=\alpha_1x_1+\varepsilon$ where $(x_1, \varepsilon)=0$ we have $C\varepsilon=Ax-\lambda_1Bx$ for $C=A-\lambda_1B$.

Thus taking into account the fact that we can use the value $\mu(x)$ for λ_1 our problem is reduced to solving the linear equation $(A-\mu(x)B)\varepsilon=(A-\mu(x)B)x$ for given x. Obviously the equation $C\varepsilon=Cx$ has a solution which is unique except for its x_1 component and we can apply the conjugate gradient method or some other successive approximation method to get the solution ε.

If we define $\xi=Ax-\mu(x)Bx$ the coefficient λ given in our approximation procedure by the formula $\frac{(x,\xi)_A-\mu(x)(x,\xi)_B}{\|\xi\|_B^2|\Delta|}=\frac{\|\xi\|^2}{\|\xi\|_B^2|\Delta|}$ is essentially the value of λ which minimize $-\|\varepsilon+\lambda C\varepsilon\|_C^2$. Thus if we continue our approximation procedure in this stage it is equivalent to solving the equation $C\varepsilon=\xi$ for a given ξ by a successive approximation method which mininize the residual in its $-C$-norm$(-\|\ \|_C^2)$ using the former residual as a linear correction factor. Obviously this approximation procedure is not always very efficient one and the most convenient procedure for solving the linear equation for given C and ξ with $|C|=0$ will be the most recommendable here. After we have solved the equation $(A-\mu(x)B)\varepsilon=\xi$ it would be better to follow our original approximation procedure using $x-\varepsilon$ as starting value until the rounding off errors dominate in the computation of $x^{(n)}$'s and $\xi^{(n)}$'s. In this case even after the $\xi=Ax-\mu(x)Bx$ become small and the value of λ given by $\frac{\|\xi\|^2}{\|\xi\|_B^2|\mu(x)-\mu(\xi)|}$ become unreliable by the roundings still we can expect the theoretical increase of $\mu(x)$ untill the ξ becomes comparable with the order of error due to roundings. This is shown by the facts that in this case we can expect $\mu(x)>\mu(\xi)$ and for any such ξ we can expect generally $\mu(x+\hat{\lambda}\xi)>\mu(x)$ for $\hat{\lambda}$ which satisfies either one of the inequalities $0\gtrless\hat{\lambda}\gtrless 2\frac{(Ax-\mu(x)Bx,\xi)}{\|\xi\|_B^2(\mu(x)-\mu(\xi))}$.

Thus even if we use the calculated value of $Ax-\mu(x)Bx$ as ξ so long as the value $\frac{\|\xi\|^2}{\|\xi\|_B^2(\mu(x)-\mu(\xi))}$ remains in the range stated above for $\hat{\lambda}$ we can expect the theoretical increase of $\mu(x)$. If we represent our ξ as a sum of $Ax-\mu(x)Bx$ and error term ε, the above stated condition for ξ is represented by the inequalities

$$(Ax-\mu(x)Bx,\ Ax-\mu(x)Bx+\varepsilon)>0$$

$$2(Ax-\mu(x)Bx,\ Ax-\mu(x)Bx+\varepsilon)\geqq\|Ax-\mu(x)Bx+\varepsilon\|^2\ .$$

From the latter inequality we get the condition

$$\|Ax-\mu(x)Bx\|^2\geqq\|\varepsilon\|^2$$

and it is clear that under this condition the former inequality also holds. This result means that untill the error ε (due to roundings) becomes comparable with the order of true residual $Ax-\mu(x)Bx$ our approximation procedure will work.

7. Convergence to x_1

In this section we consider the case $\lambda_1 > \lambda_2$.

There sometimes occurs the case where we have only to seek for the eigenvector x_1 which corresponds to the maximum eigenvalue λ_1. If we follow the computation scheme $x^{(n+1)} = x^{(n)} + \lambda_{(n)}\xi^{(n)}$ where $\xi = B^{-1}(Ax - \mu(x)Bx)$ and $\lambda_{(n)} = \frac{\|\xi^{(n)}\|_B^2}{|\mu(x^{(n)}) - \mu(\xi^{(n)})|\,\|\xi^{(n)}\|_B^2}$ we are sure to get x_1 as the limit of our sequence $\{x^{(n)}\}$ starting from arbitrary $x^{(0)} = \sum_{i=1}^{N}\alpha_i^{(0)}x_i$ with $\alpha_1^{(0)} = (x^{(0)}, x_1)_B \neq 0$. This is easily proved by using the following relations:

$$\mu(x+\lambda\xi) - \mu(x) = \frac{\lambda(2(Ax - \mu(x)Bx, \xi) - \lambda\|\xi\|_B^2(\mu(x) - \mu(\xi)))}{\|x+\lambda\xi\|_B^2},$$

$$B^{-1}(Ax - \mu(x)Bx) = \sum_{i=1}^{N}\alpha_i(\lambda_i - \mu(x))x_i .$$

We can see from the former equation that for our definition of $\lambda_{(n)}$ $\mu(x^{(n)} + \lambda_{(n)}\xi^{(n)}) > \mu(x^{(n)})$ holds for $x^{(n)}$ which is different from any one of the eigenvectors. If we put $x^{(n)} = \sum_{i=1}^{N}\alpha_i^{(n)}x_i$ we can also see from the latter equation that

$$\alpha_1^{(n)} : \alpha_2^{(n)} : \cdots : \alpha_N^{(n)} = \alpha_1^{(0)}\prod_{i=0}^{n-1}(1+\lambda_{(i)}(\lambda_1 - \mu(x^{(i)})) : \alpha_2^{(0)}\prod_{i=0}^{n-1}(1+\lambda_{(i)}(\lambda_2 - \mu(x^{(i)}))$$
$$: \cdots : \alpha_N^{(0)}\prod_{i=1}^{n-1}(1+\lambda_{(i)}(\lambda_N - \mu(x^{(i)}))$$

holds and taking into account of the fact that $\lambda_{(i)} > 0$ we can see that between those $\alpha_i^{(n-1)}$'s with λ_i greater than $\mu(x^{(n-1)})$ $\alpha_1^{(n-1)}$ has the greatest rate of increase and the convergence of $x^{(n)}$ to some x_i with $\lambda_i < \lambda_1$ is impossible. Thus we almost surely get x_1 as the first solution. For general ξ and computation scheme we cannot prove such a desirable property explicitly. In actual computation additional informations about the matrices A an B sometimes enable us to select $x^{(0)}$ which lies near x_1 and the limit of our $\{x^{(n)}\}$ is strongly supposed to be x_1. In such a case there still remains the check whether our solution really corresponds to the maximum eigenvalue λ_1 or not. For this check the following procedure may be of use:

1. Given a solution x we difine by D the difference $D = A - \mu(x)B$.

2. We take an arbitrary vector $z^{(0)}$ with $(z^{(0)}, x) = 0$ and compute $Dz^{(0)}$.

3. We compute $z^{(1)} = z^{(0)} - c(z^{(0)})Dz^{(0)}$ where $c(z) = \frac{(z, Dz)}{\|Dz\|^2}$.

4. We repeat the steps 2, 3 cyclically using $z^{(n)}$ to get $z^{(n+1)}$ successively.

5. We decide that x is the desired solution when the coefficients $c(z^{(n)})$ remain all negative until $\frac{\|z^{(n)}\|^2}{\|z\|^2}$ becomes very small.

The validity of above process is obvious from the fact that the matrix $-D$ is positive definite if $\mu(x)=\lambda_1$ and otherwise not.

8. Some generalization

Our computation procedure is mainly based on the monotone increasing property of $\mu(x^{(n)})$. Taking into account of this fact we can get many other modifications of our procedure. The following one will deserve attension.

We choose a set of linearly independent vectors $\xi_1, \xi_2, \cdots, \xi_N$ and apply the successive approximation procedure $x^{(kN+\nu+1)}=\alpha_{kN+\nu}x^{(kN+\nu)}+\alpha^{(1)}{}_{kN+\nu}\xi_{\nu+1}$ where $0\leqq\nu<N$ and $\alpha_{kN+\nu}, \alpha^{(1)}{}_{kN+\nu}$ are α, α^1 which give the maximum value to $\mu(\alpha x^{(kN+\nu)}+\alpha^1\xi_{\nu+1})$ and satisfy $\|x^{(kN+\nu+1)}\|^2=1$. As $\xi_1, \xi_2, \cdots, \xi_N$ are linearly independent there is at least one ξ_i for which $(Ax-\mu(x)Bx, \xi_i)\neq 0$ holds for given x with $Ax-\mu(x)Bx\neq 0$. Thus from the discussion made in §3 it is obvious that $\mu(x^{(kN)})<\mu(x^{((k+1)N)})$ holds and we can expect the sequence $\{x^{(nN)}\}$ to converge to some eigenvector.

If we adopt special ξ_i such as $\xi_i=(\underbrace{0 \cdots 01}_{i}0 \cdots 0)(1\leqq i\leqq N)$ the computations needed in the approximation procedure are much simplified. For example we have $(x, \xi_i)_A=\sum_{j=1}^{N}a_{ij}x_j$, $(x, \xi_i)_B=\sum_{j=1}^{N}b_{ij}x_j$ and the correction is made with one component at each step.*) After some of the solutions are given we must use $\xi'_i(\xi'_i=\xi_i-$components in eigenvectors already obtained) instead of ξ_i but in this case $(x, \xi'_i)_A=(x, \xi_i)_A$, $(x, \xi'_i)_B=(x, \xi_i)_B$ hold for x which has no component in the space spanned by the eigenvectors already given and the process still maintains its simplicity. This process will be suited for the high speed computer for its simplicity of operation especially when N is large. Further it must be noticed that the simplicity of operation means less error due to roundings. Thus this generalized procedure will persuade the former ones at the final stage of approximation. This fact is observed in our numerical example.

9. Numerical examples**)

We have used eight digits for the computation of these examples except one for which we have used ten digits.

---

*) a_{ij} and b_{ij} are (i, j) element of A and B respectively.

**) In these examples x_i's are not normalized.

1. numerical results obtained by following the computation scheme of §5

$N=4$

$$A=\begin{pmatrix} 357.95382 & 1029.0039 & 3227.2540 & 494.03866 \\ 1029.0039 & 2964.8763 & 9199.1335 & 1408.3263 \\ 3227.2540 & 9199.1335 & 29993.650 & 4590.4629 \\ 494.03866 & 1408.3263 & 4590.4629 & 702.56158 \end{pmatrix} \qquad B=\begin{pmatrix} 2630.6504 & 5353.9122 & 2489.0993 & 2695.7564 \\ 5353.9122 & 18190.863 & 8753.2363 & 12098.720 \\ 2489.0993 & 8753.2363 & 57627.101 & 1775.7445 \\ 2695.7564 & 12098.720 & 1775.7445 & 9912.8825 \end{pmatrix}$$

$\delta=10^{-4}$(10^{-2} for x_2, x_3, x_4) (In this example λ_3 and λ_4 are expected to be equal to zero.)

$\xi=Ax-\mu(x)Bx$

n	$\mu(x^{(n)})$		$x^{(n)\prime}$				$\xi^{(n)\prime}$			
0	0.47781605		1.0000000	1.0000000	1.0000000	1.0000000	−1184.3089	−6612.131	13255.099	−5458.6625
1	0.57462573		0.91301057	0.51432813	1.9736088	0.59905229	808.0657	−813.229	−234.689	239.855
2	0.58371309		1.8644643	−0.44320510	1.6972752	0.88146860	793.4247	1441.568	−1556.677	2043.9827
3	0.58743430		1.9197121	−0.34282565	1.5888805	1.0237955	135.0344	−803.369	−501.409	255.9458
10	0.60455669		1.9775058	−1.1987619	0.80446428	1.3971709	8.3154	154.8215	−261.815	271.814
20	0.60551546		1.9171232	−1.2554594	0.73278888	1.4708243	−3.1410	17.3426	−18.071	27.9014
30	0.60552464		1.9093394	−1.2590469	0.72987352	1.4793227	−0 3156	1.6825	−1.640	2.6484
40	0.60552473		1.9085991	−1.2593596	0.72963831	1.4801279	−0.0308	0.1544	−0.155	0.2467
50	0.60552477		1.9085298	−1.2593891	0.72961587	1.4802029	−0.0041	0.0144	−0.014	0.0240
60	0.60552472		1.9085225	−1.2593922	0.72961381	1.4802105	−0.0002	0.0027	0.	0.0029
64	0.60552478	$x'_1=$	1.9085222	−1.2593922	0.72961373	1.4802109	−0.0007	−0.0002	0.	0.0006
0	0.031602296		−2.8278961	3.5259452	−0.46337600	−1.9688381	−9.4593	−0.0159	0.0575	−0.04062
1	0.084171722		−4.5468275	4.4123880	−0.25277905	−3.9006289	126.3582	36.2636	−60.2620	−102.3648
2	0.092785647		−3.4397388	4.2060671	−0.30523677	−3.9198952	−23.7336	−139.5782	98.2669	−136.5936
3	0.094077672		−3.4834098	4.1695707	−0.27677062	−4.0034901	14.6038	−2.7605	−13.2084	−14.6685
9	0.094232479	$x'_2=$	−3.2826925	4.0658898	−0.27516301	−3.9487717	−0.0009	−0.0018	−0.0050	−0.0007
0	−0.000000002	$x'_3=$	0.028632	−0.012104	−0.2239351	1.4672978	0.0003	0.0008	0.0015	0.0003
0	0.000000009	$x'_4=$	4.4206932	−1.2114234	−0.29972818	1.2781463	−0.0015	−0.0035	−0.0068	−0.0013

$N=3$

$$A=\begin{pmatrix} 13 & 10 & 4 \\ 10 & 9 & 7 \\ 4 & 7 & 13 \end{pmatrix} \qquad B=\begin{pmatrix} 14 & 11 & 5 \\ 11 & 15 & 10 \\ 5 & 10 & 18 \end{pmatrix}$$

$\delta=10^{-4}$

$\xi=Ax-\mu(x)Bx$ (In this example λ_3 is expected to be equal to zero.)

n	$\mu(x^{(n)})$		$x^{(n)\prime}$			$\xi^{(n)\prime}$		
0	0.77777778		1.0000000	1.0000000	1.0000000	3.6666670	−2.0000000	−1.6666670
1	0.93016359		3.4247832	−0.32260890	−0.10217430	0.0649380	1.0390110	−1.1039490
2	0.93723258		3.4545207	0.15319216	−0.60771284	−0.0493140	−0.4026050	−0.38182450
3	0.93751659		3.4474096	0.095136644	−0.66277182	−0.0057630	0.01459500	−0.02790600
4	0.93751950		3.4448796	0.10154399	−0.67502283	0.0008400	0.01195000	0.00607260
5	0.93751964		3.4450033	0.10330435	−0.67412827	−0.0003480	0.00087100	−0.00016623
6	0.93751974		3.4448504	0.10368699	−0.67485853	0.0000490	0.00071200	0.0003657
7	0.93751975		3.4448576	0.10379180	−0.67480470	−0.000023	0.000051	−0.0000973
8	0.93751975		3.4448475	0.10381419	−0.67484742	0.000002	0.000042	0.0000211
9	0.93751969		3.4448478	0.10382040	−0.67484430	0.000000	0.000006	−0.0000054
10	0.93751975		3.4448478	0.10382317	−0.67484679	−0.000002	−0.000004	−0.0000033
11	0.93751972	$x_1'=$	3.4448475	0.10382255	−0.67484730	0.000001	0.000001	0.0000003
12	0.93751973		3.4448478	0.10382281	−0.67484722	−0.000001	−0.000001	0.0000005
13	0.93751970		3.4448472	0 10382221	−0.67484692	0.000001	−0.000003	0.0000013
14	0.93751966		3.4448474	0.10382268	−0.67484672	0.000002	0.000001	−0.0000002
15	0.93751975		3.4448492	0.10382360	−0.67484690	−0.000003	−0.000007	−0.0000041
16	0.93751976		3.4448487	0.10382251	−0.67484754	−0.000002	−0.000001	0.0000020
17	0.93751971		3.4448462	0.10382128	−0.67484508	0.000000	0.000003	−0.0000035
0	0.64132013		−0.85248950	0.94416874	1.3629036	0.43405320	−2.2964371	1.8623840
1	0.72375321	$x_2'=$	−0.05479037	−0.35763046	2.3245995	0.0000005	0.0000000	0.0000000
0	−0.00000003	$x_3'=$	−0.81081080	1.2162161	−0.40540540	−0.0000006	−0.0000006	−0.0000007

$N=4$

$\delta=10^{-4}$

$\xi=Ax-m(x)Bx$

n	$\mu(x^{(n)})$	$m(x^{(n)})$	$x^{(n)\prime}$				$\xi^{(n)\prime}$			
0	0.47781605	0.53941075	1.0000000	1.0000000	1.0000000	1.0000000	−1995.4752	−9346.7340	8903.7300	−7089.8815
1	0.57590812	0.57255966	0.86291616	0.35790424	1.6116626	0.51294400	646.3748	−506.070	49.783	327.8948
2	0.59264340	0.58565025	3.4388891	−1.6589175	1.8100609	1.8196907	656.4547	2551.786	−1008.474	2828.766
3	0.59822951	0.59194008	3.4831205	−1.4869799	1.7421107	2.0102908	−153.9539	−418.241	75.413	500.0717
10	0.60366159	0.59679608	4.0226118	−2.4781824	1.8502292	3.0177849	351.2879	657.79	−327.254	718.8999
20	0.60551877	0.60502013	4.8408842	−3.1863902	1.8707318	3 7566417	19.6858	16.229	−9.963	19.487

$N=3$ (ten digits applied)

$\delta=10^{-4}$

$\xi=Ax-m(x)Bx$

n	$\mu(x^{(n)})$	$m(x^{(n)})$		$x^{(n)\prime}$			$\xi^{(n)\prime}$		
0	0.77777778	0.77260274		1.0000000	1.0000000	1.0000000	3.8219180	−1.813699	−1.495890
1	0.93030765	0.93756999		3.8663937	−0.3602530	−0.1219000	−0.290383	0.902533	−1.331128
2	0.93742680	0.94087441		3.7293741	0.0656155	−0.7500059	−0.137011	0.167214	0.166910
3	0.93751736	0.93758760		3.6991940	0.1024486	−0.7132398	−0.008385	−0.016226	−0.023292
4	0.93751966	0.93760359		3.6952452	0.1100899	−0.7242088	−0.003525	−0.004270	0.003687
5	0.93751971	0.93751469		3.6943783	0.1111401	−0.7233020	0.000017	0.000187	−0.001144
6	0.93751971	0.93752368		3.6943829	0.1111906	−0.7236108	−0.000230	0.000348	−0.000129
7	0.93751971	0.93751819		3.6942876	0.1113346	−0.7236643	0.000044	−0.000031	−0.000162
8	0.93751971	0.93752007		3.6942982	0.1113270	−0.7237036	−0.000021	0.000031	−0.0000117
9	0.93751971	0.93751955		3.6942893	0.1113405	−0.7237087	0.000005	−0.000004	−0.0000147
10	0.93751971	0.93751974		3.6942905	0.1113395	−0.7237121	−0.000001	0.000001	−0.0000011
11	0.93751971	0.93751973	$x_1'=$	3.6942896	0.11134022	−0.72371279	−0.000001	0.000001	0.0000006

2. numerical results obtained by following the acceleration scheme of §6

We have applied the acceleration scheme of §6 to the first example of this section. We have taken $x=x^{(32)}$ and the resulting $x-\varepsilon$ differes from $x^{(64)}$ only in its final digits. ε is obtained by the conjugate gradient method.

$x=x^{(32)}$	1.9090299	−1.2591787	0.72977391	1.4796588
ε	0.00050793	0.00021387	0.000160096	−0.00055272
$x-\varepsilon$	1.9085220	−1.2593926	0.72961381	1.4802115
$x^{(64)}$	1.9085222	−1.2593922	0.72961373	1.4802109

This result is trivial in some sense, but will serve for understanding of the state of affairs.

3. checking for x_1

We have applied our checking scheme to the first example of this section with $z^{(0)}=e-\frac{(e,x)}{||x||^2}x$ where $e=(1,1,1,1)'$.

$x=x_1$

n	$z^{(n)}$				$c(z^{(n)})$
0	0.31382590	1.4527913	0.73768079	0.46781736	-7.7431189×10^{-5}
1	0.09189588	0.50178314	1.0654544	−0.21673513	-1.8783074×10^{-4}
2	0.25257432	0.72602079	0.33869748	0.12510577	-7.7841331×10^{-5}
3	0.13749493	0.27272604	0.49787468	−0.19064798	-1.9123408×10^{-4}
86	0.00000007	0.00000008	−0.00000001	−0.00000005	-8.1177666×10^{-5}

$x=x_2$

n	$z^{(n)}$				$c(z^{(n)})$
0	0.73718181	1.3255223	0.97796996	0.68385433	$+3.3685725\times10^{-5}$

4. numerical results obtained by following the generalized scheme of §8

We have applied the computation scheme of §7 with $\xi_i=(\underbrace{0\cdots010\cdots0}_{i})$'s.

$N=4$

n	$\mu(x^{(n)})$	$x^{(n)\prime}$			
4×0	0.47781605	1.0	1.0	1.0	1.0
4×1	0.55474965	−0.44906200	0.00868745	1.4867737	1.2016828
4×2	0.58791528	1.1856323	−0.42189774	1.2143895	1.0575363
4×3	0.59492885	1.5715290	−0.58511478	0.97317382	0.97841635
4×10	0.60541932	1.1377673	−0.72340286	0.44047172	0.85557884
4×20	0.60552482	1.0973896	−0.72403476	0.41954005	0.85100234
4×28	0.60552478	1.0972273	−0.72403494	0.41946252	0.85098604
$(A(x^{(4\times28)})-\mu B(x^{(4\times28)}))'=$		−0.0009	0.0023	−0.0018	0.0002

$N=3$

n	$\mu(x^{(n)})$	$x^{(n)\prime}$		
3×0	0.77777778	1.0	1.0	1.0
3×1	0.93671463	−14.460012	0.32587380	2.3603754
3×2	0.93743486	−14.999564	−0.18978745	2.7770820
3×3	0.93751435	−14.901404	−0.38296706	2.8786173
3×4	0.93751951	−14.852119	−0.43313821	2.9006579
3×5	0.93751975	−14.839896	−0.44420899	2.9052768
3×6	0.93751976	−14.837261	−0.44654104	2.9062382
3×7	77	−14.836702	−0.44702499	2.9064359
3×8	78	−14.836622	−0.44712365	2.9064820
3×9	80	−14.836622	−0.44714754	2.9064970
3×10	76	−14.836622	−0.44715273	2.9064992
3×11	74	−14.836622	−0.44715481	2.9065009
3×12	78	−14.836622	−0.44715689	2.9065026
3×13	71	−14.836542	−0.44716416	2.9064915
3×14	67	−14.836542	−0.44716419	2.9064915
3×15	67	−14.836542	−0.44716219	2.9064915
$(A(x^{(3\times 15)})-\mu B(x^{(3\times 15)}))'=$		0.000000	0.000000	0.000002

10. Statistical applications

In this section we shall discuss some of the statistical problems which will show the aspects of the so-called canonical correlation analysis.

1. canonical correlations

In this section we shall use the following notations ; $a=(a_1, a_2, \cdots a_k)'$ $b=(b_1, b_2, \cdots b_l)'$ $x=(x_1, x_2, \cdots x_k)'$ $y=(y_1, y_2, \cdots y_l)'$ where x_i's and y_j's are random variables with $E\{x_i\}=E\{y_j\}=0$, $||u||^2=Eu^2$, $(u, v)=Euv$ for random variables u and v, and $\lambda=||b'y||/||a'x||$ $\rho=(a'x, b'y)/||a'x||\,||b'y||$. We have

$$||a'x||^2=E(a'x)^2=a'E(xx')a=a'Aa$$
$$||b'y||^2=E(b'y)^2=b'E(yy')b=b'Bb$$
$$(a'x, b'y)=E(a'x)(b'y)=a'E(xy')b=a'Cb=b'E(yx')a=b'C'a$$

where $A=E(xx')$, $B=E(yy')$ and $C=E(xy')$.

Canonical correlation coefficients are defined as the stationary values of ρ with respect to variables a and b.

We shall here present two formulations of statistical problems which will show the practical meaning of the canonical correlation coefficients.

a) stationary values of $\dfrac{||a'x-b'y||^2}{||a'x||^2+||b'y||^2}$

The function $||a'x-b'y||^2/||a'x||^2+||b'y||^2$ may be considered to be showing the effect of correlations between x and y on the difference $a'x-b'y$ compared with the case where x and y are taken to be uncorrelated. When we consider the ratio $||a'x-b'y||^2/||a'x||^2+||b'y||^2 =1-\frac{2\rho\lambda}{1+\lambda^2}$ as a function of ρ (function of ax' and $b'y/||b'y||$) and λ we have $\frac{\partial}{\partial\lambda}\left(1-\frac{2\rho\lambda}{1+\lambda^2}\right)=\frac{\rho(\lambda^2-1)}{(1+\lambda^2)^2}$ and generally (except for the trivial case where $\rho=0$) at $\lambda=\pm1$ the above stated ratio takes its maximum and minimum values $1\pm\rho$. Thus to discuss the stationary values of the above stated ratio is equivalent to discussing the stationary values of $(a'x, b'y)$ under the condition $||a'x||=||b'y||$. This is the ordinary formulation of the canonical correlation. As is easily seen the above stated ratio is represented as

$$\begin{pmatrix} a \\ b \end{pmatrix}'\begin{pmatrix} A & -C \\ -C' & B \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}\Big/\begin{pmatrix} a \\ b \end{pmatrix}'\begin{pmatrix} AO \\ OB \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=1-\begin{pmatrix} a \\ b \end{pmatrix}'\begin{pmatrix} O\,C \\ C'O \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}\Big/\begin{pmatrix} a \\ b \end{pmatrix}'\begin{pmatrix} AO \\ OB \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}$$

and to get the stationary points of this function we have only to apply the numerical method described in this paper to the equation

$$\begin{pmatrix} A & -C \\ -C' & B \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\lambda'\begin{pmatrix} AO \\ OB \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} \text{ or } \begin{pmatrix} O & -C \\ -C & O \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\lambda\begin{pmatrix} AO \\ OB \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}$$

and the eigenvalue $\lambda(\lambda')$ is equal to some canonical correlation coefficient $(+1)$.

b) stationary values of $\frac{||a'x-b'y||^2}{||a'x||^2}$

This formulation corresponds to the least square method. In this case we have

$$\frac{||a'x-b'y||^2}{||a'x||^2}=1-2\rho\lambda+\lambda^2$$

and this ratio takes its minimum value $1-\rho^2$ as a function of λ at $\lambda=\rho$ and $\lambda=\rho$ is the only value of λ for which the partial derivative of $(1-2\rho\lambda+\lambda^2)$ with respect to λ vanishes. Thus the discussion of stationary values of $\frac{||a'x-b'y||^2}{||a'x||^2}$ is reduced to the discussion of stationary values of ρ^2. Using the representation

$$\rho^2=\frac{(a'Cb)^2}{(a'Aa)(b'Bb)}$$

we have for stationary points

$$(C'aa'C)b=\frac{(a'Cb)^2}{\|b'y\|^2}Bb.$$

If $C'a\neq 0$ then taking into account the fact that the matrix at the left hand side has rank one there is only one solution of the equation $(C'aa'C)b=\lambda Bb$ with $\lambda\neq 0$. Thus if we can find a, b which satisfy the equation with $\lambda\neq 0$ this b will give the maximum ρ^2 for given a. The equation may be represented in the form $(a'Cb)C'a=\lambda Bb$, thus any b which satisfies the relation $Bb=kC'a$ with some constant k will be a solution of the equation. Especially when B^{-1} exists we can take $b=B^{-1}C'a$ and the problem is reduced to the discussion of the stationary points of the ratio $(a'CB^{-1}C'a)/(a'Aa)$ and this leads to the equation $CB^{-1}C'a=\lambda Aa$. The eigenvalues of this equation will give the stationary values of ρ^2. This formulation of problem may be of some use when we are interested only in the canonical vector a and corresponding ρ^2.

2. linear discriminant function (classification into three or more groups)

Here we assume $|A|\neq 0$. We can treat the general discriminant function problem by the formulation b). We take $\hat{y}_i$ as an indicator for the i-th group i.e. when an individual belongs to the i-th group then $\hat{y}_i=1$ otherwise $\hat{y}_i=0$. We shall here consider the case where l mutually exclusive groups exist and every individual belongs to at least one of them. We shall use the notation p_i to represent $E\{\hat{y}_i\}$ or the probability that an individual belongs to the i-th group. Obviously $\sum_{i=1}^{l} p_i=1$ holds. We take y_i as given by $y_i=\hat{y}_i-p_i$ and in this case by the formulation of b) it can easily be seen that $B\bar{a}=C'a$ holds where $\bar{a}$ B and C' are given as follows;

$$\bar{a}=(\sum_{i=1}^{k}a_i(\bar{x}_i(1)-\bar{x}_i),\ \sum_{i=1}^{k}a_i(\bar{x}_i(2)-\bar{x}_i),\ \cdots\ \sum_{i=1}^{k}a_i(\bar{x}_i(l)-\bar{x}_i))$$
$$=(\sum_{i=1}^{k}a_i\bar{x}_i(1),\ \sum_{i=1}^{k}a_i\bar{x}_i(2),\ \cdots\ \sum_{i=1}^{k}a_i\bar{x}_i(l))$$

$\bar{x}_i(j)$=conditional expectation of x_i over the group j

$\bar{x}_i$=(expectation of x_i)$=\sum_{j=1}^{l}p_jx_i(j)$ (now taken to be equal to zero)

$$B=(b_{ij})\qquad b_{ij}=\delta(ij)p_i-p_ip_j\qquad (\delta(ii)=1\ \ \delta(ij)=0\ \ (i\neq j)).$$
$$C'=(c_{ij})\qquad c_{ij}=p_i\bar{x}_j(i)\ .$$

We have thus only to discuss the ratio

$$\frac{(a'Cb)}{(a'Aa)}\text{ taking } b \text{ equal to } \bar{a} \text{ or } \frac{(a'C\bar{a})}{(a'Aa)}.$$

Now $(a'C\bar{a})=\sum_{i=1}^{k}\sum_{\nu=1}^{l}a_i c_{\nu i}\sum_{j=1}^{k}a_j\bar{x}_j(\nu)=\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\nu=1}^{l}\bar{x}_i(\nu)\bar{x}_j(\nu)p_\nu a_i a_j$ and if there is at least one j_i with $\bar{x}_i(j_i)\neq\bar{x}_i=0$ for each i then $C'a\neq 0$. Thus to apply the general linear discriminant function method we have only to solve the equation

$$Da=\lambda Aa$$

where $D=(\sum_{\nu=1}^{l}\bar{x}_i(\nu)\bar{x}_j(\nu)p_\nu)$ (in practical applications care should be taken to the fact that we have put $\bar{x}_j=0$ otherwise we must replace $\bar{x}_i(\nu)$ in D with $\bar{x}_i(\nu)-\bar{x}_i$). In this case we can expect at most $l-1$ eigenvalues to be >0.

3. quantification problem

When x_i is given by the form $\hat{x}_i-E(\hat{x}_i)$ where $\hat{x}_i$ takes only values 0 or 1 A takes orather simple form and in this case $a'x$ may be considered as a score given to each individual. This is the case which is taken up by Hayashi in the so-called quantification problem.

4. trend analysis of multiple time series

As was discussed by M. S. Bartlett [8] when we want to analyse the trend pattern of linear combination of some multiple time series it may sometimes be useful to analyse the canonical correlation between the multiple time series and a set of regular curves. If we take as a set of regular curves a set of orthogonal polynomials it is very easy to analyse the canonical correlation.

Given a multiple time series $(\hat{x}_1(t), \hat{x}_2(t), \cdots \hat{x}_k(t))$ $t=0, 1, 2, \cdots N$ and a set of orthogonal polynomials of degree N $(y_1(t), y_2(t), \cdots y_l(t))$ $t=0, 1, 2, \cdots N (l\leq N)$ we can calculate A, B and C by taking the symbol E as an operator for arithmetic mean of the series. For example we put $x_i(t)=\hat{x}_i(t)-E\{\hat{x}_i\}$ where $E\{\hat{x}_i\}=\sum_{t=0}^{N}\hat{x}_i(t)/N+1$ and $A=((x_i, x_j))$ where $(x_i, x_j)=Ex_ix_j=\sum_{t=0}^{N}x_i(t)x_j(t)/N+1$.

In this case if we take the polynomials to be normalized we have $B=I$ (identity matrix) and we have only to solve the equation $(CC')a=\lambda Aa$.

Acknowledgement

The author expresses his heartily thanks to Dr. C. Hayashi for guidance to this problem and kind discussions about the method and practical applications. Thanks are also to Prof. F. M. Verzuh at M. I. T. for kind suggestions of bibliography about related topics. The numerical examples of this paper are entirely due to the work of Miss Y. Saigusa. Without these examples this paper would be less convincing.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] C. Hayashi, On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics* Vol. III (1952) pp. 69-98.

[2] ———, Multidimensional quantification—with the applications to analysis of social phenomena—. *Annals of the Institute of Statistical Mathematics* Vol. V (1954) pp. 121-143.

[3] S. Sakino and G. Kono, Forecasting of prognosis by quantiffcation. (in Japanese) *Proceedings of the Institute of Statistical Mathematics* Vol. I No. 1 (1953) pp. 17-25.

[4] F. Mosteller, Application of computing machinery to the solution of problems of the social sciences. *Proceedings of a second symposium on large-scale digital culculating machinery* (1951) pp. 323-332.

[5] H. Hotelling, Practical problems of matrix calculation, *Proceedings of the Berkeley symposium on mathematical statistics and probability* (1949) pp. 275-293.

[6] P. S. Dwyer, *Linear Computations*. Now York. John Wiley Sons, Inc. (1951) § 18.9 pp. 318-322.

[7] S. Huzino, On calculating eigenvalues by the gradient method. *Memoirs of the Faculty of Science Kyushu University*, Vol. XII, No. 1 (1958) pp. 30-39.

[8] M. S. Bartlett, A note on the statistical estimation of supply and demand relations from time series. *Econometrica* Vol. 16 (1948) pp. 323-329.