P.R. Krishnaiah, ed., *Multivariate Analysis – VI*
© Elsevier Science Publishers B.V. (1985) 485–493

PROJECTIONS AND ITERATIVE PROCEDURES

Ludger RÜSCHENDORF
University of Freiburg, West Germany

1. Conditional expectations and projections

Let $(M, \mathfrak{A}, P)$ be a probability space and $\mathfrak{A}_1 \subset \mathfrak{A}$, $1 \leqslant i \leqslant k$, be k sub σ-algebras. Define the subspace

$$F = \left\{ \sum_{i=1}^{k} f_i \,; f_i \in L_2(\mathfrak{A}_i, P), 1 \leqslant i \leqslant k \right\}$$

and $\bar{F}$ the closure of F in $L_2(\mathfrak{A}, P)$. Our aim is to describe explicitly the best approximation of $\varphi \in L_2(\mathfrak{A}, P)$ by a (more simple) element of $\bar{F}$, i.e. to determine the projection T: $L_2(\mathfrak{A}, P) \to \bar{F}$. In the special case that $\mathfrak{A}_i$ and $\mathfrak{A}_j$ are conditionally independent given $\mathfrak{A}_i \cap \mathfrak{A}_j$, $i \neq j$ (i.e.

$$E(1_{A_i} 1_{A_j} | \mathfrak{A}_i \cap \mathfrak{A}_j) = E(1_{A_i} | \mathfrak{A}_i \cap \mathfrak{A}_j) E(1_{A_j} | \mathfrak{A}_i \cap \mathfrak{A}_j)$$

for $A_i \in \mathfrak{A}_i$ and $A_j \in \mathfrak{A}_j$; cf. Loeve [10, p. 351]), this problem has the following solution:

Proposition 1. *Let for $i \neq j$, $\mathfrak{A}_i$ and $\mathfrak{A}_j$ be conditionally independent given $\mathfrak{A}_i \cap \mathfrak{A}_j$. Define for $J \subset \{1, \ldots, k\}$, $\mathfrak{A}_J = \bigcap_{j \in J} \mathfrak{A}_j$ and for $\varphi \in L_2(\mathfrak{A}, P)$*

$$T\varphi = \sum_{j=1}^{k} \sum_{\substack{J \subset \{1, \ldots, k\} \\ |J| = j}} (-1)^{j+1} E(\varphi \,|\, \mathfrak{A}_J)\,, \tag{1}$$

then $T\varphi \in F$ is the projection of φ on $F = \bar{F}$.

Proof. Clearly, for $\varphi \in L_2(\mathfrak{A}, P)$ holds $T\varphi \in F$. Furthermore, for $f_1 \in L_2(\mathfrak{A}_1, P)$, we have using $E(f_1 | \mathfrak{A}_{J \cup \{1\}}) = E(f \,|\, \mathfrak{A}_J)$ (cf. Loeve [10, p. 351]) that

$$Tf_1 = \sum_{j=1}^{k} (-1)^{j+1} \Big\{ \sum_{\substack{|J|=j \\ 1\in J}} + \sum_{\substack{|J|=j \\ 1\notin J}} \Big\} E(f_1 \mid \mathfrak{A}_J)$$

$$= f_1 + \sum_{j=1}^{k-1} (-1)^{j+1} \sum_{\substack{|J|=j-1 \\ 1\notin J}} E(f_1 \mid \mathfrak{A}_J)$$

$$+ \sum_{j=1}^{k-1} (-1)^{j+1} \sum_{\substack{|J|=j \\ 1\notin J}} E(f_1 \mid \mathfrak{A}_J) = f_1 .$$

Similarly, $Tf_i = f_i$ for $f_i \in L_2(\mathfrak{A}_i, P)$ and, therefore, $Tf = f$ for all $f \in F$ and $T^2 = T$. Using the relation $Ef_1 E(\varphi \mid \mathfrak{A}_1) = Ef_1\varphi$ and $Ef_1 E(\varphi \mid \mathfrak{A}_J) = Ef_1 E(\varphi \mid \mathfrak{A}_{J\cup\{1\}})$ for $\varphi \in L_2(\mathfrak{A}, P)$, $f_1 \in L_2(\mathfrak{A}_1, P)$, we obtain

$$Ef_1(T\varphi) = Ef_1 \sum_{j} (-1)^{j+1} \sum_{|J|=j} E(\varphi \mid \mathfrak{A}_J)$$

$$= Ef_1 \Big\{ E(\varphi \mid \mathfrak{A}_1) + \sum_{j=1}^{n} (-1)^{j+1} \sum_{\substack{|J|=j \\ 1\in J}} E(\varphi \mid \mathfrak{A}_J)$$

$$+ \sum_{j=1}^{n-1} (-1)^{j+1} \sum_{\substack{|J|=j \\ 1\notin J}} E(\varphi \mid \mathfrak{A}_J) \Big\} = Ef_1\varphi .$$

Similarly, $Ef_i(T\varphi) = Ef_i\varphi$, $1 \leqslant i \leqslant n$, and, therefore, $T\varphi$ is the orthogonal projection on F implying $F = \bar{F}$, since the range of an orthogonal projection is closed. □

Remark 1. (a) The condition that $\mathfrak{A}_i$ and $\mathfrak{A}_j$ are conditionally independent given $\mathfrak{A}_i \cap \mathfrak{A}_j$ can be shown to be equivalent to the condition that the conditional expectation operators $T_i\varphi = E(\varphi \mid \mathfrak{A}_i)$ are commuting, i.e. $T_iT_j\varphi = T_jT_i\varphi = E(\varphi \mid \mathfrak{A}_i \cap \mathfrak{A}_j)$. Furthermore, $T\varphi$ is easily shown to be identical to $(I - \Pi_{i=1}^{k} (I - T_i))\varphi$, where I is the identity operator. But for T_i commuting $\Pi_{i=1}^{k}(I - T_i)$ is the orthogonal projection of $L_2(\mathfrak{A}, P)$ on $\bigcap_{i=1}^{k} (L_2(\mathfrak{A}_i, P))^{\perp}$ and, therefore, T is the projection on F. From this point of view Proposition 1 follows from Corollary 7 of Rao and Yanai [12].

(b) T is well defined also on $L_s(\mathfrak{A}, P)$, $s \geqslant 1$, and can be shown to define the orthogonal projection on F_s, where

$$F_s = \left\{ \sum_{i=1}^{k} f_i \, ; f_i \in L_s(\mathfrak{A}_i, P), 1 \leq i \leq k \right\},$$

i.e. (1) $T\varphi \in F_s$, for $\varphi \in L_s(\mathfrak{A}, P)$, (2) $T^2 = T$ and (3) for $\varphi \in L_s(\mathfrak{A}, P)$ and $f \in F_r$, $(1/r) + (1/s) = 1$ holds $Ef(T\varphi) = Ef\varphi$.

(c) Let $X_1, \ldots, X_k$ be stochastically independent random variables on $(M, \mathfrak{A}, P)$. If we take $\mathfrak{A}_i = \mathfrak{A}(X_i)$, $1 \leq i \leq k$, we obtain from Proposition 1

$$T\varphi = \sum_{i=1}^{k} E(\varphi \mid X_i) - (k-1)E\varphi \tag{2}$$

is for $\varphi \in L_2(\mathfrak{A}, P)$ the orthogonal projection on F. This projection was considered by Hajek [6] and was shown to have important applications in asymptotic statistics.

Clearly, a better approximation of φ can be obtained by considering $\mathfrak{A}_{ij} = \mathfrak{A}(X_i, X_j)$, $1 \leq i \leq j \leq k$. Also, $\mathfrak{A}_{ij}$ and $\mathfrak{A}_{lm}$ are conditionally independent given $\mathfrak{A}_{ij} \cap \mathfrak{A}_{lm}$. By Proposition 1 the best approximation of $\varphi \in L_2(\mathfrak{A}, P)$ by functions of the type $\Sigma_{i \leq j} f_{ij}(X_i, X_j)$ is of the form

$$T\varphi = \sum_{i<j} E(\varphi \mid X_i, X_j) + \beta \sum_{i=1}^{n} E(\varphi \mid X_i) + \gamma E\varphi \,. \tag{3}$$

Since $E(K_{ij}(\varphi) \mid X_l) = 0$ for all i, j, l where $K_{ij}(\varphi) = E(\varphi \mid X_i, X_j) - E(\varphi \mid X_i) - E(\varphi \mid X_j) + E\varphi$, we obtain

$$T\varphi = \sum_{i<j} K_{ij}(\varphi) + \sum_{i=1}^{k} E(\varphi \mid X_i) - (k-1)E\varphi \,. \tag{4}$$

Equation (4) shows that the best approximation is given by means of a V-statistic and a linear statistic. This procedure can be extended to more than two variables and is known in the literature under the name of Hoeffding decomposition (cf. Karlin and Rinott [9], Rubin and Vitale [14] and Efron and Stein [5]).

2. Projections and iterative procedures

In the general case an explicit description of the projection T by means of the conditional expectations T_i, $1 \leq i \leq k$, is not possible, but T can be described by the following iterative procedure which is an analogon to the

Diliberto–Strauss [4] leveling process used by these authors to approximate continuous functions w.r.t. the Chebychev norm.

Define for $f \in L_2(\mathfrak{A}, P)$,

$$S_1(f) = E(f \mid \mathfrak{A}_1) \tag{5a}$$

and for $n = mk + r$, $1 \leqslant r \leqslant k$,

$$S_n(f) = S_{n-1}(f) + E(f - S_{n-1}(f) \mid \mathfrak{A}_r)\,. \tag{5b}$$

The idea of this procedure is to do in each step the best you can using the known conditional expectations on $\mathfrak{A}_r$, i.e. for each $g \in L_2(\mathfrak{A}_r, P)$ holds

$$E(f - (S_{n-1}(f) + g))^2 \geqslant E(f - S_n(f))^2\,. \tag{6}$$

Proposition 2. *For $f \in L_2(\mathfrak{A}, P)$, $S(f) = \lim_{n\to\infty} S_n(f)$ exists in $L_2(\mathfrak{A}, P)$ and $S(f)$ is the projection of f on $\bar{F}$.*

Proof. The proof follows from a theorem due to von Neumann [11], Wiener [15] and Halperin [8] on the iterated products of projection operators in Hilbert spaces, stating that for subspaces $H_i \subset H$, $1 \leqslant i \leqslant k$, and projections T_i of H on H_i, T^n converges to the projection on $H_1 \cap \cdots \cap H_k$ where $T = T_k T_{k-1} \cdots T_1$. Observing that $S_{mk} = I - [(I - T_k) \cdots (I - T_1)]^m$, we obtain that S_n converges to the projection on $\bar{F} = (\bigcap_{i=1}^k (L_2(\mathfrak{A}_i, P))^\perp)^\perp$. □

Burkholder and Chow [2] and more generally Rota [13] observed that for bistochastic linear operators (such as conditional expectations) even a.s. convergence holds in the alternating projection theorem. In order to obtain a.s. convergence results we restrict ourselves to the case $k = 2$; certain generalizations to the case $k \geqslant 2$ are obvious. Define

$$\rho = \rho(\mathfrak{A}_1, \mathfrak{A}_2) = \sup\{\mathrm{cor}(f_1, f_2); f_i \in L_2(\mathfrak{A}_i, P),\; E(f_i \mid \mathfrak{A}_1 \cap \mathfrak{A}_2) = 0, i = 1, 2\}\,.$$

With $F_i' = F_i \ominus (F_1 \cap F_2)$, $F_i = L_2(\mathfrak{A}_i, P)$ and $T_i\varphi = E(\varphi \mid \mathfrak{A}_i)$, $i = 1, 2$, $\rho = \rho(\mathfrak{A}_1, \mathfrak{A}_2)$ is the cosinus of the smallest angle between F_1' and F_2'. The proof of the following theorem is immediate from results of Aronszajn [1, pp. 375–379].

Theorem 3. (a) $\|(T_2T_1)^n\varphi\| \leq \rho^{2n}\|\varphi\|$, $\varphi \in F_2'$, *and* $\|(T_1T_2)^n\varphi\| \leq \rho^{2n}\|\varphi\|$, $\varphi \in F_1'$.

(b) $\rho < 1$ *is equivalent to the condition that* F *is closed in* $L_2(\mathfrak{A}, P)$.

(c) *If* $\rho < 1$ *and* S *denotes the projection on* F, *then* $S = Q_0 + Q_1 + Q_2$, *where* $Q_0\varphi = E(\varphi \mid \mathfrak{A}_1 \cap \mathfrak{A}_2)$,

$$Q_1\varphi = (T_1 - T_1T_2 + T_1T_2T_1 - \cdots)(\varphi - Q_0\varphi),$$

$$Q_2\varphi = (T_2 - T_2T_1 + T_2T_1T_2 - \cdots)(\varphi - Q_0\varphi);$$

Q_1, Q_2 *are uniformly and* a.s. *convergent.*

(d) *If* $\rho < 1$, *then* $\|S\varphi - (Q_0 + Q_{1,n} + Q_{2,n})\varphi\| \leq 2\rho^{2n-1}\|\varphi\|$, *where* $Q_{i,n}$ *denote the first* n*-terms of* Q_i, $i = 1, 2$, *and* $\lim S_n(\varphi) = S(\varphi)$ a.s. □

We remark that Aronszajn's results also provide a simple and general alternative proof of the Burkholder–Chow result. For some applications we shall need the projection on the intersection of affine spaces of $L_2(\mathfrak{A}, P)$.

Lemma 4. *Let* $G_i = c_i + F_i$, $1 \leq i \leq k$, *be affine subspaces of a Hilbert space* H *and let* T *be the projection on* $F_1 \cap \cdots \cap F_k$.

(a) *If* $G_1 \cap \cdots \cap G_k \neq \emptyset$, *then*

$$G_1 \cap \cdots \cap G_k = x_0 + F_1 \cap \cdots \cap F_k,$$

$$\text{for all } x_0 \in G_1 \cap \cdots \cap G_k.$$

(b) *If* $x_0 \in G_1 \cap \cdots \cap G_k$ *is orthogonal to* $F_1 \cap \cdots \cap F_k$, *then* $\tilde{T}x = x_0 + Tx$ *is the projection on* $G_1 \cap \cdots \cap G_k$.

Proof. (a) is trivial.

(b) For $y \in G_1 \cap \cdots \cap G_k$, $x \in H$ holds

$$\|y - x\|^2 = \|y - \tilde{T}x\|^2 + \|\tilde{T}x - x\|^2 + 2(y - \tilde{T}x, \tilde{T}x - x).$$

But $y - \tilde{T}x \in F_1 \cap \cdots \cap F_k$ and $\tilde{T}x - x = x_0 + (Tx - x)$ implies $(y - \tilde{T}x, \tilde{T}x - x) = 0$. □

Corollary 5. *Let* $G_i = c_i + F_i$, $1 \leq i \leq k$, *be affine subspaces of a Hilbert space* H *with nonempty intersection. Let* $\tilde{T}_i$ *denote the projections on* G_i *and* $\tilde{T} = \tilde{T}_k \cdots \tilde{T}_1$, *then* $(\tilde{T})^n h$ *converges for* $h \in H$ *to the projection on* $G_1 \cap \cdots \cap G_k$.

Proof. We consider only the case $k = 2$. Let $h_0 \in G_1 \cap G_2$ be orthogonal to $F_1 \cap F_2$, then by Lemma 3, $Sh = h_0 + \lim_{n\to\infty} T^n h$ is the projection of h on $G_1 \cap G_2$, where $T = T_2 T_1$ and T_i are the projections on F_i. But $\tilde{T}_i h = a_i + T_i h$, where $a_i \in G_i$ are orthogonal to F_i, $i = 1, 2$. Therefore, after some calculations we obtain

$$(\tilde{T}_2\tilde{T}_1)^n h = \sum_{\nu=0}^{n-1} (T_2T_1)^\nu a_2 + T_2 \sum_{\nu=0}^{n-1} (T_1T_2)^\nu a_1 + (T_2T_1)^n h. \tag{7}$$

For $h = h_0$, $(\tilde{T}_2\tilde{T}_1)^n h_0 = h_0$ and $(T_2T_1)^n h_0$ converges to 0. Therefore, we obtain

$$h_0 = \lim_{n\to\infty} \left(\sum_{\nu=0}^{n-1} (T_2T_1)^\nu a_2 + T_2 \sum_{\nu=0}^{n-1} (T_1T_2)^\nu a_1 \right),$$

which implies that $\lim(\tilde{T}_2\tilde{T}_1)^n h = Sh$. □

Remark 2. (a) Let $X_1, \ldots, X_k$ be i.i.d. real random variables and let $\mathfrak{A}_i = \mathfrak{A}(R_i)$, $1 \le i \le k$, where R_i is the rank of X_i. In order to determine $E(\varphi \mid \bar{F})$ for $\varphi \in L_2(\mathfrak{A}, P)$ it is sufficient to consider functions $\varphi = \varphi(R_1, \ldots, R_k)$, since φ and $E(\varphi \mid R_1, \ldots, R_k)$ have the same projection on $\bar{F}$. If $E\varphi = 0$, we obtain by simple calculation

$$E(E(\varphi \mid R_i) \mid R_j = l) = -\frac{1}{k-1} E(\varphi \mid R_i = l)$$

and

$$E(\varphi \mid R_s = l) = \frac{1}{(k-1)!} \sum_{r_1, \ldots, r_{s-1}, r_{s+1}, \ldots, r_k} \varphi(r_1, \ldots, l, \ldots, r_k).$$

From a symmetric argument and from Proposition 2, $T\varphi$ is of the form $\alpha \sum_{i=1}^k E(\varphi \mid R_i) + \beta E\varphi$. It is now easy to determine α and β from the orthogonality conditions to obtain

$$T = \frac{k-1}{k} \sum_{i=1}^{k} E(\varphi \mid R_i) - (k-2)E\varphi; \quad \varphi \in L_2(\mathfrak{A}, P). \tag{8}$$

This result was also given by Hajek and Sidak [7, p. 59]. It allows one to identify certain linear rank statistics as projections of more complicated nonlinear rank statistics.

If we consider $\mathfrak{A}_{ij} = \mathfrak{A}(R_i, R_j)$, $i \le j$, and use for $\varphi = \varphi(R_1, \ldots, R_k)$

with $E\varphi = 0$ and $g_{ij}(R_i, R_j) = E(\varphi \mid R_i, R_j)$, that

$$E(E(\varphi \mid R_1, R_2) \mid R_i = k, R_j = l)$$
$$= \frac{1}{(k-2)(k-3)}\{g_{12}(l, k) + g_{12}(k, l)\}$$
$$- \frac{k-1}{(k-2)(k-3)}\{E(\varphi \mid R_2 = k) + E(\varphi \mid R_2 = l)$$
$$+ E(\varphi \mid R_1 = k) + E(\varphi \mid R_1 = l)\}$$

for $i \neq j$, $i, j \notin \{1, 2\}$, we obtain similarly that the projection is of the form

$$T\varphi = \alpha \sum_{i<j} E(\varphi \mid R_i, R_j) + \beta \sum_{i=1}^{k} E(\varphi \mid R_i) + \gamma E\varphi\,, \tag{9}$$

where α, β and γ have to be determined by the orthogonality relations $E(T\varphi \mid R_1) = E(\varphi \mid R_1)$, $E(T\varphi \mid R_1, R_2) = E(\varphi \mid R_1, R_2)$, $ET\varphi = E\varphi$ and are independent of φ.

(b) Let R be a probability measure on $(M, \mathfrak{A})$ and let $Q_i/\mathfrak{A}_i$, $1 \leq i \leq k$, be probability measures with $Q_i \ll R_i = R/\mathfrak{A}_i$, $1 \leq i \leq k$. Let $F_i = \{f \in L_2(\mathfrak{A}, R);\ E_R(f \mid \mathfrak{A}_i) = 0\}$ and assume that

$$\frac{dQ_i}{dR_i} \in L_2(\mathfrak{A}_i, R)$$

and that

$$G_i = \frac{dQ_i}{dR_i} + F_i, \quad 1 \leq i \leq k,$$

have a nonempty intersection. Any element $f \in G = \bigcap_{i=1}^{k} G_i$ determines a signed measure on $(M, \mathfrak{A})$ with marginals Q_i, $1 \leq i \leq k$, and square integrable densities w.r.t. R.

By Proposition 2 with $T = \Pi_{i=1}^{k}(I - T_i)$, $T_i\varphi = E(\varphi \mid \mathfrak{A}_i)$, $T^n\varphi = \varphi - S_n\varphi$ converges to the projection on $F_1 \cap \cdots \cap F_k$. Therefore, by Corollary 5, $\tilde{T}^n$ converges to the projection on $G = \bigcap_{i=1}^{k} G_i$, where

$$\tilde{T} = \prod_{i=1}^{k} \tilde{T}_i \quad \text{and} \quad \tilde{T}_i\varphi = \frac{dQ_i}{dR_i} + \varphi - E(\varphi \mid \mathfrak{A}_i)\,.$$

So in this case we obtain asymptotically the projection of a measure φR to a measure with marginals Q_i and L_2 density w.r.t. R.

If we are interested in the projection of a measure R on a nonnegative measure with marginals Q_i and L_2 density w.r.t. R we would use $T' = T_{k+1}\tilde{T}$, where T_{k+1} is the (nonlinear but contractive) projection on the nonnegative elements of $L_2(\mathfrak{A}, R)$.

By a fixed point theorem as, for example, Theorem 1 of Cheney and Goldstein [3] we would obtain convergence of $(T')^n f$ of this projection but only in the weak topology on $L_2(\mathfrak{A}, R)$ if the σ-algebras $\mathfrak{A}_i$ are finite.

(c) If we take $(M, \mathfrak{A}, P) = ([0,1]^k, \mathfrak{B}^k[0,1]^k, P)$, P being any probability measure and $\mathfrak{A}_i$ to be the σ-algebra generated by the ith projection, then the leveling process $S_n(f)$ for $f = f(x_1, \ldots, x_k) \in L_2(\mathfrak{B}^k[0,1]^k, P)$ converges to the best approximation by means of functions $\Sigma_{i=1}^k f_i(x_i)$ w.r.t. $L_2(P)$ distance.

So this case corresponds to the theorem of Diliberto and Strauss [4], who consider the Chebychev norm for continuous functions. The case $P = \lambda^k$, the Lebesgue measure on $[0,1]^k$ is contained in Proposition 1.

Acknowledgement

I want to thank M. Pourahmadi for pointing out the relevance of the paper of Aronszajn.

References

[1] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68**, 337–404.

[2] Burkholder, D.L. and Chow, Y.S. (1961). Iterates of conditional expectation operators. *Proc. Amer. Math. Soc.* **12**, 490–495.

[3] Cheney, E.W. and Goldstein, A.A. (1959). Proximity maps for convex sets. *Proc. Amer. Math. Soc.* **10**, 448–450.

[4] Diliberto, S.P. and Strauss, E.G. (1951). On the approximation of functions of several variables by the sum of functions of fewer variables. *Pacific J. Math.* **1**, 195–210.

[5] Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9**, 586–596.

[6] Hajek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39**, 325–346.

[7] Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

[8] Halperin, I. (1962). The product of projection operators. *Acta Sci. Math.* **23**, 96–99.

[9] Karlin, S. and Rinott, Y. (1982). Applications of ANOVA type decompositions for

comparison of conditional variance statistics including jackknife estimates. *Ann. Statist.* **10**, 485–501.

[10] Loeve, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton, NJ.

[11] von Neumann, J. (1950). *Functional Operators*, Vol. 2. Princeton University Press, Princeton, NJ.

[12] Rao, C.R. and Yanai, H. (1979). General definition and decomposition of projectors and some applications to statistical problems. *J. Statist. Plann. Inference* **3**, 1–17.

[13] Rota, G.C. (1962). An "Alternierende Verfahren" for general positive operators. *Bull. Amer. Math. Soc.* **68**, 95–102.

[14] Rubin, H. and Vitale, R.A. (1980). Asymptotic distributions of symmetric statistics. *Ann. Statist.* **8**, 165–170.

[15] Wiener, N. (1955). On the factorization of matrices. *Comment. Math. Helv.* **29**, 97–111.