1. Introductory section

1.1. Introduction. Consider the data-bases

$$\{(X_{ij}, Y_{ij}),\ i = 1, \dots, k,\ j = 1, \dots, a(i)\} \tag{1.1}$$

consisting of $k \geq 1$ series of bipartite observations; the ith one contains $a(i) \geq 1$ items and corresponds to the ith model in a system. Such data-bases arise whenever the observed phenomenon is viewed as a structural system consisting of k subsystems in the ith one of which $a(i)$ observations are collected. For example, the body size system includes subsystems such as (height, arm length), (chest width, shoulder width), etc.; observations $(X_{11}, Y_{11}), (X_{12}, Y_{12}), \dots$ on (height, arm length), $(X_{21}, Y_{21}), (X_{22}, Y_{22}), \dots$ on (chest width, shoulder width), etc. are available. The data-bases (1.1) also correspond to the case where the relationship between two variables x, y changes from a time period to a subsequent period, e.g. y is the weight increase rate of a man and x the mean percentage of protein in food given to him daily, the relation of y to x depends on growth periods; then $(X_{11}, Y_{11}), (X_{12}, Y_{12}), \dots$ are observations on (x, y) in the baby period, $(X_{21}, Y_{21}), (X_{22}, Y_{22}), \dots$ are those in a next period, etc. Naturally global analysis of the data-bases is desired. For simplicity write $(X_{ij}, \mathbf{Y}_{ij}) = (x, \mathbf{y})$, where $\mathbf{y}$ is assumed to be vector-valued. In the ith subsystem one might seek to represent the response $\mathbf{y}$ by some spline function of the explanatory variable x, namely

$$\mathbf{y}' \approx \sum_{g=1}^{h} \mathbf{c}'_g(x) I_{S(g)}(x) \mathbf{q}_g,$$

where a prime denotes transpose, $S(1), \dots, S(h)$ are disjoint sets in the range space, quite arbitrary, of x, the $\mathbf{c}_g(\cdot)$ are known vector functions, the $I_{S(g)}(\cdot)$ set indicators and the $\mathbf{q}_g$ matrix parameters. Using block matrices $\mathbf{q} = (\mathbf{q}'_1 \dots \mathbf{q}'_h)'$ and $\mathbf{b}'(\cdot) = (\mathbf{c}'_1(\cdot) I_{S(1)}(\cdot) \dots$ $\dots \mathbf{c}'_h(\cdot) I_{S(h)}(\cdot))$ the spline representation is put in the linear form

$$\mathbf{y}' \approx \mathbf{b}'(x)\mathbf{q}.$$

Thus, assuming the explanatory variables X_{ij} $(j = 1, \dots, a(i))$ to have an arbitrary range space H_i in common $(i = 1, \dots, k)$ and the $\mathbf{Y}_{ij}$ to be $r \times 1$ response vectors, using spline approximation we shall consider the formal representation of the data-bases (1.1) by a system of k linear *correlation models*, according to the terminology in Freedman (1981):

$$\mathbf{Y}'_{ij} = \mathbf{b}'_{ij}\theta_i + \eta'_{ij}, \quad \mathbf{b}_{ij} = \mathbf{b}_i(X_{ij}), \quad i = 1, \dots, k,\ j = 1, \dots, a(i), \tag{1.2}$$

where the $\mathbf{b}_i(\cdot)$ are known $\ell(i) \times 1$ vector functions on H_i, the θ_i unknown $\ell(i) \times r$ matrix parameters and the η_{ij} residuals. In physics, chemistry, biology, etc., to determine some unknown constants one may have several series of observations (1.1) which correspond to

k experiments different in nature; then the use of spline approximation gives rise to the system of models (1.2) in which the parameters are subject to the constraint $\theta_1 = \ldots = \theta_k$.

The purpose of this paper is to establish the convergence and strong consistency of least squares estimates (LSE) of parameters in the system of models (1.2) under minimal assumptions. But we shall define the LSE in the more general situation where the regressors $\mathbf{b}_{ij}$ may be random or non-random and may depend or not on explanatory variables. Let us use the symbols

$$\ell = \ell(1) + \ldots + \ell(k),$$
$$\mathcal{M}_{p\times q} = \text{linear space of all } p \times q \text{ real matrices.}$$

With the residuals $\eta_{ij} = \mathbf{Y}_{ij} - \theta_i'\mathbf{b}_{ij},\ i = 1, \ldots, k,\ j = 1, \ldots, a(i)$, in models (1.2), the *global residual* is defined as a function of the *global matrix parameter* $\theta = (\theta_1' \ldots \theta_k')'$, namely

$$\eta(\theta) = (\eta_{11}' \ldots \eta_{1a(1)}' \vdots \ldots \vdots \eta_{k1}' \ldots \eta_{ka(k)}')'. \tag{1.3}$$

Definition 1.1. A *generalized least squares* (GLS) *value* for θ is any value $\widehat{\theta}$ which minimizes some norm of the global residual $\eta(\mathbf{t})$ as the variable $\mathbf{t}$ in $\mathcal{M}_{\ell\times r}$ varies over an affine manifold F, called the *support manifold*, containing the range Θ of the parameter θ. If unique, $\widehat{\theta} = \widehat{\theta}(F) \in F$ is called the *GLS estimator* (GLSE) for the parameter θ and in the case $F = \mathcal{M}_{\ell\times r}$, the *ordinary LSE* (OLSE).

A GLS value $\widehat{\theta} \in F$ always exists but may not be unique. In the one-model case ($k = 1$) when the regression in (1.2) has a conditional mean structure and the conditional response dispersion matrix has a variance components structure, the GLSE corresponding to the smallest manifold F containing Θ enjoys local and global optimality; see Bac-Van (1992, Theorem 5). This is the reason why we recommend the use of GLSE.

Definition 1.1 is inspired by one in multivariate fixed-design models (see Humak (1977, Satz 2.1.3, p. 34)) and is formulated progressively for the one-model case ($k = 1$) in Bac-Van (1992, 1994, 1998).

To specify the norm in Definition 1.1, let $\mathbf{z}_{ij}$ be given $r \times r$ positive definite (p.d.) matrices. Put

$$\mathbf{Z} = \operatorname{diag}(\mathbf{z}_{11}, \ldots, \mathbf{z}_{1a(1)}, \ldots, \mathbf{z}_{k1}, \ldots, \mathbf{z}_{ka(k)}), \quad s = a(1) + \ldots + a(k). \tag{1.4}$$

Then the norm for the global residual is a norm in $\mathbb{R}^{sr}$ defined by

$$\|\mathbf{u}\|_{\mathbf{Z}} = (\mathbf{u}'\mathbf{Z}\mathbf{u})^{1/2} \quad \forall \mathbf{u} \in \mathbb{R}^{sr}.$$

We are led to this norm, using a given transformation of the response $\mathbf{Y}_{ij}$ into $\mathbf{z}_{ij}^{1/2}\mathbf{Y}_{ij}$. In the situation with explanatory variables X_{ij} we shall choose an $r \times r$ p.d. matrix function $\mathbf{z}_i(\cdot)$ on H_i for each $i = 1, \ldots, k$ and put

$$\mathbf{z}_{ij} = \mathbf{z}_i(X_{ij}).$$

Then the transformation $\mathbf{Y}_{ij} \mapsto \mathbf{z}_i^{1/2}(X_{ij})\mathbf{Y}_{ij}$ is conditional given the explanatory X_{ij} paired off with $\mathbf{Y}_{ij}$.

Exploring the expression of the OLSE (see (3.18)) directly for the case of one univariate fixed-design model, namely (1.2) with $k = 1,\ r = 1$, the $\mathbf{b}_{1j}$ constant and $\mathbf{z}_{1j} = 1$, Lai *et al.* (1979) established the strong consistency of the OLSE under minimal assumptions on the design matrix; papers subsequent to this work include Chen *et al.* (1981), Bhat (1982), Chen *et al.* (1983), Wu and Wasan (1996), Jin and Chen (1996). In the random regressors case, also named stochastic regression models, to which the present paper belongs, Anderson and Taylor (1979), Christopeit and Helmes (1980), Lai and Wei (1982), Wei (1985) proved the strong consistency of the OLSE in the model ($k = 1$, $r = 1$)

$$\mathbf{Y}'_{1j} = \mathbf{b}'_{1j}\theta_1 + \eta'_{1j}, \quad j = 1, 2, \ldots,$$

(with our notations), where the $\mathbf{b}_{1j}$ are stochastic, $\mathbf{z}_{1j} = 1$; their conditions are imposed on the greatest and least eigenvalues of the random matrix $(\mathbf{b}_{11} \ldots \mathbf{b}_{1a(1)})(\mathbf{b}_{11} \ldots \mathbf{b}_{1a(1)})'$. Subsequent works on strong consistency of LSE in univariate stochastic regression models include Chang and Lin (1995), D. S. Chang and M. R. Chang (1996), with conditions imposed on the stochastic regressors and errors; in the latter work constrained LSE under linear constraints are also considered. In the system (1.2) with explanatory variables X_{ij} the regressors $\mathbf{b}_{ij}$ are functions of the explanatory observations, so to ensure the convergence and the strong consistency of GLSE we shall not try to seek deep conditions on the regressors; instead, we shall see how the true value of the parameter θ, i.e. the modelling value, influences the GLSE consistency: the point is that when in each ith model the observations (X_{ij}, Y_{ij}), $j = 1, 2, \ldots$, are identically distributed then given functions $\mathbf{b}_i(\cdot)$, there is a best modelling parameter value, viz. the global mean square (msq) regression parameter value (see Definition 2.1), and under some assumptions we shall show that the GLSE converges to the orthogonal projection of this best value on the support manifold, so the GLSE is strongly consistent if and only if the true parameter value coincides with this projection. In this way the results in Bac-Van (1994, 1998) for one-model systems will be completed and extended to systems of k models. The range Θ of θ may have an arbitrary shape, which means arbitrary constraints may be imposed on the parameters $\theta_1, \ldots, \theta_k$. A new feature of the present paper and also, as seen in Section 4, the main mathematical difficulty lies in the investigation of GLSE, instead of OLSE, in systems of $k \geq 2$ models, whereas previous works on strong consistency of LSE examined one-model systems ($k = 1$) exclusively; *all results of this paper are valid for any mutually dependent families of observations* $\{(X_{i1}, \mathbf{Y}_{i1}), (X_{i2}, \mathbf{Y}_{i2}), \ldots\}$ in several models ($i = 1, \ldots, k$). Of course with the block matrices in Subsection 3.1 we can write the system of k models (1.2) as a single model where the set of observations disintegrates into k dependent groups, but such a reduction is purely formal. In fact, the treatment of systems of $k \geq 2$ models requires the development of a mathematical tool based on sign properties of coefficients in the determinant expansion for linear combinations of matrices; this tool, which has not been used yet in the study of LSE convergence and consistency, is presented in Subsection 4.3 which shows that exploring GLSE convergence for systems of $k \geq 2$ models is indeed a problem on the interface of matrix theory and statistics.

Starting from the integral expression of $\widehat{\theta}(F) - \mathbf{t}$ for $\mathbf{t}$ arbitrary in the support manifold, which expression, by an isomorphism, has the form $A\beta$, where β is simply expressed through the global residual and A is a random linear operator, our approach to proving the GLSE convergence consists in discovering the large sample a.s. uniform boundedness of A which yields necessary and sufficient conditions for the convergence as well as for the strong consistency of GLSE in multimodel systems. These results shed light on the one-model ($k = 1$) case; cf. Theorem 2.1 and its Corollary 2.1. A comparison of the preceding approach with another standard one is given in Subsection 3.6, Remark 3.3.

The paper is organized as follows. Section 2 presents the results. Section 3 gives proofs but defers the algebraic treatment to Sections 4 and 5. Section 4 deals with a crucial lemma for exploring the expression of GLSE. Section 5 treats the positivity of sums of mixing determinants, which is basic for the mathematical method developed in Section 4.

1.2. Notations and preliminaries. In a linear space an *affine manifold* L is a set such that $\{\mathbf{u} - \mathbf{v} : \mathbf{u}, \mathbf{v} \in L\}$ is a subspace K. Then $K = L - \mathbf{t}$, $L = K + \mathbf{t}$, where $\mathbf{t}$ is a fixed element of L. The subspace K is called *parallel* to L.

Throughout the paper the following symbols are constantly used. For $\mathbf{x}, \mathbf{y} \in \mathcal{M}_{p\times q}$, $\mathbf{x} = (\mathbf{x}_1 \dots \mathbf{x}_q)$, $\mathbf{y} = (\mathbf{y}_1 \dots \mathbf{y}_p)'$ and $K \subset \mathcal{M}_{p\times q}$:

$$[\mathbf{y}] = (\mathbf{y}_1' \dots \mathbf{y}_p')', \quad [K] = \{[\mathbf{y}] : \mathbf{y} \in K\}, \tag{1.5}$$

$$\operatorname{vec} \mathbf{x} = (\mathbf{x}_1' \dots \mathbf{x}_q')'. \tag{1.6}$$

Given a $p \times q$ real matrix $\mathbf{C} = (c_{ij})$,

$$\begin{aligned} \mathbf{C}(\sigma) &= (c_{ij})_{i,j\in\sigma} \quad \text{for } \sigma \subset \{1, \dots, \min(p, q)\}, \\ \mathbf{C}_{fg} &= \text{submatrix obtained by deleting the } f\text{th row and } g\text{th column of } \mathbf{C}, \\ \|\mathbf{C}\| &= \Big(\sum c_{ij}^2\Big)^{1/2}, \end{aligned} \tag{1.7}$$

$$\begin{aligned} \mathcal{M}(\mathbf{C}) &= \text{linear space generated by the columns of } \mathbf{C}, \\ \ker \mathbf{C} &= \{\mathbf{x} \in \mathbb{R}^q : \mathbf{C}\mathbf{x} = 0\}, \\ \operatorname{Span} Q &= \text{linear hull of the set } Q \text{ in a linear space}, \\ Q^\perp &= \text{orthogonal complement of the set } Q \text{ in a Euclidean space}, \\ \sharp\varphi &= \text{cardinality of the set } \varphi, \\ \mathbf{I} &= \text{unit matrix, n.n.d.} = \text{non-negative definite}, \end{aligned} \tag{1.8}$$

$$\ell = \ell(1) + \ldots + \ell(k) \text{ in the models (1.2)}, \tag{1.9}$$

$$\varPhi = \text{the subspace of } \mathbb{R}^{\ell r} \text{ parallel to } [F] \text{ in Definition 1.1}, \tag{1.10}$$

$$\mathbf{E} = \text{a matrix whose columns form a basis for } \varPhi, \text{ with} \tag{1.11}$$

$$\mathbf{E} = (\mathbf{E}_1' \dots \mathbf{E}_k')', \quad \mathbf{E}_i = \text{a matrix of } \ell(i)r \text{ rows}. \tag{1.12}$$

For $\mathcal{M}_{p\times q}$ endowed with the inner product $(\mathbf{y}, \mathbf{z}) = [\mathbf{y}]'\mathbf{W}[\mathbf{z}]$, $\mathbf{W}$ $pq \times pq$ p.d., $\mathbf{y}, \mathbf{z} \in \mathcal{M}_{p\times q}$:

$$\begin{aligned} \|\mathbf{y}\|_{\mathbf{W}} &= ([\mathbf{y}]'\mathbf{W}[\mathbf{y}])^{1/2}, \\ \perp_{\mathbf{W}} &= \text{orthogonality symbol (use } \perp \text{ when } \mathbf{W} = \mathbf{I}). \end{aligned}$$

The orthogonal projection of $\mathbf{y} \in \mathcal{M}_{p\times q}$ on an affine manifold $L \subset \mathcal{M}_{p\times q}$ is defined as the unique element $\mathbf{x} \in L$ such that $\|\mathbf{y} - \mathbf{x}\|_{\mathbf{W}} \le \|\mathbf{y} - \mathbf{u}\|_{\mathbf{W}}$ for all $\mathbf{u} \in L$, and

$$\begin{aligned} \mathrm{Pr}_L^{\mathbf{W}} &= \text{orthogonal projector onto the affine manifold } L, \\ \mathrm{Pr}_L &= \mathrm{Pr}_L^{\mathbf{I}}. \end{aligned}$$

The following formula defines $\mathrm{Pr}_L^{\mathbf{W}}$ through the orthogonal projector onto the subspace $L-\mathbf{t}$, $\mathbf{t}\in L$:

(1.13) $$\mathrm{Pr}_L^{\mathbf{W}}\mathbf{A}-\mathbf{t}=\mathrm{Pr}_{L-\mathbf{t}}^{\mathbf{W}}(\mathbf{A}-\mathbf{t}),\quad \mathbf{A}\in\mathcal{M}_{p\times q},$$

which is easily obtained using a translation by $-\mathbf{t}$; cf. Bac-Van (1994, Proposition 3.4).

2. Results on GLSE convergence and consistency

Henceforth throughout the paper the following assumption will always be in force.

ASSUMPTION 2.1. • For $i=1,\ldots,k$, $(H_i,\mathcal{A}_i)$ are arbitrary measurable spaces,

• for $j=1,2,\ldots,$ X_{ij} are $(H_i,\mathcal{A}_i)$-valued random variables (r.v.) and $\mathbf{Y}_{ij}$ are $r\times 1$ random vector variables,

• $\mathbf{b}_i(\cdot)$ and $\mathbf{z}_i(\cdot)$ are $\mathcal{A}_i$-measurable functions on H_i; $\mathbf{b}_i(\cdot)$ is $\ell(i)\times 1$ vector-valued whereas $\mathbf{z}_i(\cdot)$ is $r\times r$ p.d. matrix-valued;

• the $\ell(i)\times r$ real matrix parameters θ_i and the global sample size

$$a=(a(1),\ldots,a(k))$$

are non-random. We put

$$(X_i,\mathbf{Y}_i)=(X_{i1},\mathbf{Y}_{i1})\quad\forall i.$$

It will turn out that the GLSE convergence and consistency are related to the notion defined below; cf. Cramér (1945, p. 272). Write

$$\forall\mathbf{y}\in\mathbb{R}^r,\quad\forall x\in H_i,\quad\|\mathbf{y}\|^2_{\mathbf{z}_i(x)}=\mathbf{y}'\mathbf{z}_i(x)\mathbf{y}.$$

DEFINITION 2.1. If the non-random matrix τ_i satisfies

$$\mathrm{E}\,\|\mathbf{Y}_i-\tau_i'\mathbf{b}_i(X_i)\|^2_{\mathbf{z}_i(X_i)}\le\mathrm{E}\,\|\mathbf{Y}_i-\mathbf{t}_i'\mathbf{b}_i(X_i)\|^2_{\mathbf{z}_i(X_i)}$$

for all $\ell(i)\times r$ non-random matrices $\mathbf{t}_i$, then the function $\tau_i'\mathbf{b}_i(X_i)$ is called the *mean square* (*msq*) *regression* of $\mathbf{Y}_i$ on X_i, and τ_i a *msq regression of* $\mathbf{Y}_i$ *on* X_i *parameter value.* $\tau=(\tau_1'\ldots\tau_k')'$ is named a *global msq regression parameter value.*

We shall repeatedly use the notion of stationary and indecomposable sequences of r.v.'s of Loève (1963, §30); here this notion is introduced to ensure that the ergodic hypothesis (*loc. cit.*, p. 423) is true. Let $(\mathit{\Gamma},\mathcal{C})$ be an arbitrary measurable space. A matrix-valued r.v. ξ, with range space isomorphic to $(\mathbb{R}^p,\mathcal{B}^p)$, is said to be *defined* on a sequence of $(\mathit{\Gamma},\mathcal{C})$-valued r.v.'s $Z_n,\ n=1,2,\ldots$ if $[\xi]=\varphi(\{Z_n\})$, where φ is a measurable function from $(\mathit{\Gamma}^\infty,\mathcal{C}^\infty)$ to $(\mathbb{R}^p,\mathcal{B}^p)$. An event A is defined on the sequence if so is I_A. The *translate* ξ_k by $k-1$ of $\xi=\varphi(Z_1,Z_2,\ldots)$ is $\xi_k=\varphi(Z_k,Z_{k+1},\ldots)$ for $k\ge 1$, the translate A_k of A is defined by $I_{A_k}=(I_A)_k$. An event A is called an *invariant event* of $\{Z_n\}$ if $A_k=A$ for all $k\ge 1$. The sequence $\{Z_n\}$ is called *stationary* if $\mathrm{P}A_k=\mathrm{P}A$ for all $k\ge 1$ for every event A defined on it and is called *indecomposable* if $\mathrm{P}A=0$ or 1 for every invariant event A of the sequence.

PROPOSITION 2.1 (Ergodic theorem). *Let ξ be a matrix-valued r.v. defined on a stationary indecomposable sequence $\{Z_n\}$ and $\{\xi_k\}$ the sequence of translates. Let the matrix $\mathrm{E}\,\xi$ exist and have finite elements. Then $h^{-1}\sum_{k=1}^h\xi_k\xrightarrow{a.s.}\mathrm{E}\,\xi$ as $h\to+\infty$.*

In what follows, under Assumption 2.1 the following conditions will often be required.

A0. For each $i = 1, \dots, k$, $\{(X_{i1}, \mathbf{Y}_{i1}), (X_{i2}, \mathbf{Y}_{i2}), \dots\}$ is a stationary and indecomposable sequence.

The strict stationarity is required for defining a msq regression parameter value of the response on the explanatory variable in the ith model.

A1. $\mathrm{E}\, \|\mathbf{z}_i^{1/2}(X_i)\mathbf{Y}_i\|^2 < \infty$ for all $i = 1, \dots, k$.

A2. $\mathrm{E}\, \|\mathbf{b}_i(X_i)\|^2 \mathrm{Tr}\, \mathbf{z}_i(X_i) < \infty$ for all $i = 1, \dots, k$, where Tr abbreviates Trace.

A3. For each $i = 1, \dots, k$, the probability distribution of $\mathbf{b}_i(X_i)$ is not concentrated in any proper subspace of $\mathbb{R}^{\ell(i)}$.

To state the convergence of GLSE we shall use the matrices $\mathbf{E}, \mathbf{E}_i$ of (1.11), (1.12) and the subspace $\varPhi$ of $\mathbb{R}^{\ell r}$ in (1.9), (1.10). Let $\mathbf{v} = (\mathbf{v}_1' \dots \mathbf{v}_k')'$ be any vector in $\mathbb{R}^{\ell r}$, $\mathbf{v}_i \in \mathbb{R}^{\ell(i) r}$ for all i. Put

$$\begin{aligned}
\mathbf{T}_{0i} &= \mathrm{E}\, (\mathbf{b}_i(X_i)\mathbf{b}_i'(X_i) \otimes \mathbf{z}_i(X_i)), \text{ a n.n.d. matrix } \forall i,\\
\mathbf{T}_0 &= \mathrm{diag}\, (\mathbf{T}_{01}, \dots, \mathbf{T}_{0k}),\\
S(F) &= \{\mathbf{v} \in \mathbb{R}^{\ell r} : \mathbf{E}_i'\mathbf{T}_{0i}\mathbf{v}_i = 0 \ \forall i = 1, \dots, k\},\\
[U(F)] &= S(F) \oplus \varPhi + [\mathbf{t}] \quad \text{with an arbitrary matrix } \mathbf{t} \in F.
\end{aligned}$$

$S(F)$ is independent of the choice of E, a matrix generating $\varPhi$, and is thus determined by F. Under Conditions A2 and A3, $\mathbf{T}_{0i}$ and $\mathbf{T}_0$ are p.d. matrices; cf. Proposition 3.2. With the inner product $\mathbf{u}'\mathbf{T}_0\mathbf{v}$ in $\mathbb{R}^{\ell r}$, $S(F) \subset \varPhi^\perp$. $U(F) \subset \mathcal{M}_{\ell \times r}$ is an affine manifold independent of $\mathbf{t} \in F$ since $[\mathbf{t}] \in [F] = \varPhi + [\mathbf{t}] \subset [U(F)]$.

Properties of $U(F)$. (i) $U(F) \supset F$, whereas $U(F) = F \Leftrightarrow \sum_{i=1}^k \mathrm{Rank}\, \mathbf{E}_i = \ell r$ or, equivalently, using ℓr orthogonal coordinate axes in $\mathbb{R}^{\ell r} = \prod_{i=1}^k \mathbb{R}^{\ell(i) r}$, the projection of $\varPhi$ on $\mathbb{R}^{\ell(i) r}$ coincides with $\mathbb{R}^{\ell(i) r}$ for all i.

(ii) $U(F) \subset \mathcal{M}_{\ell \times r}$, and

$$U(F) = \mathcal{M}_{\ell \times r} \;\Leftrightarrow\; \mathrm{Rank}\, \mathbf{E} = \sum_{i=1}^k \mathrm{Rank}\, \mathbf{E}_i \tag{2.1}$$

or, equivalently, $\varPhi$ is the product of its projections on the coordinate spaces $\mathbb{R}^{\ell(i) r}$.

(iii) The case $F \subset U(F) \subset \mathcal{M}_{\ell \times r}$ with proper inclusion really occurs if and only if $\mathrm{Rank}\, \mathbf{E} < \sum_{i=1}^k \mathrm{Rank}\, \mathbf{E}_i < \ell r$.

THEOREM 2.1 (GLSE convergence). *In the system* (1.2) *satisfying Assumption* 2.1 *and Conditions* A0–A3, *if the global msq regression parameter value* $\tau \in U(F)$ *then GLSE* $\widehat{\theta}(F) \overset{a.s.}{\to} \mathrm{Pr}_F^{\mathbf{T}_0}\tau$ *as* $a \to \infty$, *whereas* $\widehat{\theta}(F)$ *diverges a.s. when* $\tau \notin U(F)$.

The GLSE is strongly consistent if and only if together $\tau \in U(F)$ *and the projection* $\mathrm{Pr}_F^{\mathbf{T}_0}\tau$ *coincides with the true value of the parameter* θ.

By (2.1) we get the following important corollary.

COROLLARY 2.1. *For a spline model with random explanatory variables, put in the linear form*

$$\mathbf{Y}_{1j}' = \mathbf{b}_1'(X_{1j})\theta + \eta_{1j}', \quad j = 1, \dots, a,$$

satisfying Assumption 2.1 *and Conditions* A0–A3 *with* $k = 1$, *as* $a \to \infty$ *the GLSE* $\widehat{\theta}(F)$ *always converges a.s. to the orthogonal projection on* F *of the msq regression parameter value* τ *according to the inner product*

$$(\mathbf{u}, \mathbf{v}) = [\mathbf{u}]'\mathbf{T}_0[\mathbf{v}], \quad \mathbf{u}, \mathbf{v} \in \mathcal{M}_{\ell\times r}, \ \mathbf{T}_0 = \mathrm{E}\,(\mathbf{b}_1(X_{11})\mathbf{b}_1'(X_{11}) \otimes \mathbf{z}_1(X_{11})).$$

The GLSE is strongly consistent if and only if this projection coincides with the true value of the parameter θ.

When F is the whole parameter range space $\mathcal{M}_{\ell\times r}$ we get

Corollary 2.2. *Under Assumption* 2.1 *and Conditions* A0–A3 *the OLSE always converges a.s. to* τ. *The OLSE is strongly consistent if and only if the true value of the parameter* θ *coincides with* τ.

We also get

Corollary 2.3. *Under Assumption* 2.1 *and Conditions* A0–A3 *the GLSE* $\widehat{\theta}$ *is always strongly consistent in the following case:*

$$\mathrm{E}\,(\mathbf{Y}_i \mid X_i) = \theta_i'\mathbf{b}_i(X_i) \quad (\forall i) \text{ in the system } (1.2). \tag{2.2}$$

Indeed, by the minimal property of the conditional expectation, (2.2) entails that θ_i is the msq regression parameter value for all i, so $\tau = \theta \in \Theta \subset F \subset U(F)$.

According to the proof of Lemma 3.2 Conditions A1–A3 are needed to ensure that a msq regression of $\mathbf{Y}_i$ on X_i parameter value exists and is unique, $i = 1, \ldots, k$. In the one-model case, for i.i.d. observations, Condition A3 is necessary and sufficient for the a.s. existence of the GLSE for sufficiently large sample size (see Bac-Van (1994, Theorem 5.1)). Thus at least for stationary and indecomposable sequences of observations,Theorem 2.1 gives necessary and sufficient conditions for the GLSE convergence and strong consistency under minimal assumptions and characterizes these properties by msq regression.

In Definition 1.1 of GLS values, there may be an infinity of manifolds F containing the parameter range Θ. In the course of the estimation process one may wish to change the support F. Does it influence the convergence of $\widehat{\theta}(F)$? To get an answer, we shall build a class of manifolds F, using the following lemma and the definition below.

Lemma 2.1. *Let* $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$, $m \geq 1$, *be a basis for a subspace* Φ *of* $\mathbb{R}^n$ *and* $\{\mathbf{f}_{1K}, \ldots, \mathbf{f}_{mK}\}$ *its orthogonal projection on some subspace* K *of* $\mathbb{R}^n$ *according to the inner product* $\mathbf{u}'\mathbf{v}$, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. *Then the ratio*

$$\sigma = 1 - \frac{\det(\mathbf{f}_{1K} \ldots \mathbf{f}_{mK})'(\mathbf{f}_{1K} \ldots \mathbf{f}_{mK})}{\det(\mathbf{f}_1 \ldots \mathbf{f}_m)'(\mathbf{f}_1 \ldots \mathbf{f}_m)}$$

is independent of the choice of the basis for Φ. *Moreover* $0 \leq \sigma \leq 1$, *and* $\sigma = 0$ *if and only if* $\Phi \subset K$, *while* $\sigma = 1$ *if and only if* Φ *contains some non-null vector orthogonal to* K.

In Section 5 we shall prove this lemma and justify the following definition.

Definition 2.2. For $0 \leq \delta < 1$, the subspace Φ is said to be *at most* δ*-steep* relative to K if $1 \leq \dim \Phi \leq \dim K$ and if $\sigma \leq \delta$.

An affine manifold F in $\mathcal{M}_{\ell\times r}$ is said to be at most δ-steep relative to some subspace G of $\mathcal{M}_{\ell\times r}$ if, in $\mathbb{R}^n = \mathbb{R}^{\ell r}$, the subspace Φ parallel to $[F]$ is at most δ-steep relative to the subspace $K = [G]$. By abuse of language, we shall say that the affine manifold F is at most δ-steep on some coordinate space $\mathcal{M}_{\ell(i)\times r}$ of $\mathcal{M}_{\ell\times r}$, $i = 1,\dots,k$, $\ell = \ell(1)+\dots+\ell(k)$, if in $\mathbb{R}^{\ell r}$ the subspace Φ parallel to $[F]$ is at most δ-steep relative to the subspace K consisting of $\ell r\times 1$ vectors lying in the coordinate space $\mathbb{R}^{\ell(i)r}$ of $\mathbb{R}^{\ell r}$—a vector $\mathbf{x} = (\mathbf{x}_1,\dots,\mathbf{x}_k) \in \prod_{i=1}^k \mathbb{R}^{n_i}$ is said to lie in the coordinate space $\mathbb{R}^{n_i}$ if $\mathbf{x}_h = \mathbf{0}$ for all $h \neq i$.

Given numbers δ_i, $0 \le \delta_i < 1$, $i = 1,\dots,k$, consider the following class $\mathcal{C}$ of affine manifolds $F \in \mathcal{M}_{\ell\times r}$:

$$\mathcal{C} = \{F \supset \Theta : F \text{ at most } \delta_i\text{-steep on } \mathcal{M}_{\ell(i)\times r}\ \forall i\}. \tag{2.3}$$

In the one-model case ($k=1$), $\mathcal{C} = \{F : \Theta \subset F \subset \mathcal{M}_{\ell\times r}\}$ regardless of δ_1.

Theorem 2.2 (Uniform convergence of GLSE). *Under Assumption* 2.1 *and Conditions* A0–A3 *if the global parameter range Θ in* (1.2) *contains the global msq regression parameter value τ and if $\mathcal{C}$ is non-void, then*

$$\sup_{F\in\mathcal{C}} \|\widehat{\theta}(F) - \tau\| \xrightarrow{a.s.} 0 \quad \text{as } a \to \infty.$$

Theorem 2.3 (GLSE uniform consistency). *Under Assumption* 2.1 *let Conditions* A0–A3 *be fulfilled. Let there exist a manifold $F \ni \tau$, $F \in \mathcal{C}$. Then*

$$\theta = \tau \ \Leftrightarrow\ \sup_{F\in\mathcal{C}} \|\widehat{\theta}(F) - \theta\| \xrightarrow{a.s.} 0 \text{ as } a \to \infty,$$

i.e. the GLSE $\widehat{\theta}(F)$ is strongly consistent uniformly in F over the class $\mathcal{C}$ if and only if in (1.2) *the true global parameter value θ coincides with the global msq regression parameter value τ.*

In the one-model case ($k=1$) this conclusion is already stated in another form in Bac-Van (1994, Remark 5.2) for i.i.d. observations.

In the case (2.2) always $F \ni \tau$, hence

Corollary 2.4. *In the case* (2.2) *if $\mathcal{C}$ is non-void, under* A0–A3 *the GLSE $\widehat{\theta}(F)$ is always strongly consistent uniformly in F over $\mathcal{C}$.*

The belonging of F to the class $\mathcal{C}$ implies that there are sufficiently many constraints on the parameters θ_i so that the dimension of Θ does not exceed that of the range space of any θ_i.

3. Proofs of the results

3.1. Algebraic formal expressions For $i = 1,\dots,k$ consider

$$\{(\mathbf{Y}_{ij}, \mathbf{b}_{ij}), \mathbf{z}_{ij} : j = 1,\dots,a(i)\},$$

$\mathbf{Y}_{ij}$ and $\mathbf{b}_{ij}$ being respectively $r\times 1$ and $\ell(i)\times 1$ vectors, and $\mathbf{z}_{ij}$ $r\times r$ p.d. matrices. $\mathbf{Z}$ is defined by (1.4). $\otimes$ being the Kronecker product symbol, $\mathbf{I}_r$ the $r\times r$ unit matrix, we form block matrices

$$
\begin{aligned}
\mathbf{Y} &= (\mathbf{Y}'_{11} \dots \mathbf{Y}'_{1a(1)} \vdots \dots \vdots \mathbf{Y}'_{k1} \dots \mathbf{Y}'_{ka(k)})', \\
\mathbf{B}'_i &= (\mathbf{b}_{i1} \dots \mathbf{b}_{ia(i)}), \\
\mathbf{B} &= \mathrm{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k), \\
\mathbf{C} &= \mathbf{Z}^{1/2}(\mathbf{B} \otimes \mathbf{I}_r), \\
\boldsymbol{\Delta} &= \mathrm{diag}(a(1)\mathbf{I}_{\ell(1)r}, \dots, a(k)\mathbf{I}_{\ell(k)r}), \\
\mathbf{T} &= \boldsymbol{\Delta}^{-1}\mathbf{C}'\mathbf{C} = \mathrm{diag}(\mathbf{T}_1, \dots, \mathbf{T}_k),
\end{aligned} \tag{3.1}
$$

$$\mathbf{C}'\mathbf{C} = \boldsymbol{\Delta}\mathbf{T}. \tag{3.2}$$

We set

$$\ell = \ell(1) + \dots + \ell(k), \quad s = a(1) + \dots + a(k);$$

then $\mathbf{C}$ is $sr \times \ell r$, $\mathbf{T}_i$ is $\ell(i)r \times \ell(i)r$. Noting that, for any vector $\mathbf{b}$ and matrices $\mathbf{A}, \mathbf{U}, \mathbf{V}$ such that the products below exist,

$$(\mathbf{b} \otimes \mathbf{A})\mathbf{U} = \mathbf{b} \otimes \mathbf{AU}, \quad \mathbf{V}(\mathbf{b}' \otimes \mathbf{A}') = \mathbf{b}' \otimes \mathbf{VA}', \tag{3.3}$$

we have $a(i)\mathbf{T}_i = (\dots \mathbf{b}_{ij} \otimes \mathbf{I}_r \dots)(\dots \mathbf{b}_{ij} \otimes \mathbf{z}_{ij} \dots)'$, thus

$$\mathbf{T}_i = a^{-1}(i) \sum_{j=1}^{a(i)} \mathbf{b}_{ij}\mathbf{b}'_{ij} \otimes \mathbf{z}_{ij}. \tag{3.4}$$

We rewrite the global residual $\eta(\cdot)$ in (1.3) and define a vector function $\gamma(\cdot)$, which will be extensively used later, as follows:

$$\eta(\theta) = \mathbf{Y} - [\mathbf{B}\theta], \tag{3.5}$$

$$\gamma(\theta) = \boldsymbol{\Delta}^{-1}\mathbf{C}'\mathbf{Z}^{1/2}\eta(\theta) = (\gamma'_1 \dots \gamma'_k)', \tag{3.6}$$

where, using (3.1) and (1.3),

$$\gamma'_i = a^{-1}(i)(\dots \mathbf{Y}'_{ij} - \mathbf{b}'_{ij}\theta_i \dots)\left(\dots (\mathbf{b}_{ij} \otimes \mathbf{I}_r)\mathbf{z}_{ij} \dots\right)', \quad j = 1, \dots, a(i),$$

equivalently

$$\gamma'_i = \gamma'_i(\theta_i) = a^{-1}(i) \sum_{j=1}^{a(i)} (\mathbf{Y}'_{ij} - \mathbf{b}'_{ij}\theta_i)\mathbf{z}_{ij}(\mathbf{b}'_{ij} \otimes \mathbf{I}_r).$$

Note that $\mathbf{I}$ being the unit matrix, for matrices $\mathbf{A}, \mathbf{U}, \mathbf{V}, \mathbf{q}, \mathbf{z}$ with $\mathbf{z}$ symmetric and any vector $\mathbf{b}$ such that the products below exist, using (1.5) we have

$$[\mathbf{AUV}'] = (\mathbf{A} \otimes \mathbf{V})[\mathbf{U}], \tag{3.7}$$

$$\mathbf{q}'\mathbf{b} = [\mathbf{b}'\mathbf{q}] = (\mathbf{b}' \otimes \mathbf{I})[\mathbf{q}], \tag{3.8}$$

$$(\mathbf{b} \otimes \mathbf{z})\mathbf{q}'\mathbf{b} = (\mathbf{b} \otimes \mathbf{z})(\mathbf{b}' \otimes \mathbf{I})[\mathbf{q}] = (\mathbf{bb}' \otimes \mathbf{z})[\mathbf{q}]. \tag{3.9}$$

Hence

$$\gamma'_i(\theta_i) = a^{-1}(i) \sum_{j=1}^{a(i)} (\mathbf{Y}'_{ij} - \mathbf{b}'_{ij}\theta_i)(\mathbf{b}'_{ij} \otimes \mathbf{z}_{ij}),$$

$$\gamma_i(\theta_i) = a^{-1}(i)\sum_{j=1}^{a(i)}\left((\mathbf{b}_{ij}\otimes\mathbf{z}_{ij})\mathbf{Y}_{ij} - (\mathbf{b}_{ij}\mathbf{b}'_{ij}\otimes\mathbf{z}_{ij})[\theta_i]\right). \tag{3.10}$$

Note some useful expressions for γ. Put

$$\mathbf{Q}_i = a^{-1}(i)(\mathbf{b}_{i1}\otimes\mathbf{z}_{i1}\vdots\ldots\vdots\mathbf{b}_{ia(i)}\otimes\mathbf{z}_{ia(i)}),\quad \mathbf{Q} = \operatorname{diag}(\mathbf{Q}_1,\ldots,\mathbf{Q}_k).$$

Then $\gamma_i(\theta_i) = \mathbf{Q}_i(\mathbf{Y}'_{i1}\ldots\mathbf{Y}'_{ia(i)})' - \mathbf{T}_i[\theta_i]$, hence the vector $\gamma(\theta)$ linearly depends on θ, and is also a linear function of the global residual η, namely

$$\gamma(\theta) = \mathbf{QY} - \mathbf{T}[\theta] = \mathbf{Q}\eta(\theta). \tag{3.11}$$

Henceforth, viewing $\mathbf{A}\in\mathcal{M}_{p\times q}$ as a linear map, for any set $\Psi\subset\mathcal{M}_{q\times r}$ we shall often use the symbol

$$\mathbf{A}\Psi = \{\mathbf{Ax} : \mathbf{x}\in\Psi\}\subset\mathcal{M}_{p\times r}. \tag{3.12}$$

3.2. Elicitation of the GLSE. From Definition 1.1 for any $\mathbf{t}\in F\subset\mathcal{M}_{\ell\times r}$, using (1.5) consider in $\mathbb{R}^{\ell r}$ the subspace $\Phi = [F-\mathbf{t}]$ parallel to $[F]$. $\mathbf{C}$ being $sr\times\ell r$, by (3.12), $\mathbf{C}\Phi\subset\mathcal{M}(\mathbf{C})\subset\mathbb{R}^{sr}$. Generalizing the three-perpendicular theorem we can check that in $\mathbb{R}^{sr}$,

$$\operatorname{Pr}_{\mathbf{C}\Phi} = \operatorname{Pr}_{\mathbf{C}\Phi}\operatorname{Pr}_{\mathcal{M}(\mathbf{C})}. \tag{3.13}$$

Using the inner product $\mathbf{u}'\mathbf{v}$ $(\mathbf{u},\mathbf{v}\in\mathbb{R}^{sr})$ we introduce two constantly used symbols

$$\mathbf{J} = \operatorname{Pr}_{\mathbf{C}\Phi}\quad\text{in } \mathbb{R}^{sr},$$
$$\mathbf{G} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{JC}(\mathbf{C}'\mathbf{C})^{-1}. \tag{3.14}$$

Lemma 3.1. *Choose* $\mathbf{t}$ *arbitrarily in* F. *Let* β *be any vector defined by* $\mathbf{T}\beta = \gamma(\mathbf{t})$. *Then GLS values* $\widehat{\theta}$ *are defined by the equation*

$$\mathbf{C}[\widehat{\theta}-\mathbf{t}] = \mathbf{JC}\beta.$$

The GLSE $\widehat{\theta}(F)$ *exists if and only if* $\mathbf{C}'\mathbf{C}$ *is p.d., and then*

$$[\widehat{\theta}(F)-\mathbf{t}] = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{JC}(\mathbf{C}'\mathbf{C})^{-1}\boldsymbol{\Delta}\gamma(\mathbf{t}) = \mathbf{G}\boldsymbol{\Delta}\gamma(\mathbf{t}) = \mathbf{G}\boldsymbol{\Delta}\mathbf{T}\beta\quad \textit{for any } t\in F, \tag{3.15}$$

$$\beta = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{Z}^{1/2}\eta(\mathbf{t}). \tag{3.16}$$

Remark 3.1. The classic univariate multiple regression model is $\mathbf{Y} = \mathbf{B}\theta + \eta(\theta)$, where $\mathbf{Y}$ is the $a\times 1$ response data, $\mathbf{B}$ is the $a\times\ell$ design matrix, θ is the $\ell\times 1$ parameter and $\eta(\theta) = \mathbf{Y}-\mathbf{B}\theta$ is the residual; we recognize that here $k=1$, $r=1$, $a=a(1)=s$, $\boldsymbol{\Delta} = a\mathbf{I}_\ell$, and taking $\mathbf{Z} = \mathbf{I}_a$ as usual, we have $\mathbf{C} = \mathbf{B}$. Let us compute the OLSE: we have $\Phi = \mathbb{R}^\ell$, $\mathbf{C}\Phi = \mathbf{B}\mathbb{R}^\ell = \mathcal{M}(\mathbf{B})$, hence from (3.14),

$$\mathbf{J} = \operatorname{Pr}_{\mathcal{M}(\mathbf{B})}\quad\text{in } \mathbb{R}^a,\quad \mathbf{G} = (\mathbf{B}'\mathbf{B})^{-1} \tag{3.17}$$

since $\mathbf{JB} = \mathbf{B}$; from (3.6), $\boldsymbol{\Delta}\gamma(\mathbf{t}) = \mathbf{B}'\eta(\mathbf{t})$, and, by letting $\mathbf{t}=0$, (3.15) becomes the known expression

$$\widehat{\theta} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{Y}. \tag{3.18}$$

Proof of Lemma 3.1. From Definition 1.1, using the expression (3.5) for the global residual, GLS values $\widehat{\theta}$ are defined by

$$\|\mathbf{Y}-[\mathbf{B}\widehat{\theta}]\|_{\mathbf{Z}} = \min_{\mathbf{t}\in F}\|\mathbf{Y}-[\mathbf{Bt}]\|_{\mathbf{Z}},$$

which, with Notation (3.12), is equivalent to $[\mathbf{B}\widehat{\theta}] = \mathrm{Pr}^{\mathbf{Z}}_{[\mathbf{B}F]}\mathbf{Y}$. Hence GLS values $\widehat{\theta} \in F$ always exist. By a translation along $-t$ in $\mathcal{M}_{\ell\times r}$ we have the equivalent equation

$$[\mathbf{B}(\widehat{\theta} - \mathbf{t})] = \mathrm{Pr}^{\mathbf{Z}}_{[\mathbf{B}(F-t)]}(\mathbf{Y} - [\mathbf{B}\mathbf{t}]).$$

But, using (3.7), $[\mathbf{B}(\widehat{\theta} - \mathbf{t})] = (\mathbf{B} \otimes \mathbf{I}_r)[\widehat{\theta} - \mathbf{t}]$, $[\mathbf{B}(F - \mathbf{t})] = (\mathbf{B} \otimes \mathbf{I}_r)[F - \mathbf{t}] = (\mathbf{B} \otimes \mathbf{I}_r)\varPhi$, so by (3.5) we can rewrite

$$(\mathbf{B} \otimes \mathbf{I}_r)[\widehat{\theta} - \mathbf{t}] = \mathrm{Pr}^{\mathbf{Z}}_{(\mathbf{B}\otimes\mathbf{I}_r)\varPhi}\eta(\mathbf{t}).$$

Recall $\mathbf{C} = \mathbf{Z}^{1/2}(\mathbf{B} \otimes \mathbf{I}_r)$. The automorphism $\mathbf{y} \mapsto \mathbf{u} = \mathbf{Z}^{1/2}\mathbf{y}$ of $\mathbb{R}^{sr}$ transforms the inner product $\mathbf{y}'\mathbf{Z}\mathbf{z}$ into the inner product $\mathbf{u}'\mathbf{v}$, $\mathbf{u} = \mathbf{Z}^{1/2}\mathbf{y}$, $\mathbf{v} = \mathbf{Z}^{1/2}\mathbf{z}$. This yields the equivalent equation

$$\mathbf{C}[\widehat{\theta} - \mathbf{t}] = \mathrm{Pr}_{\mathbf{C}\varPhi}\mathbf{Z}^{1/2}\eta(\mathbf{t}). \tag{3.19}$$

By (3.13) this can be rewritten as

$$\mathbf{C}[\widehat{\theta} - \mathbf{t}] = \mathrm{Pr}_{\mathbf{C}\varPhi}\,\mathrm{Pr}_{\mathcal{M}(\mathbf{C})}\mathbf{Z}^{1/2}\eta(\mathbf{t}) = \mathbf{J}\mathbf{C}\beta,$$

where the vector β is defined by the orthogonality condition

$$\mathbf{C}'(\mathbf{Z}^{1/2}\eta(\mathbf{t}) - \mathbf{C}\beta) = \mathbf{0}$$

or, equivalently, by $\boldsymbol{\Delta}\gamma(\mathbf{t}) = \mathbf{C}'\mathbf{C}\beta$ due to (3.6), i.e. $\gamma(\mathbf{t}) = \mathbf{T}\beta$ by (3.2). $\mathbf{C}$ being $sr \times \ell r$, equation (3.19) has a unique solution $[\widehat{\theta}]$ if and only if $\mathrm{Rank}\,\mathbf{C} = \ell r$, i.e. $\mathbf{C}'\mathbf{C}$ is p.d.; we then have $\beta = \mathbf{C}'\mathbf{C}^{-1}\boldsymbol{\Delta}\gamma(\mathbf{t})$ and (3.16) follows from (3.6). ∎

Remark 3.2. When $\varPhi = \{\mathbf{0}\}$, F is reduced to a given element θ_0 in $\mathcal{M}_{\ell\times r}$, the global parameter θ is the constant θ_0, the above reasoning gives $\widehat{\theta} = \theta_0$.

3.3. Existence of the msq regression parameter value

Lemma 3.2. (i) *With Assumption* 2.1 *for each* $i = 1, \ldots, k$, *a msq regression of* $\mathbf{Y}_i$ *on* X_i *parameter value exists under Conditions* A1 *and* A2, *and is unique under* A1, A2 *and* A3.

(ii) *Under* A1 *and* A2, *for* τ_i *to be a msq regression of* $\mathbf{Y}_i$ *on* X_i *parameter value it is necessary and sufficient that*

$$\mathrm{E}\{(\mathbf{b}_i(X_i) \otimes \mathbf{z}_i(X_i))\mathbf{Y}_i - (\mathbf{b}_i(X_i)\mathbf{b}_i'(X_i) \otimes \mathbf{z}_i(X_i))[\tau_i]\} \quad \text{exists and vanishes.} \tag{3.20}$$

Proof. Write $\mathbf{z} = \mathbf{z}_i(X_i)$, $\mathbf{b} = \mathbf{b}_i(X_i)$. Let S_z be the set of all $r \times 1$ random vectors ζ defined up to an equivalence and such that $E\|\zeta\|^2_{\mathbf{z}} = E(\zeta'\mathbf{z}\zeta) = E\|\mathbf{z}^{1/2}\zeta\|^2 < \infty$. Then S_z is a linear space and the function $\varphi(\xi, \zeta) = E(\xi'\mathbf{z}\zeta)$ is an inner product in S_z. Further, $\mathbf{q}$ being $\ell(i) \times r$ and $\mathbf{b}$ being $\ell(i) \times 1$, from (3.8), (3.3) we have

$$\begin{aligned}
\mathbf{q}'\mathbf{b} &= [\mathbf{b}'\mathbf{q}] = (\mathbf{b}' \otimes \mathbf{I}_r)[\mathbf{q}],\\
\mathrm{E}\,\|\mathbf{z}^{1/2}\mathbf{q}'\mathbf{b}\|^2 &\le \mathrm{E}\{\|\mathbf{z}^{1/2}(\mathbf{b}' \otimes \mathbf{I}_r)\|^2\|\mathbf{q}\|^2\} = \|\mathbf{q}\|^2\mathrm{E}\,\|(\mathbf{b} \otimes \mathbf{z}^{1/2})\|^2\\
&= \|\mathbf{q}\|^2\mathrm{E}\{\mathrm{Tr}\,(\mathbf{b} \otimes \mathbf{z}^{1/2})(\mathbf{b}' \otimes \mathbf{z}^{1/2})\}\\
&= \|\mathbf{q}\|^2\mathrm{E}\{(\mathrm{Tr}\,\mathbf{b}\mathbf{b}')(\mathrm{Tr}\,\mathbf{z})\} = \|\mathbf{q}\|^2\mathrm{E}\,\|\mathbf{b}\|^2\mathrm{Tr}\,\mathbf{z}.
\end{aligned}$$

Due to Condition A2, $\mathbf{q}'\mathbf{b}$ belongs to S_z for all $\mathbf{q} \in \mathcal{M}_{\ell(i)\times r}$; in particular, each column vector of $\mathbf{b}'\otimes\mathbf{I}_r$ belongs to S_z. Thus in the space S_z, $\mathbf{q}'\mathbf{b}$ varies over the finite-dimensional subspace G generated by $\ell(i)r$ column vectors of $\mathbf{b}' \otimes \mathbf{I}_r$. By Condition A1 also $\mathbf{Y}_i \in S_z$.

In the Euclidean space S_z the square distance $\mathrm{E}\,\|\mathbf{Y}_i - \mathbf{q}'\mathbf{b}\|^2_{\mathbf{z}}$ is minimized if and only if $\mathbf{q}'\mathbf{b}$ is the orthogonal projection of $\mathbf{Y}_i$ on G. Since G is a finite-dimensional subspace of the Euclidean space S_z, the orthogonal projector from S_z onto G exists (see Chambadal and Ovaert (1968), p. 370). Thus the orthogonal projection exists, denote it by $\tau_i'\mathbf{b}$; then the existence of a msq regression parameter value is proved under A1 and A2. Further if such a value were not unique there would be some non-random $\mathbf{p} \in \mathcal{M}_{\ell(i)\times r}$, $\mathbf{p} \neq \mathbf{0}$ such that $\mathbf{p}'\mathbf{b} = \mathbf{0}$ or, equivalently, there would be some non-random $\ell(i)$-vector $\mathbf{u} \neq \mathbf{0}$ such that $\mathbf{u}'\mathbf{b} = 0$, which contradicts Condition A3. Moreover, the minimizing property of τ_i is, under A1 and A2, equivalent to $\mathbf{Y}_i - \tau_i'\mathbf{b} \perp G$, which, due to (3.9), can be written as

$$\mathrm{E}\left\{(\mathbf{Y}_i' - \mathbf{b}'\tau_i)\mathbf{z}(\mathbf{b}' \otimes \mathbf{I}_r)\right\} = 0, \quad \mathrm{E}\left\{(\mathbf{b} \otimes \mathbf{z})\mathbf{Y}_i - (\mathbf{b}\mathbf{b}' \otimes \mathbf{z})[\tau_i]\right\} = 0. \ \blacksquare$$

3.4. Infinitesimality of γ. From now on, with Assumption 2.1 we shall consider

$$\mathbf{b}_{ij} = \mathbf{b}_i(X_{ij}), \tag{3.21}$$

$$\mathbf{z}_{ij} = \mathbf{z}_i(X_{ij}). \tag{3.22}$$

LEMMA 3.3. *Under Assumption* 2.1 *and Conditions* A0–A2, *if τ is a global msq regression parameter value then $\gamma(\tau) \xrightarrow{a.s.} \mathbf{0}$ as $a \to \infty$.*

Proof. Under Condition A0, from Proposition 2.1 for non-random $\tau_i \in \mathcal{M}_{\ell(i)\times r}$, $i = 1, \ldots, k$, the function $\gamma_i(\tau_i)$, defined by (3.10) with (3.21) and (3.22), tends a.s. to zero if (3.20) holds. But from Lemma 3.2(ii), under Conditions A1 and A2, (3.20) means τ_i is a msq regression of $\mathbf{Y}_i$ on X_i parameter value. $\blacksquare$

3.5. A crucial lemma. The lemma below states an important property of the matrix $\mathbf{G\Delta}$ in (3.15) which, in view of Lemma 3.3, plays a key role in the GLSE convergence. We shall use the symbol

$$\mathcal{M}_{\mathrm{pd}}(n) = \text{set of all } n \times n \text{ p.d. matrices}$$

and first note

PROPOSITION 3.1. *In the linear space $\prod_{i=1}^{k} \mathcal{M}_{n_i\times n_i}$ endowed with an arbitrary norm and supplied with its norm topology, the product set $\prod_{i=1}^{k} \mathcal{M}_{\mathrm{pd}}(n_i)$ is open.*

Hence, together with each element, it contains some closed ball of positive radius centered at that element. From now on, the above topology will always be implied.

Proof. For any positive integer s consider the linear topological space $\mathcal{M}_{s\times s}$ with topology induced by an arbitrary norm $n(\cdot)$. By (1.6) consider the one-to-one linear map $\mathbf{u} \mapsto \operatorname{vec}\mathbf{u}$ from $\mathcal{M}_{s\times s}$ onto $\mathbb{R}^{s^2}$ inducing a norm $p(\cdot)$ by $p(\operatorname{vec}\mathbf{u}) = n(\mathbf{u})$. Then this map is an isometric isomorphism from $\mathcal{M}_{s\times s}$ onto $\mathbb{R}^{s^2}$. Hence from the norm topology viewpoint we can identify $\mathcal{M}_{s\times s}$ with $\mathbb{R}^{s^2}$. But in $\mathbb{R}^{s^2}$ every norm is equivalent to the Euclidean norm $\|\cdot\|$ (cf. Cartan (1967)), hence we can consider $n(\mathbf{u}) = \|\mathbf{u}\|$. If $\mathbf{u}_0 \in \mathcal{M}_{\mathrm{pd}}(s)$ is given, as soon as $\|\mathbf{u} - \mathbf{u}_0\|$ is sufficiently small all principal minors of $\mathbf{u}$ are close to the corresponding ones of $\mathbf{u}_0$ hence are all positive, then, by Proposition 1c.1(iv) of Rao (1973), $\mathbf{u}$ is also p.d. Therefore $\mathcal{M}_{\mathrm{pd}}(s)$ is open in the topology of $\mathcal{M}_{s\times s}$ induced by an arbitrary norm, in particular by the max-norm $n(\mathbf{u}) = \max |a_{fg}|$ for $\mathbf{u} = (a_{fg})_{s,s}$.

Thus the product set $\prod_{i=1}^k \mathcal{M}_{\mathrm{pd}}(n_i)$ is open in the product topology $\prod_{i=1}^k \tau_i$, where τ_i is the topology in $\mathcal{M}_{n_i\times n_i}$ induced by an arbitrary norm. Without loss of generality we can consider that each τ_i is induced by the max-norm; then a base for τ_i is the collection of all open cubes in $\mathcal{M}_{n_i\times n_i}$ hence a base for $\prod_{i=1}^k \tau_i$ is the collection of open cubes in $\prod_{i=1}^k \mathcal{M}_{n_i\times n_i}$, which are defined by isomorphism if we identify $\prod_{i=1}^k \mathcal{M}_{n_i\times n_i}$ with $\prod_{i=1}^k \mathbb{R}^{n_i^2}$ or $\mathbb{R}^{\sum_{i=1}^k n_i^2}$. Therefore the product topology $\prod_{i=1}^k \tau_i$ coincides with the topology induced by the norm $q(\mathbf{x}) = \max q(\mathbf{x}_i)$, where $\mathbf{x} = (\mathbf{x}_1,\ldots,\mathbf{x}_k)$ and $q(\mathbf{x}_i)$ is the max-norm of $\mathbf{x}_i \in \mathcal{M}_{n_i\times n_i}$; this last topology is also induced by any other norm in $\prod_{i=1}^k \mathcal{M}_{n_i\times n_i}$, which can be seen by identifying this product space with $\mathbb{R}^{\sum_{i=1}^k n_i^2}$. Thus the set $\prod_{i=1}^k \mathcal{M}_{\mathrm{pd}}(n_i)$ is open in the linear topological space $\prod_{i=1}^k \mathcal{M}_{n_i\times n_i}$ whose topology is induced by an arbitrary norm. ∎

In Lemma 3.4 below we shall use the following symbols, the relation of which to the ones previously used will be clear later.

Symbols 3.1. $(\forall i = 1,\ldots,k)$ $a(i)$ = *positive number*, n_i = *positive integer*, $\mathbf{T}_i = n_i \times n_i$ *p.d. matrix.*

$$\mathbf{\Delta} = \mathrm{diag}(a(1)\mathbf{I}_{n_1},\ldots,a(k)\mathbf{I}_{n_k}),$$

$$\mathbf{T} = \mathrm{diag}(\mathbf{T}_1,\ldots,\mathbf{T}_k), \quad n = \sum_{i=1}^k n_i, \quad \mathbb{R}^n = \prod_{i=1}^k \mathbb{R}^{n_i},$$

$$\Phi = \textit{any linear subspace of } \mathbb{R}^n,$$

$$\mathbf{C} = p\times n \textit{ real matrix with } \mathbf{C}'\mathbf{C} = \mathbf{\Delta}\mathbf{T},$$

$$\mathbb{R}^p \textit{ is endowed with the inner product } (\mathbf{u},\mathbf{v}) = \mathbf{u}'\mathbf{v},$$

$$\mathbf{J} = \textit{orthogonal projector from } \mathbb{R}^p \textit{ onto its subspace } \mathbf{C}\Phi \ (\textit{cf.}\ (3.12)),$$

$$B_0 = \textit{fixed closed ball in } \textstyle\prod_{i=1}^k \mathcal{M}_{\mathrm{pd}}(n_i) \ (\textit{cf. Proposition}\ 3.1),$$

$$\delta = (\delta_1,\ldots,\delta_k), \ \textit{with } 0 \le \delta_i < 1\ \forall i.$$

Lemma 3.4. *There exists a positive number $D(\Phi, B_0)$ depending on Φ and B_0 only such that*

$$(\forall(\mathbf{T}_1,\ldots,\mathbf{T}_k)\in B_0) \quad \|(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{J}\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{\Delta}\|^2 \le D(\Phi,B_0) \tag{3.23}$$

independently of $a(1),\ldots,a(k)$ and $\mathbf{C}$. Moreover, if $1 \le \dim\Phi \le \min(n_1,\ldots,n_k)$ then the class Γ_δ of subspaces Φ at most δ_i-steep relative to every coordinate space $\mathbb{R}^{n_i}$ is non-void provided $\delta_1,\ldots,\delta_k$ are sufficiently close to 1, *and there is a positive constant $\alpha(B_0,\delta_1,\ldots,\delta_k)$ depending on $B_0,\delta_1,\ldots,\delta_k$ only such that*

$$(\forall\Phi\in\Gamma_\delta)(\forall(\mathbf{T}_1,\ldots,\mathbf{T}_k)\in B_0) \quad \|(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{J}\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{\Delta}\|^2 \le \alpha(B_0,\delta_1,\ldots,\delta_k) \tag{3.24}$$

independently of $a(1),\ldots,a(k)$ and $\mathbf{C}$.

When $k = 1$, Γ_δ is just the class of all subspaces of $\mathbb{R}^n$ and the constant α is independent of δ_1. Lemma 3.4 will be proved in Section 4.

3.6. Interpretation of Lemma 3.4. Our approach to exploring the GLSE convergence consists in grounding the theory on Lemma 3.4. This is explained by some interpretation using a random linear operator which clarifies the meaning of the GLSE expression (3.15). The following formulae are useful for this purpose. Assume $\mathbf{C}'\mathbf{C}$ is p.d.; we get

$$\mathrm{Pr}_{\mathbf{C}\Phi} = \mathbf{J} = \mathbf{CGC}', \tag{3.25}$$

$$\mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}} = \mathbf{GC}'\mathbf{C} = \mathbf{G\Delta T}, \tag{3.26}$$

$$\mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}} = \mathbf{E}(\mathbf{E}'\mathbf{\Delta T E})^{-1}\mathbf{E}'\mathbf{\Delta T}, \tag{3.27}$$

$$\mathbf{G} = \mathbf{E}(\mathbf{E}'\mathbf{\Delta T E})^{-1}\mathbf{E}'. \tag{3.28}$$

Indeed, (3.27) comes from formula (1c.4.3) of Rao (1973), which also gives

$$\mathrm{Pr}_{\mathcal{M}(\mathbf{C})} = \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'. \tag{3.29}$$

Hence, $\mathbf{J}$ being symmetric and idempotent, by (3.13),

$$\mathbf{J} = \mathbf{JC}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}' = \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{JC}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}' = \mathbf{CGC}'.$$

To prove (3.26), take λ arbitrary in $\mathbb{R}^{\ell r}$. Then $\mathbf{C}\alpha = \mathbf{JC}\lambda$ means $\alpha \in \Phi$, $\mathbf{C}\lambda = \mathbf{C}\alpha + \mathbf{C}\delta$, $\mathbf{C}\delta \perp \mathbf{C}\Phi$, i.e. $\delta \perp_{\mathbf{C}'\mathbf{C}} \Phi$; since $\mathbf{C}$ is $sr \times \ell r$ and $\mathrm{Rank}\,\mathbf{C} = \ell r$ we have $\lambda = \alpha + \delta$, $\alpha \in \Phi$, $\delta \perp_{\mathbf{C}'\mathbf{C}} \Phi$, i.e. $\alpha = \mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}}\lambda$; on the other hand $\mathbf{C}\alpha = \mathbf{JC}\lambda$ entails $\mathbf{C}'\mathbf{C}\alpha = \mathbf{C}'\mathbf{JC}\lambda$, $\alpha = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{JC}\lambda$, hence $(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{JC} = \mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}}$. Finally (3.28) follows from (3.26) and (3.27) since $\mathbf{\Delta T}$ is p.d. ∎

By Lemma 3.1,

$$[\widehat{\theta}(F) - \mathbf{t}] = \mathbf{G\Delta T}\beta = \mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}}\beta, \quad \beta = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{Z}^{1/2}\eta(\mathbf{t})$$

by (3.16); in this formula $\mathbf{G\Delta T}$ is a random linear operator which depends on the explanatory observations only.

When $(\mathbf{T}_1, \ldots, \mathbf{T}_k)$ *varies over a fixed closed ball* B_0 *in* $\prod_{i=1}^{k} \mathcal{M}_{\mathrm{pd}}(n_i)$, *the uniform boundedness of the linear operator* $\mathbf{G\Delta T}$ *is equivalent to the uniform boundedness of the matrix* $\mathbf{G\Delta}$.

Indeed, viewing a $q \times q$ real matrix $\mathbf{A}$ as a linear operator, we define $\nu(A)$ = norm of the operator $A = \sup_{\mathbf{x}\in\mathbb{R}^q,\, \|\mathbf{x}\|=1} \|A\mathbf{x}\|$. Then $\nu(A) \le \|\mathbf{A}\|$, for $\|\mathbf{Ax}\| \le \|\mathbf{A}\| \cdot \|\mathbf{x}\|$. Write $\mathbf{A} = (\mathbf{A}_1 \ldots \mathbf{A}_q)$, $\mathbf{x} = (x_1 \ldots x_q)'$, and take $x_h = 1$. We see that $\|\mathbf{A}_h\| \le \nu(A)$ for all h, hence $\|\mathbf{A}\|^2 \le q\nu^2(A)$. Thus $\nu^2(A) \le \|\mathbf{A}\|^2 \le q\nu^2(A)$. Now letting $A = \mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}}$, we see that when $(\mathbf{T}_1, \ldots, \mathbf{T}_k) \in B_0$ the uniform boundedness of A is equivalent to that of the matrix $\mathbf{G\Delta T}$ and, in turn, to that of $\mathbf{G\Delta}$, for $\|\mathbf{G\Delta T}\|^2 \le \|\mathbf{G\Delta}\|^2\|\mathbf{T}\|^2$ and $\|\mathbf{G\Delta}\|^2 \le \|\mathbf{G\Delta T}\|^2\|\mathbf{T}^{-1}\|^2$. ∎

Note that, on account of $\|\mathbf{Ax}\|^2 = \mathbf{x}'\mathbf{A}'\mathbf{Ax}$, $\nu^2(A) = \sup_{\|\mathbf{x}\|=1} \|\mathbf{x}'\mathbf{A}'\mathbf{Ax}\|$ = maximum eigenvalue of $\mathbf{A}'\mathbf{A}$; cf. Rao (1973), formula (1f.2.1), p. 62.

Remark 3.3. We shall now compare our approach to proving the GLSE convergence with another standard one. Since by (3.6), $\mathbf{\Delta}\gamma(\mathbf{t}) = \mathbf{C}'\mathbf{Z}^{1/2}\eta(\mathbf{t})$, we can rewrite (3.15) as

$$[\widehat{\theta} - \mathbf{t}] = \mathbf{G\Delta}\gamma(\mathbf{t}) = \mathbf{GC}'\mathbf{Z}^{1/2}\eta(\mathbf{t}) \quad \text{for any } \mathbf{t} \in F.$$

Using $\|\mathbf{A}\mathbf{x}\| \le \nu(A)\|\mathbf{x}\|$ and, from above, $\nu^2(G^{1/2})$ = maximum eigenvalue of $\mathbf{G} = \lambda_{\max}(\mathbf{G})$ say, we get

$$\|\widehat{\theta} - \mathbf{t}\|^2 \le \lambda_{\max}(\mathbf{G})\|\mathbf{G}^{1/2}\mathbf{C}'\mathbf{Z}^{1/2}\eta(\mathbf{t})\|^2 = \lambda_{\max}(\mathbf{G})(\eta'(\mathbf{t})\mathbf{Z}^{1/2}\mathbf{J}\mathbf{Z}^{1/2}\eta(\mathbf{t})),$$

for $\mathbf{CGC}' = \mathbf{J}$ by (3.25). In the classic univariate multiple regression model $\mathbf{Y} = \mathbf{B}\theta + \eta$ one takes $\mathbf{Z} = \mathbf{I}$, for the OLSE from (3.17), Remark 3.1, and by (3.29) we have $\mathbf{J} = \mathrm{Pr}_{\mathcal{M}(\mathbf{B})} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$, $\mathbf{G} = (\mathbf{B}'\mathbf{B})^{-1}$; then the preceding inequality with $\mathbf{t} = \theta$ becomes

$$\|\widehat{\theta} - \theta\|^2 \le \lambda_{\min}^{-1}(\mathbf{B}'\mathbf{B})(\eta'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\eta).$$

Lai and Wei's (1982) approach to proving the strong consistency of OLSE consists in starting from this inequality and seeking minimal assumptions ensuring that the quadratic form $Q = \eta'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\eta$ is $o(\lambda_{\min}(\mathbf{B}'\mathbf{B}))$. The approach of the present paper for proper GLSE consists in starting from the integral expression $[\widehat{\theta}(F) - \mathbf{t}] = \mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}}\beta$ and discovering the uniform boundedness of the family $\{\mathrm{Pr}_{\Phi}^{\mathbf{C}'\mathbf{C}}\}$ of linear operators, which yields necessary and sufficient conditions for the convergence as well as for the strong consistency of GLSE in multimodel systems.

3.7. Evaluation of the GLSE error norm. When applying Lemma 3.4 to the GLSE expression we have to know if $(\mathbf{T}_1, \dots, \mathbf{T}_k) \in B_0$; to answer this we now establish the convergence of the matrices $\mathbf{T}_i$ in (3.4).

PROPOSITION 3.2. *Under Assumption* 2.1 *let, for all* $i = 1, \dots, k$, $\{X_{ij} : j = 1, 2, \dots\}$ *be a stationary and indecomposable sequence. Then under Condition* A2 *as* $a \to \infty$,

$$\mathbf{T}_i \xrightarrow{a.s.} \mathbf{T}_{0i} = \mathrm{E}\,(\mathbf{b}_i(X_i)\mathbf{b}_i'(X_i) \otimes \mathbf{z}_i(X_i)), \ \ \textit{a n.n.d. matrix}, \forall i,$$
$$\mathbf{T} \xrightarrow{a.s.} \mathbf{T}_0 = \mathrm{diag}(\mathbf{T}_{01}, \dots, \mathbf{T}_{0k}).$$

If Condition A3 *is added then the limits* $\mathbf{T}_{0i}, \mathbf{T}_0$ *are p.d. and there exists a closed ball* B_0 *of positive radius centered at* $(\mathbf{T}_{01}, \dots, \mathbf{T}_{0k})$, *contained in* $\prod_{i=1}^{k} \mathcal{M}_{\mathrm{pd}}(\ell(i)r)$, *and in the basic probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *we have*

$$(\exists \Omega_1 \in \mathcal{F},\ \mathrm{P}\,\Omega_1 = 1)(\forall \omega \in \Omega_1)(\exists a_0(\omega))(\forall a \ge a_0(\omega)) \quad (\mathbf{T}_1, \dots, \mathbf{T}_k) \in B_0.$$

Proof. We can check that Condition A2 is equivalent to the existence and finiteness (for all i) of $\mathrm{E}\,(\mathbf{b}_i(X_i)\mathbf{b}_i'(X_i) \otimes \mathbf{z}_i(X_i))$; the last matrix for each i is n.n.d. and, with Condition A3 added, is p.d.; cf. Bac-Van (1994), Propositions 5.3, 5.4. The limits follow from (3.4) and Proposition 2.1, the rest from Proposition 3.1. ∎

Lemma 3.4 and Proposition 3.2 yield this important corollary which just states the large sample a.s. uniform boundedness of the random linear operator $\mathbf{G\Delta T}$ mentioned in Subsection 3.6 and which will be used mainly for proving Theorems 2.1–2.3.

COROLLARY 3.1. *Under Assumption* 2.1 *let* $\{X_{i1}, X_{i2}, \dots\}$ *be a stationary and indecomposable sequence for each* $i = 1, \dots, k$, *and let Conditions* A2, A3 *be satisfied. Then there exists a positive constant* $D(F)$ *depending only on the manifold* F *in Definition* 1.1 *such that in the basic probability space we have*

$$(3.30) \qquad (\exists \Omega_1 \in \mathcal{F},\ \mathrm{P}\,\Omega_1 = 1)(\forall \omega \in \Omega_1)(\exists a_0(\omega))(\forall a \ge a_0(\omega)) \quad \|\mathbf{G\Delta}\|^2 \le D(F).$$

Moreover, given numbers δ_i, $0 \le \delta_i < 1$, $i = 1, \dots, k$, *there is a positive constant* $\alpha(\delta_1, \dots, \delta_k)$ *depending on* $\delta_1, \dots, \delta_k$ *only such that*

$$(\forall a \ge a_0(\omega))(\forall F \in \mathcal{C}) \quad \|\mathbf{G}\boldsymbol{\Delta}\|^2 \le \alpha(\delta_1, \dots, \delta_k), \tag{3.31}$$

the class $\mathcal{C}$ *being defined by* (2.3).

Proof. In Proposition 3.2 fix the closed ball B_0 whose existence is asserted; then as $a \ge a_0(\omega)$, $(\mathbf{T}_1, \dots, \mathbf{T}_k) \in B_0$. Apply inequalities (3.23) and (3.24) to the subspace Φ of $\mathbb{R}^{\ell r}$ parallel to $[F]$. Then $D(\Phi, B_0) = D(F)$ depends only on F and $\alpha(B_0, \delta_1, \dots, \delta_k)$ depends only on $\delta_1, \dots, \delta_k$; moreover $F \in \mathcal{C}$ entails $\Phi \in \Gamma_\delta$ in (3.24). ■

The following lemma definitely asserts the a.s. existence of GLSE for sufficiently large global sample size and also evaluates the GLSE error norm. It serves the proof of Theorems 2.2 and 2.3.

LEMMA 3.5. *Under Assumption* 2.1 *let* $\{X_{i1}, X_{i2}, \dots\}$ *be a stationary and indecomposable sequence for each* $i = 1, \dots, k$, *and let Conditions* A2, A3 *be satisfied. Then there exists a positive constant* $D(F)$ *depending only on the manifold* F *in Definition* 1.1 *such that in the basic probability space* $(\Omega, \mathcal{F}, \mathrm{P})$,

$$\begin{aligned} &(\exists \Omega_1 \in \mathcal{F},\ \mathrm{P}\,\Omega_1 = 1)(\forall \omega \in \Omega_1)(\exists a_0(\omega))(\forall a \ge a_0(\omega)) \\ &\qquad \text{the GLSE } \widehat{\theta} = \widehat{\theta}(F) \text{ exists and } \|\widehat{\theta}(F) - \mathbf{t}\|^2 \le D(F)\|\gamma(\mathbf{t})\|^2 \text{ for any } \mathbf{t} \in F. \end{aligned} \tag{3.32}$$

Moreover, given numbers δ_i, $0 \le \delta_i < 1$, $i = 1, \dots, k$, *there is a positive constant* $\alpha(\delta_1, \dots, \delta_k)$ *depending on* $\delta_1, \dots, \delta_k$ *only such that*

$$(\forall a \ge a_0(\omega)) \quad \sup_{F \in \mathcal{C}} \|\widehat{\theta}(F) - \mathbf{t}\|^2 \le \alpha(\delta_1, \dots, \delta_k)\|\gamma(\mathbf{t})\|^2 \quad \forall \mathbf{t} \in \Theta \tag{3.33}$$

if $\mathcal{C}$ *is non-void*, $\mathcal{C}$ *being defined by* (2.3).

Proof. In Proposition 3.2, as $a \ge a_0(\omega)$, $(\mathbf{T}_1, \dots, \mathbf{T}_k) \in B_0$, hence the GLSE exists since $\mathbf{C}'\mathbf{C} = \boldsymbol{\Delta}\mathbf{T}$ is p.d. Using (3.15) and applying inequalities (3.30) and (3.31) we get (3.32) and (3.33). ■

3.8. Uniform convergence and uniform consistency. Lemma 3.2 asserts the existence and uniqueness of the global msq regression parameter value $\tau = (\tau_1' \dots \tau_k')'$. As a consequence of (3.32) and (3.33), Lemma 3.5, and by Lemma 3.3 as $a \to \infty$ we have

$$\|\widehat{\theta}(F) - \tau\|^2 \le D(F)\|\gamma(\tau)\|^2 \xrightarrow{\text{a.s.}} 0 \quad \text{if } \tau \in F, \tag{3.34}$$

$$\sup_{F \in \mathcal{C}} \|\widehat{\theta}(F) - \tau\|^2 \le \alpha(\delta_1, \dots, \delta_k)\|\gamma(\tau)\|^2 \xrightarrow{\text{a.s.}} 0 \quad \text{if } \tau \in \Theta. \tag{3.35}$$

Proof of Theorem 2.2. Follows from (3.35).

Proof of Theorem 2.3. If the true global parameter value θ in (1.2) coincides with the global msq regression parameter value τ then $\Theta \ni \tau$, so (3.35) yields $\sup_{F \in \mathcal{C}} \|\widehat{\theta}(F) - \theta\| \xrightarrow{\text{a.s.}} 0$ as $a \to \infty$. Conversely if this convergence holds and if there exists $F \ni \tau$ with $F \in \mathcal{C}$ then by (3.34), $\widehat{\theta}(F) \xrightarrow{\text{a.s.}} \tau$ as $a \to \infty$, hence from the former convergence $\tau = \theta$. ■

3.9. Convergence of GLSE

3.9.1. *Some matrix formulae.* The formulae below will be used in what follows. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ be real matrices of which $\mathbf{C}$ is of order $p \times q$. Then

$$\operatorname{Rank} \mathbf{C} + \dim \ker \mathbf{C} = \text{number of columns of } \mathbf{C}, \tag{3.36}$$

$$\dim \ker \mathbf{B} \cap \mathcal{M}(\mathbf{C}) = \dim \ker \mathbf{BC} \quad \text{if } \operatorname{Rank} \mathbf{C} = q, \tag{3.37}$$

$$\ker \mathbf{C}'\mathbf{AC} = \ker \mathbf{AC} \quad \text{for n.n.d. } \mathbf{A}, \tag{3.38}$$

$$\ker \mathbf{C}'\mathbf{DC} = \ker \mathbf{C}, \quad \operatorname{Rank} \mathbf{C}'\mathbf{DC} = \operatorname{Rank} \mathbf{C} \quad \text{for p.d. } \mathbf{D}. \tag{3.39}$$

(3.36) is the well known dimension formula; cf. Chambadal and Ovaert (1968), p. 54, formula (2). To prove (3.37) note that

$$\ker \mathbf{BC} = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{BCu} = \mathbf{0}\} = \{\mathbf{u} \in \mathbb{R}^q : \mathbf{Cu} \in \ker \mathbf{B}\};$$

but $\operatorname{Rank} \mathbf{C} = q$ means that $\mathbf{C}$ defines an isomorphism of vector spaces from $\mathbb{R}^q$ onto $\mathcal{M}(\mathbf{C})$, which maps $\ker \mathbf{BC}$ onto $\mathcal{M}(\mathbf{C}) \cap \ker \mathbf{B}$. To prove (3.38) consider

$$\operatorname{Rank} \mathbf{B}'\mathbf{BC} \le \operatorname{Rank} \mathbf{BC} = \operatorname{Rank} \mathbf{C}'\mathbf{B}'\mathbf{BC} \le \operatorname{Rank} \mathbf{B}'\mathbf{BC},$$

hence $\operatorname{Rank} \mathbf{B}'\mathbf{BC} = \operatorname{Rank} \mathbf{BC}$, which implies $\ker \mathbf{B}'\mathbf{BC} = \ker \mathbf{BC}$ by (3.36); but $\mathbf{A}$ n.n.d. means $\mathbf{A} = \mathbf{B}'\mathbf{B}$ for some $\mathbf{B}$, and so

$$\ker \mathbf{C}'\mathbf{AC} = \ker \mathbf{C}'\mathbf{B}'\mathbf{BC} = \ker \mathbf{BC} = \ker \mathbf{B}'\mathbf{BC} = \ker \mathbf{AC}.$$

3.9.2. *Notations.* Recall $\mathbf{C}'\mathbf{C} = \boldsymbol{\Delta}\mathbf{T}$ from (3.2). The following symbols will be used throughout:

$$n_i = \ell(i)r, \quad n = \ell r,$$

$$\mathbf{E} = \text{an } n \times m \text{ matrix}, \quad \operatorname{Rank} \mathbf{E} = m, \quad \mathcal{M}(\mathbf{E}) = \Phi, \tag{3.40}$$

$$\mathbf{E} = (\mathbf{E}_1' \ldots \mathbf{E}_k')', \quad \mathbf{E}_i \text{ is } n_i \times m, \quad \sum_{i=1}^{k} n_i = n. \tag{3.41}$$

By Remark 3.2 we shall only consider $m = \dim \Phi \ge 1$.

3.9.3. *GLSE and projection of the msq regression parameter value.* As an application of Lemma 3.4, the proposition below states a simple relation between the GLSE $\widehat{\theta}(F)$ and the projection on F of the global msq regression parameter value τ, which serves for the proof of convergence. We put

$$\mathbf{t}_0 = \operatorname{Pr}_F^{\mathbf{T}_0} \tau \quad \text{and} \quad \xi = [\tau] - [\mathbf{t}_0]. \tag{3.42}$$

Then by (1.13),

$$\mathbf{0} = \operatorname{Pr}_\Phi^{\mathbf{T}_0}([\tau] - [\mathbf{t}_0]), \quad \text{i.e.} \quad \xi \perp_{\mathbf{T}_0} \Phi. \tag{3.43}$$

We shall let $o(1)$ denote any random matrix of fixed order tending a.s. to $\mathbf{0}$ as $a \to \infty$.

PROPOSITION 3.3. *Under Assumption* 2.1 *and Conditions* A0–A3,

$$[\widehat{\theta}(F) - \operatorname{Pr}_F^{\mathbf{T}_0} \tau] = \mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0 \xi + o(1). \tag{3.44}$$

Proof. We start from (3.15),

$$[\widehat{\theta} - \mathbf{t}] = \mathbf{G}\boldsymbol{\Delta}\gamma(\mathbf{t}) \quad \text{for any } \mathbf{t} \in F.$$

Using (3.42) take $\mathbf{t} = \mathbf{t}_0$; then by (3.11),

$$\gamma(\mathbf{t}_0) = \mathbf{QY} - \mathbf{T}[\mathbf{t}_0] = \mathbf{QY} - \mathbf{T}[\tau] + \mathbf{T}\xi = \gamma(\tau) + (\mathbf{T} - \mathbf{T}_0)\xi + \mathbf{T}_0\xi.$$

By Lemma 3.3 and Proposition 3.2, as $a \to \infty$, $\gamma(\tau) \xrightarrow{\text{a.s.}} \mathbf{0}$ and $\mathbf{T} \xrightarrow{\text{a.s.}} \mathbf{T}_0$. Hence, ξ being fixed, $\gamma(\tau) + (\mathbf{T} - \mathbf{T}_0)\xi = o(1)$. Thus

$$[\widehat{\theta} - \mathbf{t}_0] = \mathbf{G\Delta}o(1) + \mathbf{G\Delta T}_0\xi.$$

From Corollary 3.1 there exists a positive constant $D(F)$ depending on F only such that a.s. as soon as the global sample size a is sufficiently large, $\|\mathbf{G\Delta}\|^2 \le D(F)$. Hence we get (3.44). ∎

3.9.4. *Properties of* $U(F)$. Starting from the support manifold F in Definition 1.1 we shall construct an affine manifold $U(F)$ which is the region of convergence in Theorem 2.1. Let $\mathbf{v} = (\mathbf{v}_1' \dots \mathbf{v}_k')'$ be any vector in $\mathbb{R}^n$, $\mathbf{v}_i \in \mathbb{R}^{n_i}$ for all i. We shall consider the sets

$$\Phi^{\perp_{\mathbf{T}_0}} = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{E}'\mathbf{T}_0\mathbf{v} = 0\},$$

$$S(F) = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{E}_i'\mathbf{T}_{0i}\mathbf{v}_i = 0 \ \forall i = 1, \dots, k\}. \tag{3.45}$$

The matrix $\mathbf{T}_0$ being fixed, $S(F)$ is a subspace of $\mathbb{R}^n$ determined by $\mathbf{E}$. From (3.40) the columns of $\mathbf{E}$ form a basis for Φ. When the basis changes, $S(F)$ remains the same, thus $S(F)$ is entirely determined by F, for F determines Φ. With the inner product $\mathbf{u}'\mathbf{T}_0\mathbf{v}$ the first set is the orthogonal complement of Φ which will briefly be written as $\Phi^\perp$. We have $S(F) \subset \Phi^\perp$ since

$$\mathbf{E}'\mathbf{T}_0\mathbf{v} = \sum_{i=1}^{k} \mathbf{E}_i'\mathbf{T}_{0i}\mathbf{v}_i.$$

By (3.43), $\xi \in \Phi^\perp$. Take a fixed matrix $\mathbf{t} \in F$. Then by (3.42), $[\mathbf{t}_0] - [\mathbf{t}] \in \Phi$. Consider $\mathbb{R}^n = \Phi^\perp \oplus \Phi$ and put $[\tau_0] = \xi + ([\mathbf{t}_0] - [\mathbf{t}])$. Then it follows that

$$[\tau_0] \in S(F) \oplus \Phi \ \Leftrightarrow \ \xi \in S(F).$$

Now rewrite $\xi = [\tau] - [\mathbf{t}_0]$ as

$$[\tau] = \xi + ([\mathbf{t}_0] - [\mathbf{t}]) + [\mathbf{t}] = [\tau_0] + [\mathbf{t}].$$

Put $[U(F)] = S(F) \oplus \Phi + [\mathbf{t}]$. Then

$$\xi \in S(F) \ \Leftrightarrow \ \tau \in U(F). \tag{3.46}$$

$U(F)$ is an affine manifold in $\mathcal{M}_{\ell\times r}$ which remains unchanged by varying $[\mathbf{t}]$ in $[F]$, and which is entirely determined by F. It has the following properties.

(1) $U(F) \supset F$ since $[F] = \Phi + [\mathbf{t}]$; $U(F) = F$ if and only if $S(F) = \{\mathbf{0}\}$, i.e. $\ker \mathbf{E}_i'\mathbf{T}_{0i} = \{\mathbf{0}\}$ for all i. Noting that $\mathbf{T}_{0i}$ is $n_i \times n_i$ and non-singular we have $\operatorname{Rank} \mathbf{E}_i'\mathbf{T}_{0i} = \operatorname{Rank} \mathbf{E}_i'$. Then by (3.36),

$$\dim\ker \mathbf{E}_i'\mathbf{T}_{0i} = n_i - \operatorname{Rank} \mathbf{E}_i' = n_i - \operatorname{Rank} \mathbf{E}_i.$$

(2) Thus $U(F) = F$ if and only if $\operatorname{Rank} \mathbf{E}_i = n_i$ for all i; this condition is intrinsic for Φ. In other words

$$U(F) = F \ \Leftrightarrow \ \sum_{i=1}^{k} \operatorname{Rank} \mathbf{E}_i = n,$$

for, $\mathbf{E}_i$ being $n_i \times m$, $\operatorname{Rank} \mathbf{E}_i \le n_i$ for all i and $\sum_{i=1}^{k} n_i = n$.

For example, let the n coordinate axes in $\mathbb{R}^n = \prod_{i=1}^k \mathbb{R}^{n_i}$ be orthogonal. Then $U(F) = F$ if and only if the projections of the m column vectors of $\mathbf{E}$ on $\mathbb{R}^{n_i}$ span the whole space $\mathbb{R}^{n_i}$, or what is the same, the projection of Φ on $\mathbb{R}^{n_i}$ coincides with $\mathbb{R}^{n_i}$ itself for all i.

(3) In $\mathbb{R}^n$ let $\mathbf{v}_{(i)}$ denote any vector $\mathbf{v} = (\mathbf{v}_1' \dots \mathbf{v}_k')'$ such that $\mathbf{v}_f = \mathbf{0}$ for all $f \neq i$. Then $S(F)$ is the direct sum of subspaces

$$S_i(F) = \{\mathbf{v}_{(i)} : \mathbf{E}_i'\mathbf{T}_{0i}\mathbf{v}_i = 0\}, \quad i = 1, \dots, k,$$

since every $\mathbf{v} \in S(F)$ has a unique representation $\mathbf{v} = \mathbf{v}_{(1)} + \dots + \mathbf{v}_{(k)}$ with $\mathbf{v}_{(i)} \in S_i(F)$ for all i. Hence $\dim S(F) = \sum_{i=1}^k \dim S_i(F)$. But by (3.37) and (3.36),

$$\dim S_i(F) = \dim \ker \mathbf{E}_i'\mathbf{T}_{0i} = \dim \ker \mathbf{E}_i' = n_i - \operatorname{Rank} \mathbf{E}_i'.$$

Therefore by (3.40),

$$\dim S(F) = n - \sum_{i=1}^k \operatorname{Rank} \mathbf{E}_i,$$
$$\dim S(F) \oplus \Phi = \dim S(F) + \dim \Phi = \dim S(F) + m.$$

Thus

$$[U(F)] = \mathbb{R}^n \;\Leftrightarrow\; \dim S(F) \oplus \Phi = n \;\Leftrightarrow\; m = \sum_{i=1}^k \operatorname{Rank} \mathbf{E}_i.$$

The last equality means $\dim \mathcal{M}(\mathbf{E}') = \sum_{i=1}^k \dim \mathcal{M}(\mathbf{E}_i')$ or, equivalently, $\mathcal{M}(\mathbf{E}') = \bigoplus_{i=1}^k \mathcal{M}(\mathbf{E}_i')$ since $\mathcal{M}(\mathbf{E}_i')$ are subspaces of $\mathcal{M}(\mathbf{E}')$; cf. Chambadal and Ovaert (1968), p. 51, Theorem 6.3. But $\mathcal{M}(\mathbf{E}')$ and $\mathcal{M}(\mathbf{E}_i')$ are respectively isomorphic to $\mathcal{M}(\mathbf{E})$ and $\mathcal{M}(\mathbf{E}_i)$, hence we can write $\mathcal{M}(\mathbf{E}) = \bigoplus_{i=1}^k \mathcal{M}(\mathbf{E}_i)$ or, equivalently, $\mathcal{M}(\mathbf{E}) = \prod_{i=1}^k \mathcal{M}(\mathbf{E}_i)$; cf. Dunford and Schwartz (1958), p. 38. Therefore, $U(F) = \mathcal{M}_{\ell \times r}$, the global parameter range space, if and only if Φ is the product of its projections on the coordinate spaces $\mathbb{R}^{n_i}$—we use orthogonal coordinate axes in $\mathbb{R}^n = \prod_{i=1}^k \mathbb{R}^{n_i}$.

For example, in $\mathbb{R}^3 = \mathbb{R}^1 \times \mathbb{R}^2$ if Φ is a line or a plane then the above condition means Φ is perpendicular to either $\mathbb{R}^1$ or $\mathbb{R}^2$.

(4) Since $m = \operatorname{Rank} \mathbf{E}' \leq \sum_{i=1}^k \operatorname{Rank} \mathbf{E}_i'$,

$$U(F) \subset \mathcal{M}_{\ell \times r} \text{ properly } \;\Leftrightarrow\; m < \sum_{i=1}^k \operatorname{Rank} \mathbf{E}_i. \tag{3.47}$$

The case $F \subset U(F) \subset \mathcal{M}_{\ell \times r}$ with proper inclusion really occurs when and only when $m < \sum_{i=1}^k \operatorname{Rank} \mathbf{E}_i < n$.

3.9.5. *Proof of Theorem 2.1.* From (3.44),

$$\widehat{\theta}(F) \text{ converges a.s. } \;\Leftrightarrow\; \text{ so does } \mathbf{G\Delta T}_0\xi. \tag{3.48}$$

Write $\xi = (\xi_1' \dots \xi_k')'$, $\xi_i \in \mathbb{R}^{n_i}$. By (3.41), $\mathbf{E}' = (\mathbf{E}_1' \dots \mathbf{E}_k')$. Recall $\mathbf{\Delta T}_0 = \operatorname{diag}(a(1)\mathbf{T}_{01}, \dots, a(k)\mathbf{T}_{0k})$. Then we have the following similar formulae:

$$\mathbf{E}'\mathbf{\Delta T E} = \sum_{i=1}^k a(i)\mathbf{E}_i'\mathbf{T}_i\mathbf{E}_i, \tag{3.49}$$

$$\mathbf{E}'\mathbf{T}\mathbf{E} = \sum_{i=1}^{k} \mathbf{E}_i'\mathbf{T}_i\mathbf{E}_i, \tag{3.50}$$

$$\mathbf{E}'\boldsymbol{\Delta}\mathbf{T}_0\xi = \sum_{i=1}^{k} a(i)\mathbf{E}_i'\mathbf{T}_{0i}\xi_i.$$

On account of (3.28), $\mathbf{G} = \mathbf{E}(\mathbf{E}'\boldsymbol{\Delta}\mathbf{T}\mathbf{E})^{-1}\mathbf{E}'$, so

$$\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi = \mathbf{E}(\mathbf{E}'\boldsymbol{\Delta}\mathbf{T}\mathbf{E})^{-1}\sum_{i=1}^{k} a(i)\mathbf{E}_i'\mathbf{T}_{0i}\xi_i. \tag{3.51}$$

If $\tau \in U(F)$ then by (3.46), $\xi \in S(F)$, hence, by (3.45), $\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi = \mathbf{0}$, thus (3.44) yields $\widehat{\theta}(F) \xrightarrow{\text{a.s.}} \Pr_F^{\mathbf{T}_0}\tau$ as $a \to \infty$.

For later use the following consideration is general. From the above $\mathbf{E}'\mathbf{T}_0\xi = \sum_{i=1}^{k}\mathbf{E}_i'\mathbf{T}_{0i}\xi_i$, hence by (3.43),

$$\sum_{i=1}^{k} \mathbf{E}_i'\mathbf{T}_{0i}\xi_i = \mathbf{0}. \tag{3.52}$$

Thus $\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi = \mathbf{0} \to \mathbf{0}$ as $a \to \infty$ so that $a(1) = \ldots = a(k)$.

Let $k \geq 2$. When $a(2) = \ldots = a(k) = a(1)/2$, from (3.52), (3.49) and (3.50),

$$\sum_{i=1}^{k} a(i)\mathbf{E}_i'\mathbf{T}_{0i}\xi_i = a(2)\mathbf{E}_1'\mathbf{T}_{01}\xi_1,$$

$$\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi = \mathbf{E}(a^{-1}(2)\mathbf{E}'\boldsymbol{\Delta}\mathbf{T}\mathbf{E})^{-1}\mathbf{E}_1'\mathbf{T}_{01}\xi_1 \quad \text{by (3.51)},$$

$$a^{-1}(2)(\mathbf{E}'\boldsymbol{\Delta}\mathbf{T}\mathbf{E}) = \mathbf{E}_1'\mathbf{T}_1\mathbf{E}_1 + \mathbf{E}'\mathbf{T}\mathbf{E} \xrightarrow{\text{a.s.}} \mathbf{E}_1'\mathbf{T}_{01}\mathbf{E}_1 + \mathbf{E}'\mathbf{T}_0\mathbf{E},$$

since $\mathbf{T} \xrightarrow{\text{a.s.}} \mathbf{T}_0$ as $a \to \infty$. $\mathbf{T}_0$ and $\mathbf{E}'\mathbf{T}_0\mathbf{E}$ being p.d., it follows that

$$\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi \xrightarrow{\text{a.s.}} \mathbf{E}(\mathbf{E}_1'\mathbf{T}_{01}\mathbf{E}_1 + \mathbf{E}'\mathbf{T}_0\mathbf{E})^{-1}\mathbf{E}_1'\mathbf{T}_{01}\xi_1$$

as $a \to \infty$ so that $a(2) = \ldots = a(k) = a(1)/2$.

Now consider the case $\tau \notin U(F)$, i.e., by (3.46), $\xi \notin S(F)$ or, equivalently, by (3.45), $\mathbf{E}_i'\mathbf{T}_{0i}\xi_i \neq \mathbf{0}$ for some i. Without loss of generality we can consider $\mathbf{E}_1'\mathbf{T}_{01}\xi_1 \neq \mathbf{0}$. Then

$$(\mathbf{E}_1'\mathbf{T}_{01}\mathbf{E}_1 + \mathbf{E}'\mathbf{T}_0\mathbf{E})^{-1}\mathbf{E}_1'\mathbf{T}_{01}\xi_1 = \zeta, \quad \text{say},$$

is a non-null vector. Hence $\mathbf{E}\zeta \neq \mathbf{0}$ by (3.40). Thus from the above for $k \geq 2$ as $a \to \infty$,

$$\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi \xrightarrow{\text{a.s.}} \begin{cases} \mathbf{E}\zeta \neq \mathbf{0} & \text{if } a(2) = \ldots = a(k) = a(1)/2, \\ \mathbf{0} & \text{if } a(1) = \ldots = a(k). \end{cases}$$

Therefore $\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi$ diverges a.s., and so does $\widehat{\theta}(F)$ by (3.48). Theorem 2.1 is proved.

Remark 3.4. In Lemma 3.4 and Corollary 3.1 the assertion on boundedness of $\mathbf{G}\boldsymbol{\Delta}$ is reasonably mild, since $\mathbf{G}\boldsymbol{\Delta}$ and $\Pr_{\Phi}^{\boldsymbol{\Delta}\mathbf{T}} = \mathbf{G}\boldsymbol{\Delta}\mathbf{T}$ diverge a.s. when $m < \sum_{i=1}^{k}\operatorname{Rank}\mathbf{E}_i$. Indeed, by (3.47) this is the case when $U(F) \subset \mathcal{M}_{\ell\times r}$ properly, hence $\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0\xi$ diverges a.s. as soon as $\tau \notin U(F)$; then $\mathbf{G}\boldsymbol{\Delta}\mathbf{T}_0$, $\mathbf{G}\boldsymbol{\Delta}$ and $\mathbf{G}\boldsymbol{\Delta}\mathbf{T}$ also diverge a.s.

But in the one-model case ($k = 1$) $\boldsymbol{\Delta} = a\mathbf{I}_n$, hence by (3.26), (3.27) and Proposition 3.2, we have

$$\mathbf{G}\boldsymbol{\Delta}\mathbf{T} = \mathbf{E}(\mathbf{E}'\boldsymbol{\Delta}\mathbf{T}\mathbf{E})^{-1}\mathbf{E}'\boldsymbol{\Delta}\mathbf{T} = \mathbf{E}(\mathbf{E}'\mathbf{T}\mathbf{E})^{-1}\mathbf{E}'\mathbf{T} \xrightarrow{\text{a.s.}} \mathbf{E}(\mathbf{E}'\mathbf{T}_0\mathbf{E})^{-1}\mathbf{E}'\mathbf{T}_0$$

as $a \to \infty$, so $\mathbf{G}\boldsymbol{\Delta}$ also converges a.s.

4. Algebraic tools

This section aims at proving Lemma 3.4, the crucial tool for establishing the GLSE convergence and consistency. Lemma 3.4 is trivial when $\varPhi = \mathbb{R}^n$, which corresponds to the OLSE, since we then have $\mathbf{C}\varPhi = \mathcal{M}(\mathbf{C})$, $\mathbf{J} = \mathrm{Pr}_{\mathbf{C}\varPhi} = \mathrm{Pr}_{\mathcal{M}(\mathbf{C})}$, $\mathbf{JC} = \mathbf{C}$, hence, on account of $\mathbf{C}'\mathbf{C} = \mathbf{\Delta T}$,

$$\|(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{JC}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{\Delta}\|^2 = \|\mathbf{T}^{-1}\|^2 = \sum_{i=1}^{k} \|\mathbf{T}_i^{-1}\|^2.$$

Further, consider the one-model case ($k = 1$). From Remark 3.4 under the conditions of Proposition 3.2, $\mathbf{G\Delta}$ converges a.s. as $a \to \infty$. To prove the algebraic Lemma 3.4 the reasoning in Bac-Van (1994), Proposition 4.4, leads to the inequality

$$\|(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{JC}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{\Delta}\|^2 \le n(\mathrm{Tr}\,\mathbf{T}^{-1})^2\Big(\sum_{f,g=1}^{k} a(g)a^{-1}(f)\Big),$$

the right-hand side of which in the case $k = 1$ is reduced to $n(\mathrm{Tr}\,\mathbf{T}^{-1})^2$. Thus Lemma 3.4 is proved when $k = 1$. The only difficulty resides in $\varPhi \ne \mathbb{R}^n$, $k \ge 2$, which corresponds to proper GLSE in systems of k models.

We shall use a coordinate approach. Throughout the proof the following symbols are to be kept in mind.

Symbols 4.1. Recall the p.d. matrices: $\mathbf{T}_i$ being $n_i \times n_i$, $\mathbf{T} = \mathrm{diag}(\mathbf{T}_1, \ldots, \mathbf{T}_k)$ and $\mathbf{\Delta} = \mathrm{diag}(a(1)\mathbf{I}_{n_1}, \ldots, a(k)\mathbf{I}_{n_k})$ being $n \times n$ with positive numbers $a(i)$.

We let $\mathbf{E}$ denote an $n \times m$ real block matrix, $\mathbf{E} = (\mathbf{E}_1' \ldots \mathbf{E}_k')'$, where $\mathbf{E}_i$ is $n_i \times m$, by specifying the columns $\mathbf{E} = (\mathbf{e}_1 \ldots \mathbf{e}_m)$, $\mathbf{E}_i = (\mathbf{e}_{1i} \ldots \mathbf{e}_{mi})$; next,

$$\mathbf{M}_i = \mathbf{E}_i'\mathbf{T}_i\mathbf{E}_i, \tag{4.1}$$

$$\mathbf{M}_i = \mathbf{P}(i)\mathbf{\Lambda}(i)\mathbf{P}'(i) \quad \text{(spectral decomposition)}, \tag{4.2}$$

where

$$\mathbf{P}(i) = \text{orthogonal matrix (arbitrary when } \mathbf{M}_i = \mathbf{0}),$$

$$\mathbf{\Lambda}(i) = \mathrm{diag}(\lambda_1(i), \ldots, \lambda_m(i)), \quad \lambda_1(i) \ge \ldots \ge \lambda_m(i) \ge 0,$$

$$\mathbf{M} = \sum_{i=1}^{k} a(i)\mathbf{M}_i, \tag{4.3}$$

$$\mathbf{F}(i) = a^{-1}(i)\mathbf{P}'(i)\mathbf{M}\mathbf{P}(i). \tag{4.4}$$

Then

$$\mathbf{M} = \mathbf{E}'\mathbf{\Delta T E}, \tag{4.5}$$

hence, when $\mathrm{Rank}\,\mathbf{E} = m$, $\mathbf{M}$ is p.d. and so is $\mathbf{F}(i)$.

The proof consists of the following steps.

4.1. Three identities. We first seek an expression for the orthogonal projector in Lemma 3.4. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a basis for a subspace G of $\mathbb{R}^p$. We let μ_{fg} denote the (f, g) entry of the matrix

$$\{(\mathbf{u}_1 \ldots \mathbf{u}_m)'(\mathbf{u}_1 \ldots \mathbf{u}_m)\}^{-1} = (\mu_{fg}). \tag{4.6}$$

Consider the matrix

$$\mathbf{J} = \sum_{f,g=1}^{m} \mu_{fg}\mathbf{u}_f\mathbf{u}_g'. \tag{4.7}$$

We always assume that $\mathbb{R}^p$ is supplied with the inner product $(\mathbf{u},\mathbf{v}) = \mathbf{u}'\mathbf{v}$. We can directly check that

$$\mathbf{J}' = \mathbf{J}, \quad \mathbf{J}^2 = \mathbf{J}, \quad \mathbf{J}\mathbf{v} = \mathbf{0} \quad \text{for } \mathbf{v} \perp G, \quad \mathbf{J}\mathbf{u}_h = \mathbf{u}_h,$$

$h = 1, \dots, m$, so for any $\mathbf{y} = \mathbf{u} + \mathbf{v} \in \mathbb{R}^p$ with $\mathbf{u} \in G$, $\mathbf{v} \perp G$ we have $\mathbf{J}\mathbf{y} = \mathbf{u} \in G$. Thus $\mathbf{J}$ is the orthogonal projector of $\mathbb{R}^p$ onto G.

We now deduce identities concerning the objects in Lemma 3.4. Using the $p \times n$ real matrix $\mathbf{C}$ with $\mathbf{C}'\mathbf{C} = \mathbf{\Delta}\mathbf{T}$, from (4.5) we get

$$\mathbf{M} = (\mathbf{e}_1 \dots \mathbf{e}_m)'\mathbf{C}'\mathbf{C}(\mathbf{e}_1 \dots \mathbf{e}_m).$$

In Lemma 3.4 let $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ be a basis for the subspace Φ of $\mathbb{R}^n$. Then $\{\mathbf{C}\mathbf{e}_1, \dots, \mathbf{C}\mathbf{e}_m\}$ is a basis $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ for the subspace $G = \mathbf{C}\Phi$ of $\mathbb{R}^p$ (see notation (3.12)). From (4.6) we see that

$$\mathbf{M}^{-1} = (\mu_{fg}), \tag{4.8}$$

and the expression (4.7) for the orthogonal projector $\mathbf{J}$ onto the subspace $G = \mathbf{C}\Phi$ of $\mathbb{R}^p$ becomes $\mathbf{J} = \sum_{f,g=1}^{m} \mu_{fg}\mathbf{C}\mathbf{e}_f\mathbf{e}_g'\mathbf{C}'$. Hence we get the first identity

$$(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{J}\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{\Delta} = \sum_{f,g=1}^{m} \mu_{fg}\mathbf{e}_f\mathbf{e}_g'\mathbf{\Delta}. \tag{4.9}$$

The second is given below.

PROPOSITION 4.1. *If $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ is an o.n. system in $\mathbb{R}^n$ according to the inner product $(\mathbf{u},\mathbf{v}) = \mathbf{u}'\mathbf{v}$, then*

$$\Big\| \sum_{f,g=1}^{m} \mu_{fg}\mathbf{e}_f\mathbf{e}_g'\mathbf{\Delta} \Big\|^2 = \sum_{i=1}^{k} \|\mathbf{F}^{-1}(i)\mathbf{P}'(i)\mathbf{E}_i'\|^2. \tag{4.10}$$

On the right-hand side, every summand with $\operatorname{Rank}\mathbf{E}_i = 0$, *if any, vanishes.*

Proof. First note $(\sum \mu_{gf}\mathbf{e}_g\mathbf{e}_f')^2 = \sum_{f,g,q} \mu_{gf}\mu_{fq}\mathbf{e}_g\mathbf{e}_q'$. Then using $\operatorname{Tr}\mathbf{A}\mathbf{A}' = \|\mathbf{A}\|^2$ repeatedly, noting $\mathbf{\Delta}\mathbf{e}_g = (a(1)\mathbf{e}_{g1}' \dots a(k)\mathbf{e}_{gk}')'$, we have

$$\begin{aligned}\Big\| \sum_{f,g=1}^{m} \mu_{fg}\mathbf{e}_f\mathbf{e}_g'\mathbf{\Delta} \Big\|^2 &= \operatorname{Tr}\mathbf{\Delta}\Big(\sum \mu_{gf}\mathbf{e}_g\mathbf{e}_f'\Big)^2\mathbf{\Delta} = \sum_{f=1}^{m} \operatorname{Tr}\sum_{g,q} \mu_{fg}\mathbf{\Delta}\mathbf{e}_g\mu_{fq}\mathbf{e}_q'\mathbf{\Delta} \\ &= \sum_{f=1}^{m} \Big\| \sum_{g=1}^{m} \mu_{fg}\mathbf{\Delta}\mathbf{e}_g \Big\|^2 = \sum_{i=1}^{k} a^2(i) \sum_{f=1}^{m} \Big\| \sum_{g=1}^{m} \mu_{fg}\mathbf{e}_{gi}' \Big\|^2.\end{aligned}$$

From (4.8) and (4.4), $(\mu_{fg}) = \mathbf{M}^{-1} = a^{-1}(i)\mathbf{P}(i)\mathbf{F}^{-1}(i)\mathbf{P}'(i)$. Using $\|\mathbf{P}\mathbf{A}\|^2 = \|\mathbf{A}\|^2$ for $\mathbf{P}$ orthogonal, we further have

$$a^2(i) \sum_{f=1}^{m} \Big\| \sum_{g=1}^{m} \mu_{fg}\mathbf{e}_{gi}' \Big\|^2 = a^2(i)\|\mathbf{M}^{-1}(\mathbf{e}_{1i} \dots \mathbf{e}_{mi})'\|^2 = \|\mathbf{F}^{-1}(i)\mathbf{P}'(i)\mathbf{E}_i'\|^2.$$

Then (4.10) follows. If $\operatorname{Rank}\mathbf{E}_i = 0$, $\mathbf{M}_i = \mathbf{0}$ then from (4.2), $\mathbf{P}(i)$ is an arbitrary orthogonal matrix, hence the relevant summand on the right-hand side of (4.10) vanishes. ■

We now have the third identity.

PROPOSITION 4.2. *Assume that* $\nu(i) = \operatorname{Rank}\mathbf{E}_i \geq 1$. *For the subspace (see* (1.8)*)* $\mathcal{M}(\mathbf{T}_i^{1/2}\mathbf{E}_i)$ *of* $\mathbb{R}^{n_i}$ *let* $\{\varphi_1, \ldots, \varphi_{\nu(i)}\}$ *be an o.n. basis according to the inner product* $(\mathbf{u}, \mathbf{v}) = \mathbf{u}'\mathbf{v}$. *Then for any* $p \times m$ *matrix* $\mathbf{U}$,

$$\mathbf{U}\mathbf{E}_i' = \sum_{j=1}^{\nu(i)} (\varphi_j' \otimes \mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}\varphi_j)\mathbf{T}_i^{-1/2}. \tag{4.11}$$

Proof. Complete $\{\varphi_1, \ldots, \varphi_{\nu(i)}\}$ to an o.n. basis $\{\varphi_1, \ldots, \varphi_{n_i}\}$ for $\mathbb{R}^{n_i}$. Using (3.7) note that, for any $n_i \times 1$ vector $\mathbf{v}$, $\sum_{f=1}^{n_i}(\varphi_f' \otimes \varphi_f)\mathbf{v} = \sum_{f=1}^{n_i}(\varphi_f'\mathbf{v})\varphi_f = \mathbf{v}$, hence $\sum_{f=1}^{n_i}(\varphi_f' \otimes \varphi_f) = \mathbf{I}_{n_i}$. Then

$$\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2} = \sum_{j=1}^{n_i} \mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}(\varphi_j' \otimes \varphi_j).$$

Let $\mathbf{g}$ be any $m \times 1$ vector. With (1.6) we have $\operatorname{vec}(\mathbf{A}\mathbf{B}\mathbf{C}') = (\mathbf{C} \otimes \mathbf{A}) \operatorname{vec}\mathbf{B}$ for matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, hence

$$(\varphi_j \otimes \varphi_j')\mathbf{T}_i^{1/2}\mathbf{E}_i\mathbf{g} = \begin{cases} \operatorname{vec}(\varphi_j'\mathbf{T}_i^{1/2}\mathbf{E}_i\mathbf{g}\varphi_j') & \text{for } j \leq \nu(i), \\ \mathbf{0} & \text{for } j > \nu(i), \end{cases}$$

thus, by transposing,

$$\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}(\varphi_j' \otimes \varphi_j) = \mathbf{0} \quad \text{for } j > \nu(i).$$

But from (3.3),

$$\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}(\varphi_j' \otimes \varphi_j) = \varphi_j' \otimes \mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}\varphi_j,$$

therefore we get (4.11). ■

4.2. Majorizing. Due to (4.10) we shall try to majorize the summands in its right-hand side, which is an important step in the proof of Lemma 3.4.

PROPOSITION 4.3. *Assume* $\nu(i) = \operatorname{Rank}\mathbf{E}_i \geq 1$ *and* $\operatorname{Rank}\mathbf{E} = m$. *Let* $\mathbf{N}_i$ *be some* $\nu(i) \times \nu(i)$ *p.d. principal submatrix of* $\mathbf{M}_i$. *Then there exist two positive functions* $d(\cdot)$ *and* $w_\nu(\cdot)$, *defined and continuous on* $\mathcal{M}_{m\times m}$ *and* $\mathcal{M}_{\mathrm{pd}}(\nu)$ *respectively, such that*

$$\|\mathbf{F}^{-1}(i)\mathbf{P}'(i)\mathbf{E}_i'\|^2 \leq d(\mathbf{M}_i)w_{\nu(i)}(\mathbf{N}_i)(\operatorname{Tr}\mathbf{T}_i^{-1}) \sum_{1\leq f\leq\nu(i),\, 1\leq g\leq m} \varphi_{fg}^2(i), \tag{4.12}$$

where $\varphi_{fg}(i)$ *is the* (f, g) *entry of* $\mathbf{F}^{-1}(i)$. *In particular, when* $\nu(i) = m$ *there exists a positive function* $c(\cdot)$, *defined and continuous on* $\mathcal{M}_{\mathrm{pd}}(m)$, *such that*

$$\|\mathbf{F}^{-1}(i)\mathbf{P}'(i)\mathbf{E}_i'\|^2 \leq c(\mathbf{M}_i) \operatorname{Tr}\mathbf{T}_i^{-1}. \tag{4.13}$$

Proof. Using notation (1.7), define the functions

$$\begin{aligned} d(\mathbf{A}) &= m^2 \operatorname{Tr}(\mathbf{A}^2) \quad \text{for } \mathbf{A} \in \mathcal{M}_{m\times m}, \\ w_\nu(\mathbf{D}) &= \sum_{j=1}^{\nu} \operatorname{Tr}(\mathbf{D}_j^{-1}) \quad \text{for } \mathbf{D} \in \mathcal{M}_{\mathrm{pd}}(\nu),\ \mathbf{D}_j = \mathbf{D}(\{1, \ldots, j\}), \\ c(\mathbf{B}) &= d(\mathbf{B})w_m(\mathbf{B})(\operatorname{Tr}\mathbf{B}^{-1})^2 \quad \text{for } \mathbf{B} \in \mathcal{M}_{\mathrm{pd}}(m). \end{aligned}$$

The topology in $\mathcal{M}_{m\times m}$ is assumed to be induced by an arbitrary norm or, equivalently, by the Euclidean norm. Then, from Proposition 3.1, $\mathcal{M}_{\mathrm{pd}}(\nu)$ is an open set in $\mathcal{M}_{\nu\times\nu}$ and the functions $d(\cdot), w_\nu(\cdot), c(\cdot)$ are continuous on their respective domains of definition.

Write $\nu(i)=\nu$ briefly. Since $\operatorname{Rank}\mathbf{T}_i^{1/2}\mathbf{E}_i = \operatorname{Rank}\mathbf{E}_i=\nu$, among the m columns of $\mathbf{T}_i^{1/2}\mathbf{E}_i$ there are ν linearly independent vectors, denoted by $\mathbf{v}_1,\ldots,\mathbf{v}_\nu$. Put

$$\mathbf{N}_i=(\mathbf{v}_1\ldots\mathbf{v}_\nu)'(\mathbf{v}_1\ldots\mathbf{v}_\nu),\quad \mathbf{D}_j=(\mathbf{v}_1\ldots\mathbf{v}_j)'(\mathbf{v}_1\ldots\mathbf{v}_j),\quad j=1,\ldots,\nu.$$

Since $\{\mathbf{v}_1,\ldots,\mathbf{v}_\nu\}$ is a basis for $\mathcal{M}(\mathbf{T}_i^{1/2}\mathbf{E}_i)$ (see (1.8)), the Gram–Schmidt orthogonalization process gives an o.n. basis $\{\varphi_1,\ldots,\varphi_\nu\}$, namely

$$\varphi_j=(\mathbf{v}_1\ldots\mathbf{v}_j)(d_{j1}\ldots d_{jj})',\quad j=1,\ldots,\nu,$$

where, δ_{pq} denoting the (p,q) entry of $\mathbf{D}_j^{-1}$ for $j\ge 1$,

$$d_{jq}=(\det\mathbf{D}_{j-1})^{-1/2}(\det\mathbf{D}_j)^{1/2}\delta_{jq},\quad q=1,\ldots,j.$$

By convention, take $\det\mathbf{D}_0=1$ to ensure the formula consistency. Since $\mathbf{D}_j$'s are p.d. we have

$$\delta_{jq}^2<\delta_{jj}\delta_{qq}=(\det\mathbf{D}_{j-1})(\det\mathbf{D}_j)^{-1}\delta_{qq},$$

hence $d_{jq}^2<\delta_{qq}$ for $q<j$, whereas

$$d_{jj}^2=(\det\mathbf{D}_{j-1})^{-1}(\det\mathbf{D}_j)\delta_{jj}^2=\delta_{jj}.$$

Thus $\sum_{q=1}^{j} d_{jq}^2\le\operatorname{Tr}(\mathbf{D}_j^{-1})$. Therefore, for any $p\times m$ matrix $\mathbf{U}$ we get

$$\|\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}\varphi_j\|^2\le\|\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}(\mathbf{v}_1\ldots\mathbf{v}_j)\|^2\operatorname{Tr}(\mathbf{D}_j^{-1}).$$

From (4.1), $\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}(\mathbf{v}_1\ldots\mathbf{v}_j)$ is a submatrix of $\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}\mathbf{T}_i^{1/2}\mathbf{E}_i=\mathbf{U}\mathbf{M}_i$, hence

$$\|\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}\varphi_j\|^2\le\|\mathbf{U}\mathbf{M}_i\|^2\operatorname{Tr}(\mathbf{D}_j^{-1}),\quad j=1,\ldots,\nu.$$

From (4.11), noting $\|\varphi_j'\|^2=1$ and $\|\mathbf{A}_1+\ldots+\mathbf{A}_\nu\|^2\le\nu(\|\mathbf{A}_1\|^2+\ldots+\|\mathbf{A}_\nu\|^2)$ for matrix sum norm, we then have

$$\begin{aligned}\|\mathbf{U}\mathbf{E}_i'\|^2&\le\nu\sum_{j=1}^{\nu}\|\mathbf{U}\mathbf{E}_i'\mathbf{T}_i^{1/2}\varphi_j\|^2\|\mathbf{T}_i^{-1/2}\|^2\le m\|\mathbf{U}\mathbf{M}_i\|^2\|\mathbf{T}_i^{-1/2}\|^2\sum_{j=1}^{\nu}\operatorname{Tr}(\mathbf{D}_j^{-1})\\&=m\|\mathbf{U}\mathbf{M}_i\|^2w_\nu(\mathbf{N}_i)\operatorname{Tr}(\mathbf{T}_i^{-1}).\end{aligned}$$

By (4.2) put $\mathbf{U}=\mathbf{F}^{-1}(i)\mathbf{P}'(i)$. Noting $\|\mathbf{P}'(i)\|^2=m$ we get

$$\begin{aligned}\|\mathbf{U}\mathbf{M}_i\|^2&=\|\mathbf{F}^{-1}(i)\mathbf{\Lambda}(i)\mathbf{P}'(i)\|^2\le m\|\mathbf{F}^{-1}(i)\mathbf{\Lambda}(i)\|^2=m\sum_{f,g}\varphi_{fg}^2(i)\lambda_f^2(i)\\&\le m(\operatorname{Tr}\mathbf{M}_i^2)\sum_{f,g}\varphi_{fg}^2(i),\quad 1\le f\le\nu,\ 1\le g\le m,\end{aligned}$$

since $\operatorname{Tr}\mathbf{M}_i^2=\sum_{f=1}^{\nu}\lambda_f^2(i)$. Thus,

$$\|\mathbf{F}^{-1}(i)\mathbf{P}'(i)\mathbf{E}_i'\|^2\le d(\mathbf{M}_i)w_\nu(\mathbf{N}_i)(\operatorname{Tr}\mathbf{T}_i^{-1})\sum\varphi_{fg}^2(i),$$

i.e. we get (4.12). Now, from (4.4) and (4.3) we have

$$\mathbf{F}(i)=\mathbf{G}_1+\ldots+\mathbf{G}_k,$$

where $\mathbf{G}_1 = \mathbf{\Lambda}(i)$ and the other summands are n.n.d. matrices

$$a^{-1}(i)a(h)\mathbf{P}'(i)\mathbf{M}_h\mathbf{P}(i), \quad h = 1, \ldots, k, \ h \neq i.$$

In the case $\nu = \operatorname{Rank}\mathbf{M}_i = \operatorname{Rank}\mathbf{G}_1 = m$, from Remark 4.1 below we have

$$\|\mathbf{F}^{-1}(i)\|^2 \leq (\operatorname{Tr}\mathbf{\Lambda}^{-1}(i))^2 = (\operatorname{Tr}\mathbf{M}_i^{-1})^2,$$

therefore, using $\mathbf{N}_i = \mathbf{M}_i$ due to $\nu = m$, (4.12) becomes

$$\|\mathbf{F}^{-1}(i)\mathbf{P}'(i)\mathbf{E}_i'\|^2 \leq d(\mathbf{M}_i)w_m(\mathbf{M}_i)(\operatorname{Tr}\mathbf{T}_i^{-1})(\operatorname{Tr}\mathbf{M}_i^{-1})^2 = c(\mathbf{M}_i)\operatorname{Tr}\mathbf{T}_i^{-1}.$$

Proposition 4.3 is proved.

REMARK 4.1. For a p.d. matrix $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ and $m \times m$ n.n.d. $\mathbf{G}_2, \ldots, \mathbf{G}_k$ let $\mathbf{F} = \mathbf{\Lambda} + \sum_{i=2}^k \mathbf{G}_i$. Then

$$\|\mathbf{F}^{-1}\|^2 \leq (\operatorname{Tr}\mathbf{\Lambda}^{-1})^2.$$

Indeed, the cases $m = 1$ and $\mathbf{F} = \mathbf{\Lambda}$ are trivial. Consider $k, m \geq 2$. For $z > 0$ write $\mathbf{H} = \sum_{i=2}^k \mathbf{G}_i$, $\mathbf{F}(z) = \mathbf{\Lambda} + z\mathbf{H}$, $\mathbf{F}^{-1}(z) = (\varphi_{fg}(z))$; when z is small,

$$\varphi_{ff}(z) = \frac{\det\mathbf{F}_{ff}(z)}{\det\mathbf{F}(z)} \approx \frac{\det\mathbf{\Lambda}_{ff}}{\det\mathbf{\Lambda}} = \lambda_f^{-1}, \quad f = 1, \ldots, m.$$

For $z > h > 0$, $(\mathbf{\Lambda} + z\mathbf{H}) - (\mathbf{\Lambda} + (z-h)\mathbf{H})$ is n.n.d. hence (see Rao (1973), p. 70, Problem 9(iii)) $(\mathbf{\Lambda}+(z-h)\mathbf{H})^{-1} - (\mathbf{\Lambda}+z\mathbf{H})^{-1}$ is n.n.d., therefore $\varphi_{ff}(z-h) - \varphi_{ff}(z) \geq 0$, so $\varphi_{ff}(z)$ is non-increasing in z, thus $\varphi_{ff}(z) \uparrow \lambda_f^{-1}$ as $z \downarrow 0$, hence $0 < \varphi_{ff}(z) \leq \lambda_f^{-1}$ for all f. Now write $\mathbf{F}^{-1} = (\varphi_{fg})$ to get

$$\|\mathbf{F}^{-1}\|^2 = \sum_{f,g} \varphi_{fg}^2 < \sum_{f,g} \varphi_{ff}\varphi_{gg} \leq \Big(\sum \lambda_f^{-1}\Big)^2 = (\operatorname{Tr}\mathbf{\Lambda}^{-1})^2.$$

The following corollary, which will be used for proving (3.24) in Lemma 3.4, follows from (4.10) and (4.13).

COROLLARY 4.1. *If the system* $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ *is o.n. in* $\mathbb{R}^n$ *for the inner product* $(\mathbf{u}, \mathbf{v}) = \mathbf{u}'\mathbf{v}$ *and* $\operatorname{Rank}\mathbf{E}_i = m$ *for all* $i = 1, \ldots, k$, *then there exists a positive function* $c(\cdot)$, *defined and continuous on* $\mathcal{M}_{\mathrm{pd}}(m)$, *such that*

$$\Big\| \sum_{f,g=1}^m \mu_{fg}\mathbf{e}_f\mathbf{e}_g'\mathbf{\Delta} \Big\|^2 \leq \sum_{i=1}^k c(\mathbf{M}_i)\operatorname{Tr}\mathbf{T}_i^{-1}.$$

Due to (4.12) to prove the boundedness of the summands on the right-hand side of (4.10) we are led to proving the boundedness of $\sum \varphi_{fg}^2(i)$, $\varphi_{fg}(i)$ being the (f,g) entry of $\mathbf{F}^{-1}(i)$: this is just the most involved step in the proof of Lemma 3.4. Noting that, by (4.3) and (4.4), $\mathbf{F}(i)$ is a linear combination of n.n.d. matrices, we have to investigate the inverse of such a matrix combination; the study of such inverses, though laborious, is unavoidable when we explore the GLSE expression, so we shall digress a little in order to examine the determinant expansion for linear matrix combinations.

4.3. Linear matrix combinations, determinant expansion. The main tool for proving Lemma 3.4 is just the sign properties of coefficients in the determinant expansion for a linear combination of matrices and is presented in this subsection. Let there be given $m \times m$ real matrices $\mathbf{A}_1, \dots, \mathbf{A}_k$, $m, k \geq 1$. Consider the integers $m_1, \dots, m_k \geq 0$, $m_1 + \dots + m_k = m$ and the polynomial $\det(\sum_{i=1}^k t_i \mathbf{A}_i)$ in the numerical variables $t_1, \dots, t_k$; we shall use the symbol

$$\mathbf{A}_1(m_1) \circ \dots \circ \mathbf{A}_k(m_k) = \text{ coefficient of } t_1^{m_1} \dots t_k^{m_k} \text{ in } \det\Big(\sum_{i=1}^k t_i \mathbf{A}_i\Big).$$

Thus we have the determinant expansion

$$\det\Big(\sum_{i=1}^k t_i \mathbf{A}_i\Big) = \sum_{m_1, \dots, m_k} t_1^{m_1} \dots t_k^{m_k} \mathbf{A}_1(m_1) \circ \dots \circ \mathbf{A}_k(m_k), \tag{4.14}$$

and get

Proposition 4.4. *$\mathbf{A}_1(m_1) \circ \dots \circ \mathbf{A}_k(m_k)$ is obtained by deleting m_i rows from $\mathbf{A}_i$, $i = 1, \dots, k$, and taking the sum of $m!/(m_1! \dots m_k!)$ determinants, each formed with the deleted rows, their original ordinals being conserved.*

Proof. Write $\mathbf{A}_i = (a_{jh}(i))$, $\sum_i t_i \mathbf{A}_i = (\sum_i t_i a_{jh}(i))$. In $\det \sum t_i \mathbf{A}_i$ the term with $t_1^{m_1} \dots t_k^{m_k}$ is generated by the following process:

(i) Take a partition $\{D_1, \dots, D_k\}$ of $\{1, \dots, m\}$ with $\sharp D_i = m_i$, $i = 1, \dots, k$.

(ii) When forming $\det \sum t_i \mathbf{A}_i$, consider the product $\pm \prod_{j=1}^m \sum_i t_i a_{jh_j}(i)$ corresponding to some permutation $\{h_1, \dots, h_m\}$ of $\{1, \dots, m\}$, and when forming this product take the summand element $t_i a_{jh_j}(i)$ for every $j \in D_i$; then a term with $t_1^{m_1} \dots t_k^{m_k}$ is produced.

(iii) Summing these terms over all permutations $\{h_1, \dots, h_m\}$ gives a determinant whose rows with ordinals in D_i are from $t_i \mathbf{A}_i$, $i = 1, \dots, k$. This determinant gives a term with $t_1^{m_1} \dots t_k^{m_k}$, corresponding to the chosen partition, namely $t_1^{m_1} \dots t_k^{m_k} \det(\mathbf{r}'_1 \dots \mathbf{r}'_m)'$, where for $j \in D_i$, $\mathbf{r}_j$ is the jth row of $\mathbf{A}_i$.

(iv) Summing these terms over all partitions $\{D_1, \dots, D_k\}$ with $\sharp D_i = m_i$ given, $i = 1, \dots, k$, gives the term with $t_1^{m_1} \dots t_k^{m_k}$ of $\det \sum t_i \mathbf{A}_i$, which is just $t_1^{m_1} \dots t_k^{m_k} \mathbf{A}_1(m_1) \circ \dots \circ \mathbf{A}_k(m_k)$. ∎

We shall present three sign properties of coefficients in Propositions 4.5–4.7. The first of these is basic, it concerns the positivity of coefficients which plays a key role in the study of the inverse of a linear matrix combination.

Proposition 4.5. (i) *If $\mathbf{A}_1(m_1) \circ \dots \circ \mathbf{A}_k(m_k) \neq 0$ then*

$$\operatorname{Rank}(\dots \mathbf{A}'_i \dots)_{i \in \varphi} \geq \sum_{i \in \varphi} m_i \quad \forall \varphi \subset \{1, \dots, k\},\ \varphi \neq \emptyset. \tag{4.15}$$

On the right-hand side stands a block matrix formed with blocks $\mathbf{A}_i$, the index i ranging through the subset φ from left to right.

(ii) *For n.n.d. $\mathbf{A}_1, \dots, \mathbf{A}_k$ the coefficient $\mathbf{A}_1(m_1) \circ \dots \circ \mathbf{A}_k(m_k)$ is non-negative, and it is positive if and only if* (4.15) *holds.*

Proof of (i). Let (4.15) be false, i.e.

$$(\exists \varphi \subset \{1, \dots, k\},\ \varphi \neq \emptyset) \quad \operatorname{Rank}(\dots \mathbf{A}'_i \dots)_{i \in \varphi} \leq \sum_{i \in \varphi} m_i - 1.$$

Then any set of $\sum_\varphi m_i$ rows, consisting of m_i rows drawn from each $\mathbf{A}_i$ with $i \in \varphi$, is linearly dependent. Thus every determinant formed with m_i rows from each $\mathbf{A}_i$, $i = 1, \dots, k$, vanishes and by Proposition 4.4, $\mathbf{A}_1(m_1) \circ \dots \circ \mathbf{A}_k(m_k) = 0$. ∎

The proof of (ii), long and involved, is deferred to Section 5.

The following proposition explores the index set of positive coefficients in the polynomial $\det(\sum_{i=1}^k t_i \mathbf{M}_i)$.

PROPOSITION 4.6. *For* $i = 1, \dots, k$, $k \geq 2$ *consider the real matrices:* $\mathbf{T}_i$ $n_i \times n_i$ *p.d.*; $\mathbf{E}_i$ $n_i \times m$, $\mathbf{E} = (\mathbf{E}'_1 \dots \mathbf{E}'_k)'$ *with* $\operatorname{Rank} \mathbf{E}_1 > 0$ *and* $\operatorname{Rank} \mathbf{E} = m \geq 2$; $\mathbf{M}_i = \mathbf{E}'_i \mathbf{T}_i \mathbf{E}_i$ $m \times m$ *n.n.d. Put*

(4.16) $\quad u = (\psi, \{s_i : i \in \psi\})$ *for each set* $\psi \subset \{2, \dots, k\}$ *and*

for positive integers s_i *such that* $\sum_{i \in \psi} s_i \leq m - 1$.

Setting $s_1 = m - \sum_{i \in \psi} s_i$, *consider the coefficient*

$$d_u = \mathbf{M}_1(s_1) \circ \dots \circ \mathbf{M}_i(s_i) \circ \dots, \quad i \in \psi. \tag{4.17}$$

Then the set $U = \{u : \ d_u > 0\}$ *is non-void, finite, independent of* $\mathbf{T}_1, \dots, \mathbf{T}_k$ *and is entirely determined by the matrix* $\mathbf{E}$.

Proof. Consider positive numbers t_i with $t_1 = 1$. Set

$$\mathbf{H} = \sum_{i=1}^k t_i \mathbf{M}_i = \mathbf{E}' \operatorname{diag}(t_i \mathbf{T}_i) \mathbf{E}.$$

Then $\operatorname{Rank} \mathbf{H} = \operatorname{Rank} \mathbf{E} = m$, hence $\mathbf{H}$ is p.d. In the polynomial $\det \mathbf{H}$ in $t_2, \dots, t_k$ each term of degree less than m corresponds biunivocally to one index u, and the uth term $d_u \prod_{i \in \psi} t_i^{s_i}$ has coefficient d_u which is non-negative by Proposition 4.5(ii). Set

$$z = (t_2^2 + \dots + t_k^2)^{1/2} \quad \text{and} \quad \nu = \operatorname{Rank} \mathbf{M}_1 = \operatorname{Rank} \mathbf{E}_1 \geq 1,$$

and consider the spectral decomposition

$$\mathbf{M}_1 = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}', \quad \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_\nu, \dots), \quad \lambda_1 \geq \dots \geq \lambda_\nu > 0.$$

Put $\mathbf{Q}_i = \mathbf{P}'\mathbf{M}_i\mathbf{P}$ and $\mathbf{G}_0 = \sum_{i=2}^k (t_i/z)\mathbf{Q}_i$. Then

$$\mathbf{P}'\mathbf{H}\mathbf{P} = \boldsymbol{\Lambda} + \sum_{i=2}^k t_i \mathbf{Q}_i = \boldsymbol{\Lambda} + z\mathbf{G}_0 \quad \text{is p.d.}$$

Disregard the expression of $\mathbf{G}_0$. Then in the formal expansion of $\det \mathbf{H} = \det(\boldsymbol{\Lambda} + z\mathbf{G}_0)$ the term with $z^{m-\nu}$, by (4.14), is $z^{m-\nu}\boldsymbol{\Lambda}(\nu) \circ \mathbf{G}_0(m-\nu)$ whose coefficient, by Proposition 4.4, equals the determinant formed by piling the first ν rows of $\boldsymbol{\Lambda}$ on the last $m - \nu$ rows of $\mathbf{G}_0$; hence this term is positive for $\nu = m$, whereas for $\nu < m$ with notation (1.7) it is $z^{m-\nu}\lambda_1 \dots \lambda_\nu \det \mathbf{G}_0(\{\nu+1, \dots, m\})$ and it is also positive since $\det(z\mathbf{G}_0(\{\nu+1, \dots, m\}))$ is a principal minor of the p.d. matrix $\boldsymbol{\Lambda} + z\mathbf{G}_0$. But this positive term is just the sum of terms of degree $m - \nu$ in $t_2, \dots, t_k$ in the polynomial $\det \mathbf{H} = \det(\boldsymbol{\Lambda} + \sum_{i=2}^k t_i \mathbf{Q}_i)$. These terms are non-negative, hence there exists a positive one among them. Therefore the set U of those indices u for which the coefficients d_u are positive is non-void. Further, from

Proposition 4.5(ii), d_u is positive if and only if

$$\operatorname{Rank}(\dots \mathbf{M}_i \dots)_{i\in\varphi} \geq \sum_{i\in\varphi} s_i \quad \forall \varphi \subset \psi \cup \{1\},\ \varphi \neq \emptyset.$$

By (3.39), $\ker \mathbf{M}_i = \ker \mathbf{E}_i$, hence

$$\ker(\dots \mathbf{M}_i' \dots)_{i\in\varphi}' = \ker(\dots \mathbf{E}_i' \dots)_{i\in\varphi}'.$$

Thus by (3.36), d_u is positive if and only if

$$\operatorname{Rank}(\dots \mathbf{E}_i' \dots)_{i\in\varphi}' \geq \sum_{i\in\varphi} s_i \quad \forall \varphi \subset \psi \cup \{1\},\ \varphi \neq \emptyset.$$

This condition is quite independent of $\mathbf{T}_1, \dots, \mathbf{T}_k$, so the set $U = \{u : d_u > 0\}$ is entirely determined by the matrix $\mathbf{E}$. ■

We further deal with the vanishing of coefficients in the determinant expansion for a linear combination of n.n.d. matrices and the cofactors of its elements.

PROPOSITION 4.7. *Let $\mathbf{Q}_2, \dots, \mathbf{Q}_k$ be $m \times m$ n.n.d. matrices, $k, m \geq 2$ and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ with $\operatorname{Rank} \mathbf{\Lambda} = \nu \geq 1$, $\lambda_1, \dots, \lambda_\nu > 0$. Let $\mathbf{F} = \mathbf{\Lambda} + \sum_{i=2}^k t_i \mathbf{Q}_i$ with real $t_2, \dots, t_k$. Consider any set $\{s_i : i \in \psi\}$, where $\psi \subset \{2, \dots, k\}$ and s_i are positive integers with $\sum_{i\in\psi} s_i \leq m-1$. Then, if in the expansion of $\det \mathbf{F}$ the coefficient of $\prod_{i\in\psi} t_i^{s_i}$ vanishes, so does the same coefficient in the expansion of $\det \mathbf{F}_{fg}$ for fixed (f, g), $f = 1, \dots, \nu$ and $g = 1, \dots, m$.*

Proof. Put $s_1 = m - \sum_{i\in\psi} s_i \geq 1$, $r_1 = s_1 - 1$ and $\mathbf{Q}_1 = \mathbf{\Lambda}$. Put $\mathbf{R}_i = \mathbf{Q}_{i,fg}$, the submatrix obtained by deleting the fth row and gth column of $\mathbf{Q}_i$. Then $\mathbf{R}_1 = \mathbf{\Lambda}_{fg}$. The coefficient of $\prod_{i\in\psi} t_i^{s_i}$ in the expansion of $\det \mathbf{F}$ is $\mathbf{Q}_1(s_1) \circ \dots \circ \mathbf{Q}_i(s_i) \circ \dots,\ i \in \psi$, which from Proposition 4.5(ii) vanishes if and only if

$$(\exists \varphi \subset \psi \cup \{1\},\ \varphi \neq \emptyset) \quad \operatorname{Rank}(\dots \mathbf{Q}_i \dots)_{i\in\varphi} \leq \sum_{i\in\varphi} s_i - 1. \tag{4.18}$$

Similarly, in the expansion of $\det \mathbf{F}_{fg} = \det(\mathbf{R}_1 + \sum_{i=2}^k t_i \mathbf{R}_i)$ the coefficient of $\prod_{i\in\psi} t_i^{s_i}$ is

$$\mathbf{R}_1(r_1) \circ \dots \circ \mathbf{R}_i(s_i) \circ \dots, \quad i \in \psi. \tag{4.19}$$

Starting from Assumption (4.18), we shall consider three cases.

(a) $\varphi \subset \psi$. Since $(\dots \mathbf{R}_i' \dots)_{i\in\varphi}$ is a submatrix of $(\dots \mathbf{Q}_i' \dots)_{i\in\varphi}$, from (4.18) we have

$$\operatorname{Rank}(\dots \mathbf{R}_i' \dots)_{i\in\varphi} \leq \sum_{i\in\varphi} s_i - 1,$$

which, by Proposition 4.5(i), entails the vanishing of (4.19).

(b) $\varphi = \{1\}$. Now (4.18) becomes $\operatorname{Rank} \mathbf{\Lambda} \leq s_1 - 1$. Since $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_\nu, \dots)$, $\nu \geq 1$ and $\lambda_1 \dots \lambda_\nu \neq 0$, for $f = 1, \dots, \nu$ we have $\operatorname{Rank} \mathbf{\Lambda}_{fg} \leq \operatorname{Rank} \mathbf{\Lambda} - 1$, hence $\operatorname{Rank} \mathbf{R}_1 \leq r_1 - 1$. Thus (4.19) vanishes by Proposition 4.5(i).

(c) $\{1\} \subset \varphi$ and $\varphi \cap \psi \neq \emptyset$. Then (4.18) is rewritten as

$$\operatorname{Rank}(\mathbf{\Lambda} \dots \mathbf{Q}_i \dots)_{i\in\varphi\cap\psi} \leq \sum_{i\in\varphi\cap\psi} s_i + s_1 - 1 = \sum_{i\in\varphi\cap\psi} s_i + r_1 \leq m - 1$$

since $\sum_{i\in\psi} s_i + s_1 = m$. Thus we have

$$m - n = \mathrm{Rank}(\mathbf{\Lambda} \ldots \mathbf{Q}_i \ldots)'_{i\in\varphi\cap\psi} \le \sum_{i\in\varphi\cap\psi} s_i + r_1, \quad \text{where } n \ge 1. \tag{4.20}$$

Let $\mathbf{C}_1, \ldots, \mathbf{C}_m$ denote the successive columns of the matrix $(\mathbf{\Lambda}' \ldots \mathbf{Q}'_i \ldots)'_{i\in\varphi\cap\psi}$. Then a vector $(z_1 \ldots z_m)'$ belongs to $\ker(\mathbf{\Lambda}' \ldots \mathbf{Q}'_i \ldots)'_{i\in\varphi\cap\psi}$ if and only if $z_1 = \ldots = z_\nu = 0$ and $(z_{\nu+1} \ldots z_m)' \in \ker(\mathbf{C}_{\nu+1} \ldots \mathbf{C}_m)$. Hence, from (4.20) using (3.36) we have

$$n = \dim\ker(\mathbf{\Lambda} \ldots \mathbf{Q}_i \ldots)'_{i\in\varphi\cap\psi} = \dim\ker(\mathbf{C}_{\nu+1} \ldots \mathbf{C}_m).$$

Then from (3.36) we have

$$m - \nu - n = \mathrm{Rank}(\mathbf{C}_{\nu+1} \ldots \mathbf{C}_m) \ge 0.$$

Therefore in $\mathbf{C}_{\nu+1} \ldots \mathbf{C}_m$ there are exactly $m - \nu - n$ linearly independent columns forming a submatrix $\mathbf{C}(m - \nu - n)$. Thus by (4.20),

$$m - n = \mathrm{Rank}(\mathbf{\Lambda} \ldots \mathbf{Q}_i \ldots)'_{i\in\varphi\cap\psi} = \mathrm{Rank}(\mathbf{C}_1 \ldots \mathbf{C}_m) = \mathrm{Rank}(\mathbf{C}_1 \ldots \mathbf{C}_\nu \vdots \mathbf{C}(m - \nu - n)).$$

Therefore we have all the $m - n$ columns linearly independent in the last matrix. In it and in the matrix $(\mathbf{\Lambda}' \ldots \mathbf{Q}'_i \ldots)'_{i\in\varphi\cap\psi}$ the deletion of the fth column, $1 \le f \le \nu$, thus gives rise to two new matrices with ranks both equal to $m - n - 1$. A fortiori, further deleting some rows, for $1 \le g \le m$ on account of (4.20) we have

$$\mathrm{Rank}(\mathbf{\Lambda}'_{gf} \ldots \mathbf{Q}'_{i,\,gf} \ldots)'_{i\in\varphi\cap\psi} \le (m - n) - 1 \le \sum_{i\in\varphi\cap\psi} s_i + r_1 - 1. \tag{4.21}$$

From (4.14), in the expansion of

$$\det \mathbf{F}_{gf} = \det\Big(\mathbf{\Lambda}_{gf} + \sum_{i=2}^{k} t_i \mathbf{Q}_{i,\,gf}\Big),$$

the coefficient of $\prod_{i\in\psi} t_i^{s_i}$ is $\mathbf{\Lambda}_{gf}(r_1) \circ \ldots \circ \mathbf{Q}_{i,\,gf}(s_i) \circ \ldots$, $i \in \psi$, which, because of (4.21), vanishes by Proposition 4.5(i). Since $\mathbf{F}$ is symmetric we have $\det \mathbf{F}_{fg} = \det \mathbf{F}_{gf}$, thus the coefficient of $\prod_{i\in\psi} t_i^{s_i}$ in the expansion of $\det \mathbf{F}_{fg}$ vanishes for $1 \le f \le \nu$ and $1 \le g \le m$. ∎

4.4. Proof of Lemma 3.4. The most involved task concerns $\sum \varphi_{fg}^2(i)$ in (4.12); it can now be achieved.

PROPOSITION 4.8. *Assume* $\nu = \mathrm{Rank}\,\mathbf{E}_1 \ge 1$, $\mathrm{Rank}\,\mathbf{E} = m \ge 2$. *Put* $\mathbf{F}^{-1}(1) = (\varphi_{fg})_{m,m}$ (see (4.4)). *Then, for all* $(\mathbf{T}_1, \ldots, \mathbf{T}_k)$ *in a fixed closed ball* B_0 *in* $\prod_{i=1}^k \mathcal{M}_{\mathrm{pd}}(n_i)$, *the sum* $\sum \varphi_{fg}^2$ ($f = 1, \ldots, \nu$ *and* $g = 1, \ldots, m$) *is bounded by a positive constant depending on* B_0 *and* $\mathbf{E}$ *only.*

Proof. From (4.5), $\mathbf{M}$ is p.d. Write

$$\mathbf{H} = a^{-1}(1)\mathbf{M} = \mathbf{M}_1 + \sum_{i=2}^{k} t_i \mathbf{M}_i,$$

where $t_i = a^{-1}(1)a(i) > 0$, $i = 2, \ldots, k$. From (4.2) and (4.4) write

$$\mathbf{P}(1) = \mathbf{P}, \quad \mathbf{\Lambda}(1) = \mathbf{\Lambda}, \quad \mathbf{F}(1) = \mathbf{F}, \quad \mathbf{Q}_i = \mathbf{P}'\mathbf{M}_i\mathbf{P}.$$

Then

$$\mathbf{F} = \mathbf{P}'\mathbf{H}\mathbf{P} = \mathbf{\Lambda} + \sum_{i=2}^{k} t_i \mathbf{Q}_i.$$

First let $k \geq 2$. Consider

$$(-1)^{f+g}\varphi_{fg} = \frac{\det \mathbf{F}_{fg}}{\det \mathbf{F}}.$$

For each set $\psi \subset \{2, \dots, k\}$ and for positive integers s_i such that $\sum_{i\in\psi} s_i \leq m-1$, the coefficient of $\prod_{i\in\psi} t_i^{s_i}$ in the expansion of $\det \mathbf{F} = \det \mathbf{H}$ is given by (4.17), namely

$$d_u = \mathbf{M}_1(s_1) \circ \dots \circ \mathbf{M}_i(s_i) \circ \dots, \; i \in \psi,$$

where $s_1 = m - \sum_{i\in\psi} s_i$ and the index u is defined by (4.16),

$$u = (\psi, \{s_i : i \in \psi\}).$$

The set $U = \{u : d_u > 0\}$, by Proposition 4.6, is non-void. By Proposition 4.5(ii) all summands in the expansion of $\det \mathbf{H} = \det \mathbf{F}$ are non-negative, hence by putting

$$\delta_u = d_u \prod_{i\in\psi} t_i^{s_i}$$

we have

$$\det \mathbf{F} \geq \sum_{u\in U} \delta_u > 0.$$

By Proposition 4.7, for $f = 1, \dots, \nu$ and $g = 1, \dots, m$, in the expansion of $\det \mathbf{F}_{fg}$ the coefficient of $\prod_{i\in\psi} t_i^{s_i}$ can be different from zero only if $u \in U$. Let $\mathbf{E}$ be given and (f, g) fixed. Then for

$$r_1 = s_1 - 1, \quad \mathbf{R} \in \mathcal{M}_{m\times m}, \quad \mathbf{D}_i = \mathbf{R}'\mathbf{M}_i\mathbf{R}$$

we define the functions $z_u(\mathbf{R}, \mathbf{T}_1, \dots, \mathbf{T}_k)$ through the matrices $\mathbf{D}_{i,\,fg}$ obtained by deleting the fth row and gth column of $\mathbf{D}_i$:

$$d_u z_u = \mathbf{D}_{1,\,fg}(r_1) \circ \dots \circ \mathbf{D}_{i,\,fg}(s_i) \circ \dots, \quad i \in \psi, \; u \in U.$$

In particular, choosing $\mathbf{R} = \mathbf{P}$ we have

$$\mathbf{F} = \mathbf{P}'\mathbf{M}_1\mathbf{P} + \sum_{i=2}^{k} t_i \mathbf{P}'\mathbf{M}_i\mathbf{P} = \mathbf{D}_1 + \sum_{i=2}^{k} t_i \mathbf{D}_i,$$

hence in the expansion of $\det \mathbf{F}_{fg}$ the possibly non-null coefficient of $\prod_{i\in\psi} t_i^{s_i}$ is just $d_u z_u$. Moreover, as $\mathbf{F}_{fg}$ is of order $(m-1)\times(m-1)$ and $r_1 + \sum_{i\in\psi} s_i = m-1$, $r_1 \geq 0$, the expansion of $\det \mathbf{F}_{fg}$ for $1 \leq f \leq \nu$ and $1 \leq g \leq m$ is

$$\det \mathbf{F}_{fg} = \sum_{u\in U} d_u z_u \prod_{i\in\psi} t_i^{s_i} = \sum_{u\in U} z_u \delta_u,$$

where $z_u = z_u(\mathbf{P}, \mathbf{T}_1, \dots, \mathbf{T}_k)$. Therefore,

$$|\varphi_{fg}| \leq \frac{\sum_{u\in U} |z_u|\delta_u}{\sum_{u\in U} \delta_u}, \quad 1 \leq f \leq \nu \text{ and } 1 \leq g \leq m.$$

Now, given the matrix $\mathbf{E}$, for fixed (f, g) and for each $u \in U$, the function $z_u(\mathbf{R}, \mathbf{T}_1, \dots, \mathbf{T}_k)$ is defined and is obviously continuous on the product set $\mathcal{M}_{m\times m} \times \prod_{i=1}^{k} \mathcal{M}_{\mathrm{pd}}(n_i)$ whose topology is induced by an arbitrary norm on the product space $\mathcal{M}_{m\times m} \times \prod_{i=1}^{k} \mathcal{M}_{n_i\times n_i}$ (see Proposition 3.1). A fortiori $z_u(\mathbf{R}, \mathbf{T}_1, \ldots, \mathbf{T}_k)$ is continuous on the closed and bounded, hence compact, subset $\{\mathbf{R} \in \mathcal{M}_{m\times m} : \|\mathbf{R}\|^2 = m\} \times B_0$; the existence of closed balls B_0 in $\prod_{i=1}^{k} \mathcal{M}_{\mathrm{pd}}(n_i)$ being ensured by Proposition 3.1. On this compact set, $|z_u|$ has a finite upper bound $K(f, g, \mathbf{E}, B_0, u)$. By Proposition 4.6 the set U is non-void, finite, independent of $\mathbf{T}_1, \ldots, \mathbf{T}_k$ and is entirely determined by the matrix $\mathbf{E}$. Hence we can put

$$K(f, g, \mathbf{E}, B_0) = \max_{u\in U} K(f, g, \mathbf{E}, B_0, u) < \infty.$$

In particular, letting $\mathbf{R} = \mathbf{P}$ orthogonal, we have

$$|z_u(\mathbf{P}, \mathbf{T}_1, \ldots, \mathbf{T}_k)| \le K(f, g, \mathbf{E}, B_0) \quad \forall u \in U,$$

hence

$$|\varphi_{fg}| \le K(f, g, \mathbf{E}, B_0) \quad \text{for } 1 \le f \le \nu,\ 1 \le g \le m \text{ and } (\mathbf{T}_1, \ldots, \mathbf{T}_k) \in B_0.$$

Thus for all $(\mathbf{T}_1, \ldots, \mathbf{T}_k) \in B_0$,

$$\sum \varphi_{fg}^2 \le K(\mathbf{E}, B_0) = \sum_{f,g} K^2(f, g, \mathbf{E}, B_0), \quad f = 1, \ldots, \nu,\ g = 1, \ldots, m,$$

where the constant $K(\mathbf{E}, B_0)$ depends only on B_0 and $\mathbf{E}$.

In the case $k = 1$ we have $\mathbf{F} = \mathbf{\Lambda}$ and $\operatorname{Rank} \mathbf{\Lambda} = \nu = m$, hence

$$\sum \varphi_{fg}^2 = \|\mathbf{F}^{-1}\|^2 \quad (1 \le f, g \le m),$$

and from (4.2) we have

$$\|\mathbf{F}^{-1}\|^2 = \|\mathbf{\Lambda}^{-1}\|^2 = \operatorname{Tr}(\mathbf{\Lambda}^{-1})^2 = \operatorname{Tr}(\mathbf{M}_1^{-1})^2$$

since $\mathbf{\Lambda} = \mathbf{\Lambda}(1)$. Given $\mathbf{E} = \mathbf{E}_1$, $\operatorname{Tr}(\mathbf{M}_1^{-1})^2$ is a continuous function of $\mathbf{T}_1$ on $\mathcal{M}_{\mathrm{pd}}(n_1)$. Therefore $\sum_{1\le f,g\le m} \varphi_{fg}^2 = \operatorname{Tr}(\mathbf{M}_1^{-1})^2$ has a finite upper bound $K(\mathbf{E}, B_0)$ when $\mathbf{T}_1$ varies over a closed ball $B_0 \subset \mathcal{M}_{\mathrm{pd}}(n_1)$. ■

Due to the above important result we can assert the boundedness of each summand in (4.10).

Proposition 4.9. *Let* $\operatorname{Rank} \mathbf{E} = m \ge 1$ *and* $\operatorname{Rank} \mathbf{E}_j \ge 1$ *for some given* $j = 1, \ldots, k$. *Let* $(\mathbf{T}_1, \ldots, \mathbf{T}_k)$ *vary over a fixed closed ball* B_0 *in* $\prod_{i=1}^{k} \mathcal{M}_{\mathrm{pd}}(n_i)$. *Then* $\|\mathbf{F}^{-1}(j)\mathbf{P}'(j)\mathbf{E}_j'\|^2$ *is bounded by a positive number* K_j *depending on* $\mathbf{E}$ *and* B_0 *only.*

Proof. By renumbering the coordinate spaces $\mathbb{R}^{n_i}$, the corresponding matrices $\mathbf{T}_i$ and numbers $a(i)$, we can assume $j = 1$ without loss of generality. Then we are starting with the same conditions as in Proposition 4.8, so the same notations will be used.

First consider $m \ge 2$. Let us apply inequality (4.12). Let $\mathbf{E}$ be given and $(\mathbf{T}_1, \ldots, \mathbf{T}_k)$ vary over B_0. The sum $\sum \varphi_{fg}^2$, $1 \le f \le \nu$, $1 \le g \le m$, on the right-hand side of (4.12), by Proposition 4.8, is bounded by a positive constant $K = K(\mathbf{E}, B_0)$. Since $\mathbf{M}_1 = \mathbf{E}_1'\mathbf{T}_1\mathbf{E}_1$ by (4.1) and the function $d(\mathbf{M}_1)$ is continuous in $\mathbf{M}_1$ it follows that given $\mathbf{E}$ the factor $d(\mathbf{M}_1) \operatorname{Tr} \mathbf{T}_1^{-1}$ is continuous in $\mathbf{T}_1$. On the other hand

$$\operatorname{Rank} \mathbf{M}_1 = \operatorname{Rank} \mathbf{E}_1 = \nu \ge 1,$$

hence, given $\mathbf{E}$, there exists some fixed set $\sigma \subset \{1, \ldots, m\}$ with $\sharp\sigma = \nu$ such that $\operatorname{Rank}(\ldots \mathbf{e}_{f1} \ldots)_{f\in\sigma} = \nu$ and then

$$\mathbf{N}_1 = (\ldots \mathbf{e}_{f1} \ldots)'\mathbf{T}_1(\ldots \mathbf{e}_{f1} \ldots) \quad (f \in \sigma)$$

is a $\nu \times \nu$ p.d. principal submatrix of $\mathbf{M}_1$ for any p.d. matrix $\mathbf{T}_1$. Thus, the function $w_\nu(\mathbf{N}_1)$, being continuous in $\mathbf{N}_1$ on $\mathcal{M}_{\mathrm{pd}}(\nu)$, is continuous in $\mathbf{T}_1$ on $\mathcal{M}_{\mathrm{pd}}(n_1)$. Therefore $d(\mathbf{M}_1)w_\nu(\mathbf{N}_1)\operatorname{Tr}\mathbf{T}_1^{-1}$ is continuous in $\mathbf{T}_1$ on $\mathcal{M}_{\mathrm{pd}}(n_1)$, hence it is continuous in $(\mathbf{T}_1, \ldots, \mathbf{T}_k)$ on the product set $\prod_{i=1}^k \mathcal{M}_{\mathrm{pd}}(n_i)$, so on the compact subset B_0 it has a positive upper bound $K' = K'(\mathbf{E}, B_0)$. Put $K_1(\mathbf{E}, B_0) = KK'$. From (4.12) we have

$$\|\mathbf{F}^{-1}(1)\mathbf{P}'(1)\mathbf{E}_1'\|^2 \le K_1(\mathbf{E}, B_0) \quad \forall(\mathbf{T}_1, \ldots, \mathbf{T}_k) \in B_0.$$

In the case $m = 1$ we have $\nu = \operatorname{Rank}\mathbf{E}_1 = 1 = m$. Then (4.13) applies. Given $\mathbf{E}$, the positive function $c(\mathbf{M}_1)\operatorname{Tr}\mathbf{T}_1^{-1}$ is continuous in $\mathbf{T}_1$, hence continuous in $(\mathbf{T}_1, \ldots, \mathbf{T}_k)$, thus on B_0 it has a positive upper bound $K_1(\mathbf{E}, B_0)$, therefore the above inequality still holds true. ∎

The boundedness of our norm now follows.

Proposition 4.10. *If $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ is an o.n. system in $\mathbb{R}^n$ for the inner product $(\mathbf{u}, \mathbf{v}) = \mathbf{u}'\mathbf{v}$ and if $(\mathbf{T}_1, \ldots, \mathbf{T}_k)$ varies over a fixed closed ball B_0 in $\prod_{i=1}^k \mathcal{M}_{\mathrm{pd}}(n_i)$, then $\|\sum_{f,g=1}^m \mu_{fg}\mathbf{e}_f\mathbf{e}_g'\boldsymbol{\Delta}\|^2$ is bounded by a positive constant K_0 depending on $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ and B_0 only.*

Proof. On the right-hand side of (4.10), every summand with $\operatorname{Rank}\mathbf{E}_i \ge 1$ is, by Proposition 4.9, bounded by a positive constant $K_i = K_i(\mathbf{E}, B_0)$, whereas that with $\operatorname{Rank}\mathbf{E}_i = 0$ vanishes. If we put $K_0 = \sum K_i$ for those i with $\operatorname{Rank}\mathbf{E}_i \ge 1$, the assertion follows. ∎

We are in a position to prove Lemma 3.4. We let $m = \dim\Phi$.

Proof of (3.23). When $m = 0$, $\Phi = \{\mathbf{0}\}$ hence $\mathbf{J} = \mathbf{0}_{p\times p}$ and (3.23) is trivial. Thus we consider $m \ge 1$. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ be an arbitrary o.n. basis of Φ according to the inner product $(\mathbf{u}, \mathbf{v}) = \mathbf{u}'\mathbf{v}$ in $\mathbb{R}^n$. From (4.9) and Proposition 4.10 there is a positive constant $K_0(\mathbf{e}_1, \ldots, \mathbf{e}_m, B_0)$ such that

$$\|(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{J}\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\boldsymbol{\Delta}\|^2 \le K_0(\mathbf{e}_1, \ldots, \mathbf{e}_m, B_0)$$

if $(\mathbf{T}_1, \ldots, \mathbf{T}_k) \in B_0$. Consider the set $\{\Phi\}$ of all linear subspaces Φ of $\mathbb{R}^n$ with $\dim\Phi \ge 1$. For each Φ there is the corresponding set $O(\Phi)$ of all o.n. bases of the subspace Φ. By the axiom of choice we can construct on $\{\Phi\}$ a function $E(\Phi)$ whose value at Φ is some definitely chosen element $E(\Phi)$ in $O(\Phi)$. Taking $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\} = E(\Phi)$, put $D(\Phi, B_0) = K_0(E(\Phi), B_0)$. Then (3.23) holds. ∎

Before proving (3.24) note the following fact.

Let Φ and K be subspaces of a Euclidean vector space, of which Φ is finite-dimensional. Let Φ_K be the orthogonal projection of Φ on K. Then

$$\dim\Phi_K = \dim\Phi - \dim(\Phi \cap K^\perp). \tag{4.22}$$

Indeed, let A be the orthogonal projector from Φ onto K. Then $\operatorname{Im}A$ = image of $A = A\Phi = \Phi_K$ and

$$\ker A = \{\mathbf{u} \in \Phi : A\mathbf{u} = \mathbf{0}\} = \Phi \cap K^\perp.$$

By the dimension formula (see Chambadal and Ovaert (1968), p. 54, formula (2))

$$\dim \operatorname{Im} A + \dim \ker A = \dim \Phi,$$

we get the desired equality.

Proof of (3.24). There are m linearly independent vectors $\mathbf{e}_{1i}, \ldots, \mathbf{e}_{mi}$ in every $\mathbb{R}^{n_i}$, for $1 \le m \le \min n_i$. According to Symbols 4.1, put

$$\mathbf{E}_i = (\mathbf{e}_{1i} \ldots \mathbf{e}_{mi}), \quad \mathbf{E} = (\mathbf{e}_1 \ldots \mathbf{e}_m) = (\mathbf{E}_1' \ldots \mathbf{E}_k')'. \tag{4.23}$$

Then $\mathbf{e}_1, \ldots, \mathbf{e}_m$ are linearly independent vectors in $\mathbb{R}^n = \prod_{i=1}^k \mathbb{R}^{n_i}$. Put $\Phi = \operatorname{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$. Let R_i be the set of $n \times 1$ vectors $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$ lying in the coordinate space $\mathbb{R}^{n_i}$ of $\mathbb{R}^n$, i.e. $\mathbf{x}_h \in \mathbb{R}^{n_h}$, $\mathbf{x}_h = \mathbf{0}$ for all $h \ne i$. Then by (4.22),

$$\dim(\Phi \cap \mathbb{R}_i^\perp) = \dim \Phi - \dim \operatorname{Span}\{\mathbf{e}_{1i}, \ldots, \mathbf{e}_{mi}\} = 0.$$

Thus the class

$$\Gamma = \{\Phi \subset \mathbb{R}^n : 1 \le \dim \Phi \le \min n_i,\ \Phi \cap R_i^\perp = \{\mathbf{0}\}\ \forall i\}$$

is non-void. In each subspace $\Phi \in \Gamma$ now choose any o.n. basis which is still denoted by $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, and the representation (4.23) remains in use; then according to Lemma 2.1 consider

$$\sigma_i = 1 - \det{(\mathbf{e}_{1i} \ldots \mathbf{e}_{mi})'}(\mathbf{e}_{1i} \ldots \mathbf{e}_{mi}).$$

It follows that $\Phi \cap R_i^\perp = \{\mathbf{0}\}$ means

$$\sigma_i < 1 \quad \text{or, equivalently,} \quad \operatorname{Rank} \mathbf{E}_i = m. \tag{4.24}$$

Hence

$$\Gamma = \{\Phi : \Phi \text{ has an o.n. basis } \{\mathbf{e}_1, \ldots, \mathbf{e}_m\},\ 1 \le m \le \min n_i,\ \sigma_i < 1, \forall i\},$$

so using the symbol

$$\Gamma_\delta = \{\Phi : 1 \le \dim \Phi \le \min n_i,\ \sigma_i \le \delta_i,\ \forall i\}$$

we have $\Gamma = \bigcup \Gamma_\delta$, where the union extends over all $\delta = (\delta_1, \ldots, \delta_k)$, $0 \le \delta_i < 1$ for all i. By Definition 2.2, Γ_δ is just the class of subspaces Φ at most δ_i-steep relative to every coordinate space $\mathbb{R}^{n_i}$. Since Γ is non-void there is $\delta_0 = (\delta_{01}, \ldots, \delta_{0k})$, $0 \le \delta_{0i} < 1$ for all i, such that Γ_{δ_0} is non-void. Then for every $\delta \ge \delta_0$ we have $\Gamma_\delta \supset \Gamma_{\delta_0}$, so Γ_δ is non-void.

From (4.9), (4.24) and Corollary 4.1, for every $\Phi \in \Gamma$ we get

$$\|(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{J}\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\boldsymbol{\Delta}\|^2 \le \sum_{i=1}^k c(\mathbf{M}_i) \operatorname{Tr} \mathbf{T}_i^{-1}, \tag{4.25}$$

where $c(\mathbf{M}_i) = c(\mathbf{E}_i'\mathbf{T}_i\mathbf{E}_i)$ (see (4.1)) is a function continuous in $\mathbf{M}_i$ on $\mathcal{M}_{\mathrm{pd}}(m)$. But $\mathbf{M}_i \in \mathcal{M}_{\mathrm{pd}}(m)$ whenever $\mathbf{T}_i \in \mathcal{M}_{\mathrm{pd}}(n_i)$ and $\operatorname{Rank} \mathbf{E}_i = m$. Therefore using (4.24) we see that $c(\mathbf{M}_i)$ is a continuous function of $(\mathbf{E}_i, \mathbf{T}_i)$ on the product space

$$\{\mathbf{E}_i : \sigma_i < 1\} \times \mathcal{M}_{\mathrm{pd}}(n_i).$$

Hence by (4.23), $c(\mathbf{M}_i)$ is a continuous function of $(\mathbf{E}, \mathbf{T}_1, \ldots, \mathbf{T}_k)$ on the product space

$$\{\mathbf{E} : \sigma_i < 1\ \forall i = 1, \ldots, k\} \times \prod_{i=1}^k \mathcal{M}_{\mathrm{pd}}(n_i) \tag{4.26}$$

and then so is the sum $\sum_{i=1}^k c(\mathbf{M}_i)\,\mathrm{Tr}\,\mathbf{T}_i^{-1}$. But the set

$$E_\delta = \{(\mathbf{e}_1 \dots \mathbf{e}_m) : \sigma_i \le \delta_i\ \forall i,\ \text{the vectors } \mathbf{e}_1, \dots, \mathbf{e}_m \text{ are o.n.}\}$$

is bounded and closed in $\mathcal{M}_{n\times m}$ since the orthonormality conditions are

$$\mathbf{e}_j'\mathbf{e}_j = 1 \quad \text{and} \quad \mathbf{e}_j'\mathbf{e}_p = 0 \quad \forall j = 1, \dots, m,\ p = j+1, \dots, m.$$

Thus E_δ is compact, and hence so is $E_\delta \times B_0$ as a product of compact spaces. But $E_\delta \times B_0$ is included in the set (4.26), hence the positive function $\sum_{i=1}^k c(\mathbf{M}_i)\,\mathrm{Tr}\,\mathbf{T}_i^{-1}$, being continuous on $E_\delta \times B_0$, attains on this compact set its supremum denoted by $\alpha_m(B_0, \delta)$ as this depends only on $m, \delta_1, \dots, \delta_k$ and B_0. Now consider an arbitrary subspace $\Phi \in \Gamma_\delta$ and let $m = \dim \Phi$. Then for any o.n. basis $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ of Φ we have $(\mathbf{e}_1 \dots \mathbf{e}_m) \in E_\delta$, hence the right-hand side of (4.25) does not exceed

$$\alpha(B_0, \delta) = \max\{\alpha_m(B_0, \delta) : m = 1, \dots, \min n_i\}$$

provided $(\mathbf{T}_1, \dots, \mathbf{T}_k) \in B_0$. Thus (3.24) is proved. ■

The proof of Lemma 3.4 is complete.

5. Positivity of sums of mixing determinants

In this section we shall prove the basic property stated in Proposition 4.5(ii) for sums of mixing determinants, the name stemming from Proposition 4.4. As a geometric application, Lemma 2.1 will then be proved and Definition 2.2 justified.

5.1. Proof of Proposition 4.5(ii). In the particular case $k = m$, $m_1 = \dots = m_k = 1$, Proposition 4.5(ii) follows from (i) and the following.

PROPOSITION 5.1. *Let* $\mathbf{A}_1, \dots, \mathbf{A}_m$ *be* $m \times m$ *n.n.d. matrices. If*

$$\mathrm{Rank}(\dots \mathbf{A}_i' \dots)_{i\in\psi} \ge \sharp\psi \quad \forall \psi \subset \{1, \dots, m\},\ \psi \ne \emptyset \tag{5.1}$$

then $\mathbf{A}_1(1) \circ \dots \circ \mathbf{A}_m(1) > 0$.

Proof. We always write

$$\mathbf{A}_1 \circ \dots \circ \mathbf{A}_m = \mathbf{A}_1(1) \circ \dots \circ \mathbf{A}_m(1)$$

for short. By assumption we have $\mathrm{Rank}\,\mathbf{A}_i \ge 1$, $i = 1, \dots, m$. Let $\mu = \mathrm{Rank}\,\mathbf{A}_1 \ge 1$. Consider the spectral decomposition

$$\mathbf{A}_1 = \mathbf{P}\mathbf{\Lambda}\mathbf{P}', \quad \mathbf{P} = (\mathbf{P}_1 \dots \mathbf{P}_m) \text{ orthogonal},$$
$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_m), \quad \lambda_1 \ge \dots \ge \lambda_\mu > 0.$$

Put $\mathbf{Q}_i = \mathbf{P}'\mathbf{A}_i\mathbf{P}$. For any numbers $t_1, \dots, t_m$ we then have

$$\det\Big(\sum_{i=1}^m t_i\mathbf{A}_i\Big) = \det\Big(\sum_{i=1}^m t_i\mathbf{Q}_i\Big).$$

Thus from (4.14) in the expansion of the right-hand side, $\mathbf{A}_1 \circ \dots \circ \mathbf{A}_m$ is also the coefficient of $t_1 \dots t_m$. For $m \ge 2$ from Proposition 4.4 this coefficient is the sum of $m!$ mixing determinants, each formed by deleting one row from each $\mathbf{Q}_i$, $i = 1, \dots, m$.

But $\mathbf{Q}_1 = \Lambda$, hence we only consider those mixing determinants formed with the jth row deleted from Λ, $j = 1, \ldots, \mu$, whereas the remaining rows are from $\mathbf{Q}_2, \ldots, \mathbf{Q}_k$. The cofactors of λ_j in such mixing determinants are formed by deleting one row from each matrix $\mathbf{Q}_2(\sigma_j), \ldots, \mathbf{Q}_k(\sigma_j)$, $\mathbf{Q}_i(\sigma_j)$ being, according to (1.7), the $(m-1) \times (m-1)$ principal submatrix of $\mathbf{Q}_i$ corresponding to $\sigma_j = \{1, \ldots, m\} - \{j\}$. Thus

$$\mathbf{A}_1 \circ \ldots \circ \mathbf{A}_m = \sum_{j=1}^{\mu} \lambda_j \mathbf{Q}_2(\sigma_j) \circ \ldots \circ \mathbf{Q}_m(\sigma_j). \tag{5.2}$$

In general we have

$$\mathbf{Q}_i(\sigma) = (\ldots \mathbf{P}_f \ldots)'_{f \in \sigma} \mathbf{A}_i (\ldots \mathbf{P}_f \ldots)_{f \in \sigma} \quad \text{for } \sigma \subset \{1, \ldots, m\}.$$

Since $\mathbf{Q}_i$ is n.n.d., so is $\mathbf{Q}_i(\sigma)$. Further let K_i denote $\ker \mathbf{A}_i$ and φ any subset of $\{2, \ldots, m\}$. Thus

$$\ker(\ldots \mathbf{A}'_i \ldots)'_{i \in \varphi} = \bigcap_{i \in \varphi} K_i.$$

Then from (3.38) and (3.37) we have

$$\begin{aligned}
\dim \ker(\ldots \mathbf{Q}_i(\sigma)' \ldots)'_{i \in \varphi} &= \dim \bigcap_{i \in \varphi} \ker \mathbf{Q}_i(\sigma) = \dim \bigcap_{i \in \varphi} \ker \mathbf{A}_i (\ldots \mathbf{P}_f \ldots)_{f \in \sigma} \\
&= \dim \ker(\ldots \mathbf{A}'_i \ldots)'_{i \in \varphi} (\ldots \mathbf{P}_f \ldots)_{f \in \sigma} \\
&= \dim \Big(\bigcap_{i \in \varphi} K_i \Big) \cap \operatorname{Span}\{\mathbf{P}_f : f \in \sigma\}.
\end{aligned}$$

But G and H being arbitrary finite-dimensional subspaces of some linear space, we have

$$\dim \operatorname{Span}(G \cup H) = \dim G + \dim H - \dim(G \cap H) \tag{5.3}$$

(cf. Chambadal and Ovaert (1968), p. 453, formula (1)), in particular $\dim(G \cap H) = \dim G + \dim H - m$ if $\dim \operatorname{Span}(G \cup H) = m$. Hence, the index f ranging over $1, \ldots, m$, we get

$$\dim \ker \begin{pmatrix} \vdots \\ \mathbf{Q}_i(\sigma_j) \\ \vdots \end{pmatrix}_{i \in \varphi} = \begin{cases} \dim \bigcap_{i \in \varphi} K_i & \text{if } \bigcap_{i \in \varphi} K_i \subset \operatorname{Span}\{\mathbf{P}_f : f \neq j\}, \\ \dim \bigcap_{i \in \varphi} K_i - 1 & \text{otherwise.} \end{cases} \tag{5.4}$$

Let us reason by induction. Using (3.36) we make the following assumption, equivalent to (5.1):

Assumption. $\mathbf{A}_1, \ldots, \mathbf{A}_m$ are $m \times m$ n.n.d. matrices satisfying the condition

$$\dim \bigcap_{i \in \psi} K_i \leq m - \sharp\psi \quad \forall \psi \subset \{1, \ldots, m\},\ \psi \neq \emptyset. \tag{5.5}$$

Let $m \geq 2$. We adopt the following

Induction hypothesis. For arbitrary $(m-1) \times (m-1)$ n.n.d. matrices $\mathbf{B}_1, \ldots, \mathbf{B}_{m-1}$ we have $\mathbf{B}_1 \circ \ldots \circ \mathbf{B}_{m-1} > 0$ if

$$\dim \ker(\ldots \mathbf{B}'_i \ldots)'_{i \in \psi} \leq m - 1 - \sharp\psi \quad \text{for every non-void set } \psi \subset \{1, \ldots, m-1\}.$$

On account of Proposition 4.5(i) and formula (5.2), this hypothesis has three consequences:

(5.6) if $\mathbf{B}_1, \ldots, \mathbf{B}_{m-1}$ are $(m-1) \times (m-1)$ n.n.d. matrices then $\mathbf{B}_1 \circ \ldots \circ \mathbf{B}_{m-1} \geq 0$,

(5.7) $\mathbf{B}_1 \circ \ldots \circ \mathbf{B}_{m-1} = 0$ if and only if

$$(\exists \psi \subset \{1, \ldots, m-1\},\ \psi \neq \emptyset) \quad \dim \ker(\ldots \mathbf{B}_i' \ldots)'_{i \in \psi} \geq m - \sharp\psi,$$

(5.8) $\mathbf{A}_1 \circ \ldots \circ \mathbf{A}_m \geq 0$.

Under Assumption (5.5) we intend to prove $\mathbf{A}_1 \circ \ldots \circ \mathbf{A}_m > 0$. In view of (5.8) we start from the converse

(5.9) *Supposition*: $\mathbf{A}_1 \circ \ldots \circ \mathbf{A}_m = 0$,

which by (5.2) and (5.6) is equivalent to

$$\mathbf{Q}_2(\sigma_j) \circ \ldots \circ \mathbf{Q}_m(\sigma_j) = 0 \quad \text{for } j = 1, \ldots, \mu,$$

and again, by (5.7), equivalent to

$$(\forall j = 1, \ldots, \mu)(\exists \varphi \subset \{2, \ldots, m\},\ \varphi \neq \emptyset) \quad \dim \ker(\ldots \mathbf{Q}_i(\sigma_j)' \ldots)'_{i \in \varphi} \geq m - \sharp\varphi.$$

Now, from (5.4), if $\bigcap_{i \in \varphi} K_i$ were not included in $\mathrm{Span}\{\mathbf{P}_f : f \neq j\}$, from Assumption (5.5) we would have

$$\dim \ker(\ldots \mathbf{Q}_i(\sigma_j)' \ldots)'_{i \in \varphi} = \dim \bigcap_{i \in \varphi} K_i - 1 \leq m - \sharp\varphi - 1.$$

Therefore, on account of Assumption (5.5), Supposition (5.9) is equivalent to

(5.10) $(\forall j = 1, \ldots, \mu)(\exists \varphi \subset \{2, \ldots, m\},\ \varphi \neq \emptyset)$

$$\bigcap_{i \in \varphi} K_i \subset \mathrm{Span}\{\mathbf{P}_f : f \neq j\} \quad \text{and} \quad \dim \bigcap_{i \in \varphi} K_i = m - \sharp\varphi.$$

For short reference we shall say that a subset φ of $\{2, \ldots, m\}$ *suits* some integer $j,\ 1 \leq j \leq \mu$, if simultaneously

(5.11) $$\varphi \neq \emptyset, \quad \bigcap_{i \in \varphi} K_i \subset \mathrm{Span}\{\mathbf{P}_f : f \neq j\}, \quad \dim \bigcap_{i \in \varphi} K_i = m - \sharp\varphi.$$

Then note the following properties.

(a) *If φ suits j and φ' suits $j' \neq j$ then $\varphi \cup \varphi'$ suits both j and j'.*

Indeed, by Assumption (5.5) the subspaces $K_i,\ i = 2, \ldots, m$, of $\mathbb{R}^m$ satisfy the condition

$$\dim \bigcap_{i \in \varphi} K_i \leq m - \sharp\varphi \quad \forall \varphi \subset \{2, \ldots, m\}$$

with the agreement that $\bigcap_{i \in \emptyset} K_i = \mathbb{R}^m$. Then we have

$$\begin{aligned} m - \sharp\varphi \cap \varphi' &\geq \dim \bigcap_{i \in \varphi \cap \varphi'} K_i \geq \dim \mathrm{Span}\Big\{\Big(\bigcap_{i \in \varphi} K_i\Big) \cup \Big(\bigcap_{i \in \varphi'} K_i\Big)\Big\} \\ &= \dim \bigcap_{i \in \varphi} K_i + \dim \bigcap_{i \in \varphi'} K_i - \dim \bigcap_{i \in \varphi \cup \varphi'} K_i \quad \text{by (5.3)} \\ &\geq (m - \sharp\varphi) + (m - \sharp\varphi') - (m - \sharp\varphi \cup \varphi') = m - \sharp\varphi \cap \varphi', \end{aligned}$$

since $\sharp\varphi\cup\varphi' = \sharp\varphi+\sharp\varphi'-\sharp\varphi\cap\varphi'$. Thus from the fourth inequality it follows $\dim\bigcap_{i\in\varphi\cup\varphi'} K_i = m - \sharp\varphi\cup\varphi'$. Moreover, both $\bigcap_{i\in\varphi} K_i$ and $\bigcap_{i\in\varphi'} K_i$ contain $\bigcap_{i\in\varphi\cup\varphi'} K_i$ hence so do $\mathrm{Span}\{\mathbf{P}_f : f \neq j\}$ and $\mathrm{Span}\{\mathbf{P}_f : f \neq j'\}$. Thus (a) follows.

(b) *If there exists a set φ that suits a certain $j \le \mu = \mathrm{Rank}\,\mathbf{A}_1$, then necessarily $\mu \ge 2$ and $\bigcap_{i\in\varphi} K_i$ is not included in K_1.*

Indeed, $\dim\bigcap_{i\in\varphi} K_i = m - \sharp\varphi$ whereas $\dim K_1 \cap \bigcap_{i\in\varphi} K_i \le m - \sharp\varphi - 1$ by Assumption (5.5). Hence $\bigcap_{i\in\varphi} K_i$ cannot be included in K_1. On the other hand, from the spectral decomposition of $\mathbf{A}_1$, noting that

$$K_1 = \ker\mathbf{A}_1 = \{\mathbf{P}_1,\ldots,\mathbf{P}_\mu\}^\perp = \bigcap_{k=1}^{\mu}\{\mathbf{P}_k\}^\perp = \bigcap_{k=1}^{\mu}\mathrm{Span}\{\mathbf{P}_f : f \neq k\}, \tag{5.12}$$

we have $K_1 \subset \mathrm{Span}\{\mathbf{P}_f : f \neq j\}$ since $j \le \mu$. Since $\mathrm{Rank}\,\mathbf{A}_1 \ge 1$ we have $\dim K_1 \le m-1$. If $\dim K_1 = m - 1$, we would have

$$K_1 = \mathrm{Span}\{\mathbf{P}_f : f \neq j\} \supset \bigcap_{i\in\varphi} K_i \quad \text{since } \varphi \text{ suits } j.$$

Thus $\dim K_1 < m - 1$, i.e. $\mu = \mathrm{Rank}\,\mathbf{A}_1 \ge 2$.

(c) *If φ suits j then there exists another j' such that no set φ', if any, suiting j' is included in φ.*

Indeed, from (b), $\bigcap_{i\in\varphi} K_i$ is not included in K_1, also $\mu \ge 2$. Since, by (5.11), $\bigcap_{i\in\varphi} K_i \subset \mathrm{Span}\{\mathbf{P}_f : f \neq j\}$ there exists from (5.12) some $j' \neq j,\ 1 \le j' \le \mu$, such that $\bigcap_{i\in\varphi} K_i$ is not included in $\mathrm{Span}\{\mathbf{P}_f : f \neq j'\}$. Then by (5.11) no $\varphi' \subset \varphi$ suits j' since $\bigcap_{i\in\varphi'} K_i \supset \bigcap_{i\in\varphi} K_i$. Thus (c) follows.

Now, by (5.10) and (5.11) Supposition (5.9) is equivalent to

$$(\forall j = 1,\ldots,\mu)(\exists\varphi \subset \{2,\ldots,m\}) \quad \varphi \text{ suits } j. \tag{5.13}$$

For $\mu = 1$, from (b), (5.13) is impossible. Let (5.13) be true for some $\mu \ge 2$. Consider some $j \le \mu$ and let φ be a set of maximal cardinality that suits j. By (5.13) for any $j' \le \mu$ there exists φ' suiting j' and from (c) there exists $j' \neq j$ such that $\sharp\varphi'\cup\varphi > \sharp\varphi$; from (a), $\varphi'\cup\varphi$ suits j, which contradicts the maximality of $\sharp\varphi$. Therefore (5.13) is always impossible. Thus the induction hypothesis and Assumption (5.5) entail that $\mathbf{A}_1 \circ \ldots \circ \mathbf{A}_m > 0$. On the other hand for $m = 2$ the induction hypothesis is trivially true: indeed, for $m-1 = 1$, the n.n.d. matrix $\mathbf{B}_1$ is a non-negative number and obviously $\mathbf{B}_1 \circ \ldots \circ \mathbf{B}_{m-1} = \mathbf{B}_1$ is positive if $\dim\ker\mathbf{B}_1 \le m - 1 - 1 = 0$. So Proposition 5.1 is proved.

Let us now prove Proposition 4.5(ii) in the general case.

From Proposition 4.4, the sum $\mathbf{A}_1(m_1) \circ \ldots \circ \mathbf{A}_k(m_k)$ does not involve those $\mathbf{A}_i$ with $m_i = 0$. Also, if (4.15) is satisfied by every set φ containing no elements i with $m_i = 0$, then it is satisfied by an arbitrary set φ. Thus, without loss of generality, we shall assume that $m_1,\ldots,m_k$ are positive. Put $l_0 = 0,\ldots,l_i = m_1+\ldots+m_i,\ldots,l_k = m$. Consider the partition $\{1,\ldots,m\} = \bigcup_{i=1}^{k}\Delta_i$, where $\Delta_i = \{l_{i-1}+1,\ldots,l_i\}$. To every set φ in $\{1,\ldots,k\}$ there corresponds a subset, called the *Δ-set*, of $\{1,\ldots,m\}$, of the form $\Delta = \bigcup_{i\in\varphi}\Delta_i$, where $\Delta = \emptyset$ if and only if $\varphi = \emptyset$. Since an intersection of Δ-sets is again a Δ-set, to every set ψ in $\{1,\ldots,m\}$ there corresponds a minimal Δ-set containing ψ. From $\mathbf{A}_1,\ldots,\mathbf{A}_k$ let us generate n.n.d. matrices $\mathbf{B}_1, \ldots, \mathbf{B}_m$ by putting $\mathbf{B}_f = \mathbf{A}_i$ for all $f \in \Delta_i,\ i = 1, \ldots, k$. Then for any set $\psi \subset \{1, \ldots, m\}$ and for the minimal set $\Delta = \bigcup_{i \in \varphi} \Delta_i$ containing ψ we have

$$\mathrm{Rank}(\ldots \mathbf{A}_i' \ldots)_{i \in \varphi} = \mathrm{Rank}(\ldots \mathbf{B}_f' \ldots)_{f \in \Delta} = \mathrm{Rank}(\ldots \mathbf{B}_f' \ldots)_{f \in \psi}.$$

Note that $\sum_{i \in \varphi} m_i = \sharp \Delta$. Then in particular taking $\psi = \Delta$, we see that the condition

$$(\forall \psi \subset \{1, \ldots .m\},\ \psi \neq \emptyset) \quad \mathrm{Rank}(\ldots \mathbf{B}_f' \ldots)_{f \in \psi} \geq \sharp \psi \tag{5.14}$$

entails

$$\mathrm{Rank}(\ldots \mathbf{A}_i' \ldots)_{i \in \varphi} \geq \sum_{i \in \varphi} m_i \quad \forall \varphi \subset \{1, \ldots, k\},\ \varphi \neq \emptyset,$$

which is (4.15). Conversely, (4.15) entails (5.14) since $\sharp \Delta \geq \sharp \psi$. Thus (4.15) is equivalent to (5.14).

On the other hand, from the proof of Proposition 4.4, $\mathbf{A}_1(m_1) \circ \ldots \circ \mathbf{A}_k(m_k)$ is a sum of determinants, each corresponding to a partition of the set $\{1, \ldots, m\}$ into k parts D_i $(i = 1, \ldots, k)$ with $\sharp D_i = m_i$, and the jth row of a summand determinant is just the jth row of $\mathbf{A}_i$ when $j \in D_i$; in particular each permutation of the set $\{1, \ldots, m\}$, considered as a partition of this set into m parts, gives a summand determinant of $\mathbf{B}_1(1) \circ \ldots \circ \mathbf{B}_m(1)$, the jth row of which is the jth row of $\mathbf{B}_f$ when j is the fth element of the permutation considered. To any permutation of $\{1, \ldots, m\}$ there corresponds a partition $\{D_1, \ldots, D_k\}$ with $\sharp D_i = m_i$ as follows: put $D_i = \{j\}$, where j is the fth element of the permutation with $f \in \Delta_i$; but $f \in \Delta_i$ entails $\mathbf{B}_f = \mathbf{A}_i$, hence the $m_1! \ldots m_k!$ permutations generated by the partition $\{D_1, \ldots, D_k\}$ give the corresponding summand determinants of $\mathbf{B}_1(1) \circ \ldots \circ \mathbf{B}_m(1)$ all equal to the summand determinant of $\mathbf{A}_1(m_1) \circ \ldots \circ \mathbf{A}_k(m_k)$ that corresponds to the same partition. Thus we get

$$m_1! \ldots m_k! \mathbf{A}_1(m_1) \circ \ldots \circ \mathbf{A}_k(m_k) = \mathbf{B}_1 \circ \ldots \circ \mathbf{B}_m.$$

Therefore, from Propositions 4.5(i) and 5.1, $\mathbf{A}_1(m_1) \circ \ldots \circ \mathbf{A}_k(m_k)$ is non-negative and its positiveness is equivalent to (5.14), i.e. to (4.15). The proof of Proposition 4.5(ii) is complete.

5.2. A multidimensional geometric characteristic. Proposition 4.5(ii) enables us to establish a geometric fact in multidimensional vector spaces, namely Lemma 2.1.

Proof of Lemma 2.1. Let $\{\mathbf{g}_1, \ldots, \mathbf{g}_m\} \subset \mathbb{R}^n$ be another basis of Φ. If $\mathbf{A}$ is the $m \times m$ matrix of change of basis then $(\mathbf{g}_1 \ldots \mathbf{g}_m) = (\mathbf{f}_1 \ldots \mathbf{f}_m)\mathbf{A}$. Consider

$$\det(\mathbf{g}_1 \ldots \mathbf{g}_m)'(\mathbf{g}_1 \ldots \mathbf{g}_m) = (\det \mathbf{A})^2 \det(\mathbf{f}_1 \ldots \mathbf{f}_m)'(\mathbf{f}_1 \ldots \mathbf{f}_m);$$

projecting on K we have a similar equality for $\det(\mathbf{g}_{1K} \ldots \mathbf{g}_{mK})'(\mathbf{g}_{1K} \ldots \mathbf{g}_{mK})$. But $\det \mathbf{A} \neq 0$ since $\mathrm{Rank}(\mathbf{g}_1 \ldots \mathbf{g}_m) = m$. Then we see that σ is independent of the choice of the basis.

Write $\mathbf{f}_i = \mathbf{f}_{iK} + \mathbf{t}_{iK},\ \mathbf{f}_{iK} \in K,\ \mathbf{t}_{iK} \perp K$, $i = 1, \ldots, m$. Then

$$(\mathbf{f}_1 \ldots \mathbf{f}_m)'(\mathbf{f}_1 \ldots \mathbf{f}_m) = (\mathbf{f}_i'\mathbf{f}_j) = (\mathbf{f}_{iK}'\mathbf{f}_{jK} + \mathbf{t}_{iK}'\mathbf{t}_{jK}), \quad i, j = 1, \ldots, m.$$

Consider the n.n.d. matrices $\mathbf{B} = (\mathbf{f}_{iK}'\mathbf{f}_{jK})$ and $\mathbf{C} = (\mathbf{t}_{iK}'\mathbf{t}_{jK})$. From (4.14) we have

$$\det(\mathbf{f}_1 \ldots \mathbf{f}_m)'(\mathbf{f}_1 \ldots \mathbf{f}_m) = \det(\mathbf{B} + \mathbf{C}) = \det \mathbf{B} + \sum \mathbf{B}(r) \circ \mathbf{C}(s),$$

$r = 0, \dots, m-1$, $r+s=m$. By Proposition 4.5(ii), $\mathbf{B}(r) \circ \mathbf{C}(s) \geq 0$, hence $0 \leq \sigma \leq 1$; moreover, when $\det \mathbf{B} > 0$ then $\mathbf{B}(r) \circ \mathbf{C}(s) = 0$ if and only if $\operatorname{Rank} \mathbf{C} \leq s-1$, since we have neither $\operatorname{Rank} \mathbf{B} \leq r-1$ nor $\operatorname{Rank}(\mathbf{B} \vdots \mathbf{C}) \leq r+s-1$. Thus $\det (\mathbf{f}_1 \dots \mathbf{f}_m)'(\mathbf{f}_1 \dots \mathbf{f}_m) = \det \mathbf{B} > 0$ if and only if $\operatorname{Rank} \mathbf{C} \leq s-1$ for $s = 1, \dots, m$, i.e. if and only if $\mathbf{C} = \mathbf{0}$ or, equivalently, $\mathbf{t}_{iK} = \mathbf{0}$ for $i = 1, \dots, m$. Therefore $\sigma = 0$ if and only if

$$\mathbf{f}_i = \mathbf{f}_{iK} \ \forall i = 1, \dots, m \quad \text{or, equivalently,} \quad \Phi \subset K.$$

On the other hand, $\sigma = 1$ if and only if

$$\det(\mathbf{f}_{1K} \dots \mathbf{f}_{mK})'(\mathbf{f}_{1K} \dots \mathbf{f}_{mK}) = 0, \quad \text{i.e.} \quad \operatorname{Rank}(\mathbf{f}_{1K} \dots \mathbf{f}_{mK}) < m,$$

which means $\dim \Phi_K < m$, where Φ_K is the orthogonal projection of Φ on K. By (4.22) we see that $\sigma = 1$ if and only if $\dim(\Phi \cap K^\perp) > 0$. ∎

Since $K \subset \mathbb{R}^n$ we have $\dim K = n - \dim K^\perp$, which by (5.3) entails

$$\dim(\Phi \cap K^\perp) = \dim \Phi + (n - \dim K) - \dim \operatorname{Span}(\Phi \cup K^\perp) \geq \dim \Phi - \dim K,$$

for $\Phi \cup K^\perp \subset \mathbb{R}^n$. Hence if $\dim \Phi > \dim K$ then, from the above, always $\sigma = 1$, whereas if $\dim \Phi \leq \dim K$, σ can in fact vary from zero to one, and then σ can be used as a *measure of steepness* of the subspace Φ with respect to the subspace K in $\mathbb{R}^n$, which justifies Definition 2.2.

Acknowledgments. This work was supported by the Natural Science Council of Viet Nam.

The author wishes to thank the referee of the earlier draft of the paper for valuable comments which have led to important improvements in the presentation.

References

T. W. Anderson and J. Taylor (1979), *Strong consistency of least squares estimators in dynamic models*, Ann. Statist. 7, 484–489.

N. Bac-Van (1992), *Estimation and prediction in regression models with random explanatory variables*, Dissertationes Math., 317.

N. Bac-Van (1994), *Strong consistency of least squares estimates in polygonal regression with random explanatory variables*, Acta Math. Vietnam. 19 (1), 111–137.

N. Bac-Van (1998), *Strong consistency of least squares estimates and mean square regression*, ibid. 23 (2), 195–205.

B. R. Bhat (1982), *Strong consistency of least squares estimates and conditional consistency*, in: Statistics and Probability (essays in honor of C. R. Rao), North-Holland, Amsterdam, 105–108.

H. Cartan (1967), *Calcul différentiel*, Hermann, Paris.

L. Chambadal and J. L. Ovaert (1968), *Algèbre linéaire et algèbre tensorielle*, Dunod, Paris.

D. S. Chang and M. R. Chang (1996), *A note on hypothesis testing in stochastic regression models*, Statist. Sinica 6, 491–498.

D. S. Chang and G. C. Lin (1995), *Stochastic regression models with heteroscedastic disturbance*, Ann. Inst. Statist. Math. 47, 351–369.

G. J. Chen, T. L. Lai and C. Z. Wei (1981), *Convergence systems and strong consistency of least squares estimates in regression models*, J. Multivariate Anal. 11, 319–333.

G. J. Chen, T. L. Lai and C. Z. Wei (1983), *Convergence systems and strong consistency of least squares estimates in regression models*, Kexue Tongbao (Science Bulletin) (Chinese Academy of Science) 28 (1), 16–20.

N. Christopeit and K. Helmes (1980), *Strong consistency of least squares estimators in linear regression models*, Ann. Statist. 8, 778–788.

H. Cramér (1945), *Mathematical Methods of Statistics*, Almqvist and Wiksells, Uppsala.

N. Dunford and J. T. Schwartz (1958), *Linear Operators*, *Part I*: *General Theory*, Interscience Publ., London.

D. A. Freedman (1981), *Bootstrapping regression models*, Ann. Statist. 9, 1218–1228.

K. M. S. Humak (1977), *Statistische Methoden der Modellbildung*, Band I, Akademie-Verlag, Berlin.

M. Z. Jin and X. R. Chen (1996), *Necessary and sufficient conditions for strong consistency of least squares estimates of linear regression coefficients*, Chinese Ann. Math. Ser. A 17, 123–130.

T. L. Lai, H. Robbins and C. Z. Wei (1979), *Strong consistency of least squares estimates in multiple regression II*, J. Multivariate Anal. 9, 343–361.

T. L. Lai and C. Z. Wei (1982), *Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems*, Ann. Statist. 10, 154–166.

M. Loève (1963), *Probability Theory*, 3rd ed., Van Nostrand, New York.

C. R. Rao (1973), *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

C. Z. Wei (1985), *Asymptotic properties of least squares estimates in stochastic regression models*, Ann. Statist. 13, 1498–1508.

T. J. Wu and M. T. Wasan (1996), *Weighted least squares estimates in linear regression models for processes with uncorrelated increments*, Stochastic Process. Appl. 64, 273–286.