Теорія Ймовір. та Матем. Статист.
Вип. 83, 2010

Theor. Probability and Math. Statist.
No. 83, 2011, Pages 175–190
S 0094-9000(2012)00850-8
Article electronically published on February 2, 2012

CONDITIONS FOR THE CONSISTENCY OF THE TOTAL LEAST SQUARES ESTIMATOR IN AN ERRORS-IN-VARIABLES LINEAR REGRESSION MODEL

UDC 519.21

S. V. SHKLYAR

Abstract. A homoscedastic errors-in-variables linear regression model is considered. The total least squares estimator is studied. New conditions for the consistency and strong consistency of the total least squares estimator are proposed. These conditions are weaker than those proposed by Kukush and Van Huffel (Metrika **59** (2004), 75–97).

1. Introduction

Consider a linear vector regression errors-in-variables model

$$\begin{cases} b_i = a_i^0 X_0 + \tilde{b}_i, \\ a_i = a_i^0 + \tilde{a}_i. \end{cases} \tag{1}$$

Here $a_i^0 \in \mathbb{R}^{1\times n}$ are unknown nonrandom vectors, $X_0 \in \mathbb{R}^{n\times d}$ is a matrix to be estimated, and $\tilde{b}_i$ and $\tilde{a}_i$ are vectors of random errors, $i = 1, \ldots, m$.

This model can be rewritten in the matrix form as follows:

$$A_0 X_0 = B_0, \qquad A = A_0 + \tilde{A}, \qquad B = B_0 + \widetilde{B}, \tag{2}$$

where A_0, B_0, $\tilde{A}$, and $\tilde{B}$ are the matrices constituted from the rows a_i^0, b_i^0, $\tilde{a}_i$, and $\tilde{b}_i$, respectively.

In fact, (1) is a functional model, since the vectors a_i^0 are nonrandom. Model (2) is used in one of the approaches to solve overdetermined systems of linear equations $AX \approx B$. A widely used estimator of the parameter X_0 for such a model is the so-called *total least squares estimator*.

Sufficient conditions for the consistency of the total least squares estimators are given in the papers [3, 4, 5, 7] under various assumptions concerning the model of observations. The so-called structured total least squares estimator is studied in [8]. The construction of this estimator is based on an assumption that the true matrices as well as matrices of observations have a specific structure. For example, the construction in [8] is suitable for Toeplitz or Hankel matrices having the block structure.

New conditions for the consistency of the total least squares estimator are given in the current paper. These conditions are weaker than those given in [7].

The paper is organized as follows. Section 2 describes the model of observations and defines the total least squares estimator. In Section 3, we recall known conditions for the

---

2010 *Mathematics Subject Classification.* Primary 62H12, 62J05; Secondary 62F10, 62F12, 62J10.

Key words and phrases. Linear regression, errors in variables, total least squares estimator, orthogonal regression.

consistency of this estimator. In Section 4, we prove theorems concerning the consistency under weaker conditions. We recall some known results needed for the proof in Section 5.

2. The model and estimator

2.1. **Model.** We assume that the true nonrandom matrices A_0 and B_0 are such that

$$\underset{m\times n}{A_0} \cdot \underset{n\times d}{X_0} = \underset{m\times d}{B_0}.$$

We further assume that A_0 and B_0 are observed with random errors, $\tilde{A}$ and $\widetilde{B}$, respectively; that is, we observe, in fact, the matrices A and B given by

$$A = A_0 + \tilde{A}, \qquad B = B_0 + \widetilde{B}.$$

Our aim is to estimate the parameter X_0 from the observations.

Now we rewrite the model in an implicit form. Let $C_0 \in \mathbb{R}^{m\times(n+d)}$, $\widetilde{C} \in \mathbb{R}^{m\times(n+d)}$, and $C \in \mathbb{R}^{m\times(n+d)}$ be the $m \times (n+d)$ matrices such that

$$C_0 = [A_0\ B_0], \qquad \widetilde{C} = [\tilde{A}\ \widetilde{B}], \qquad C = [A\ B].$$

Let $X_{\text{ext}}^0 = \left(\begin{smallmatrix} X_0 \\ -I_d \end{smallmatrix}\right)$. Then

$$\underset{m\times(n+d)}{C_0} \cdot \underset{(n+d)\times d}{X_{\text{ext}}^0} = \underset{m\times d}{0}.$$

The entries of the matrix $\widetilde{C}$ (the errors of observations) are denoted by δ_{ij} and its rows are denoted by $\tilde{c}_i$; namely, $\widetilde{C} = (\delta_{ij})_{i=1}^{m}{}_{j=1}^{n+d}$ and $\tilde{c}_i = (\delta_{ij})_{j=1}^{n+d}$.

We assume the following conditions.

(G.1) The rows $\tilde{c}_i$ of the matrix $\widetilde{C}$ are jointly independent;

(G.2) $\mathsf{E}\,\widetilde{C} = 0$ and, for all $i = 1, \dots, m$, the covariance matrix of the vector $\tilde{c}_i$ is equal to $\operatorname{Var} \tilde{c}_i = \Sigma$;

(G.3) $\operatorname{rank}(\Sigma X_{\text{ext}}^0) = d$.

Example 1 (univariate scalar regression)**.** For $i = 1, \dots, m$, let

$$\begin{cases} x_i = \xi_i + \delta_i; \\ y_i = \beta_0 + \beta_1 \xi_i + \epsilon_i. \end{cases}$$

The sequence $\{(x_i, y_i), y = 1, \dots, m\}$ is observed. One needs to estimate the parameters β_0 and β_1 from the observations.

2.2. **Total least squares estimator.** This estimator is defined as a solution of the problem

$$\begin{cases} \|\operatorname{pinv}(\Sigma^{1/2})\,\Delta^\top\|_F \to \min; \\ (I - P_\Sigma)\Delta^\top = 0; \\ \operatorname{rank}(C - \Delta) \le n. \end{cases} \tag{3}$$

Here $\operatorname{pinv}(\Sigma)$ denotes the pseudoinverse matrix to Σ and P_Σ is the orthogonal projector to the column space Σ, so that $P_\Sigma = \Sigma \operatorname{pinv}(\Sigma)$. The estimator is evaluated from the equations

$$(C - \Delta)\hat{X}_{\text{ext}} = 0, \tag{4}$$

$$(C - \Delta)\begin{pmatrix} \hat{X} \\ -I \end{pmatrix} = 0. \tag{5}$$

The columns of $\hat{X}_{\mathrm{ext}}$ have to constitute a basis of the invariant subspace of $C^\top C$ with respect to Σ that corresponds to d minimal generalized eigenvalues,

$$\exists M \in \mathbb{R}^{d\times d}: \quad C^\top C \hat{X}_{\mathrm{ext}} = \Sigma \hat{X}_{\mathrm{ext}} M.$$

Two real symmetric matrices A and B are called a definite pair if there exist two real numbers α and β such that $\alpha A + \beta B$ is a positive definite matrix. We write $\langle A, B\rangle$ in this case.

Below is a list of possible problems that may arise when solving the minimization problem (3).

a) The pair of matrices $\langle C^\top C, \Sigma\rangle$ is degenerate (the matrices $C^\top C$ and Σ may have a common eigenvector corresponding to the eigenvalue 0). Even if the point of minimum is unique, it may change discontinuously with respect to small changes of C.

b) The eigenvalues are nonseparable: when ordering the generalized eigenvalues of the matrix $C^\top C$ in the ascending order, the d-th eigenvalue coincides with the $(d+1)$-th one. Consider the following two cases:
   1) $\operatorname{rank} C < n$. Then problem (3) has a trivial solution $\Delta = 0$, but the eigenspace of the matrix $C - \Delta$ corresponding to the eigenvalue 0 in (4) is $(n+d-\operatorname{rank} C)$-dimensional (its dimension is greater than n).
   2) $\operatorname{rank} C > n$. Then there is an infinite number of points of minimum Δ of problem (3).

c) The system of linear equations (5) may have no solution for such a number Δ that gives the minimum to problem (3).

d) If $\operatorname{rank} \Sigma < d$, then the constraints in (3) may be inconsistent. However, condition (G.3) implies that $\operatorname{rank} \Sigma \geq d$.

We will prove that if the assumptions of at least one of Theorems 4.1, 4.2, or 4.3 hold, then $\langle C^\top C, \Sigma\rangle$ is a definite pair almost surely for a sufficiently large number m. A pair of real symmetric (or complex Hermitian) matrices is called a *definite pair* if a certain linear combination of these matrices is a positive definite matrix.

To avoid the case where system (5) is inconsistent we use the following idea. Given a point of minimum Δ we find $\widehat{X}_{\mathrm{ext}}$ from equation (4) such that the columns of $\widehat{X}_{\mathrm{ext}}$ are linearly independent (of course, the matrices Δ and $\widehat{X}_{\mathrm{ext}}$ should be Borel functions of observations (note that measurable solutions of problem (3) exist); or, if we allow the use of "randomized estimators", then Δ and $\widehat{X}_{\mathrm{ext}}$ should be random matrices).

In what follows we will prove (under the conditions of Theorems 4.1, 4.2, or 4.3) that

$$\left\| \sin \angle(\widehat{X}_{\mathrm{ext}}, X^0_{\mathrm{ext}}) \right\| \to 0, \qquad m \to \infty.$$

The convergence in the latter relation is understood in probability or almost surely depending on a specific setting of the problem. Here

$$\left\| \sin \angle(A, B) \right\| = \left\| P_A P_B^\perp \right\|,$$

where P_A is an orthogonal projector to the space of columns of the matrix A, $P_B^\perp$ is the complementary projector defined by

$$P_B^\perp = I - P_B,$$

and $\|M\|$ is the operator norm of the matrix M defined as the maximum singular value.

Then we show that the lower $d\times d$ block of $\widehat{X}_{\mathrm{ext}}$ is a nonsingular matrix; that is, there is a linear transform of its columns such that the matrix $\widehat{X}_{\mathrm{ext}}$ has the form of $\left(\begin{smallmatrix}\widehat{X}\\ -I_d\end{smallmatrix}\right)$ provided that

$$\left\|\sin\angle(\widehat{X}_{\mathrm{ext}}, X^0_{\mathrm{ext}})\right\| < \frac{1}{\sqrt{1+\|X_0\|^2}}$$

and that the columns of the matrix $\widehat{X}_{\mathrm{ext}}$ are linearly independent.

Proposition 2.1. *Suppose that conditions* (**G.2**) *and* (**G.3**) *hold. If the matrix $A_0^\top A_0$ is positive definite, then $\langle C^\top C,\Sigma\rangle$ is a definite pair almost surely. In other words,*

$$\mathsf{P}(C^\top C+\Sigma>0)=1.$$

Proof. 1. The result is obvious if the matrix Σ is nonsingular. Thus we consider the case where the matrix Σ is singular. By $F=\left(\begin{smallmatrix}F_1\\ F_2\end{smallmatrix}\right)$, we denote a $(n+d)\times(n+d-\operatorname{rank}(\Sigma))$ matrix whose columns form a basis of the subspace $\operatorname{Ker}(\Sigma)=\{x\colon \Sigma x=0\}$.

2. We prove that the columns of the matrix $[I_n\ X_0]\,F$ are linearly independent. If this is not the case, then there exists a vector $v\in\mathbb{R}^{n+d-\operatorname{rank}(\Sigma)}\setminus\{0\}$ such that

$$\begin{gathered}[I_n \quad X_0]\,Fv=0,\\ F_1v=-X_0F_2v,\\ -Fv=\left(\begin{smallmatrix}X_0\\ -I_d\end{smallmatrix}\right)F_2v=X^0_{\mathrm{ext}}F_2v,\\ 0=-\Sigma Fv=\Sigma X^0_{\mathrm{ext}}\cdot F_2v.\end{gathered}\tag{6}$$

Moreover, $Fv\neq0$, since $v\neq0$ and the columns of F are linearly independent, whence $F_2v\neq0$ according to (6).

The result we just obtained contradicts condition (**G.3**), since (**G.3**) implies that the columns of the matrix $\Sigma X^0_{\mathrm{ext}}$ are linearly independent. Thus the columns of the matrix $[IX^0_{\mathrm{ext}}]\,F$ are linearly independent, indeed.

3. Condition (**G.2**) implies that $\tilde{C}F=0$ almost surely. Indeed, $\mathsf{E}\,\tilde{c}_i=0$ and $\operatorname{Var}[\tilde{c}_iF]=0$ for all $i=1,2,\dots,m$.

4. It remains to prove the implication

$$\text{“}A_0^\top A_0>0\text{” and “}\tilde{C}F=0\text{”}\quad\Longrightarrow\quad\text{“}C^\top C+\Sigma>0\text{”}.$$

The matrices $C^\top C$ and Σ are positive semi-definite. Assume that

$$x^\top(C^\top C+\Sigma)x=0.\tag{7}$$

We will show that $x=0$. If this is the case, then $C^\top C+\Sigma>0$. Equality (7) holds only if $Cx=0$ and $\Sigma x=0$. Thus $x\in\operatorname{Ker}\Sigma$ and hence there exists a vector $v\in\mathbb{R}^{n+d-\operatorname{rank}\Sigma}$ such that $x=Fv$. This implies that

$$0=A_0^\top Cx=A_0^\top(C_0+\tilde{C})x=A_0^\top C_0Fv+A_0^\top\tilde{C}Fv=A_0^\top A_0\left[I_n\quad X_0\right]Fv+0.$$

The matrix $A_0^\top A_0$ is nonsingular and the columns of the matrix $[I_n\ X_0]\,F$ are linearly independent, whence we deduce that the columns of the matrix $A_0^\top A_0\,[I_n\ X_0]\,F$ are linearly independent, as well. Therefore $v=0$ and $x=Fv=0$. □

3. Some known results concerning the consistency

Theorem 3.1 (Gallo [4]). *Let $d=1$. Suppose that conditions* (**G.1**)–(**G.3**) *hold. We further assume that*

$$\begin{gathered}m^{-1/2}\lambda_{\min}(A_0^\top A_0)\to\infty,\qquad m\to\infty;\\ \frac{\lambda^2_{\min}(A_0^\top A_0)}{\lambda_{\max}(A_0^\top A_0)}\to\infty,\qquad m\to\infty.\end{gathered}$$

Let the errors of observations have identical distributions such that the fourth moment is finite. Then the total least squares estimator is consistent, that is,

$$\widehat{X} \xrightarrow{\mathsf{P}} X_0, \qquad m \to \infty.$$

Theorem 3.2 (Kukush and Van Huffel [7])**.** *Let conditions* (**G.1**)–(**G.3**) *hold. Assume further that*

$$\sup_{\substack{i\ge 1\\ j=1,\dots,n+d}} \mathsf{E}\,\delta_{ij}^4 < \infty;$$

$$m^{-1/2}\lambda_{\min}(A_0^\top A_0) \to \infty, \qquad m \to \infty;$$

$$\frac{\lambda_{\min}^2(A_0^\top A_0)}{\lambda_{\max}(A_0^\top A_0)} \to \infty, \qquad m \to \infty. \tag{8}$$

Then $\widehat{X} \xrightarrow{\mathsf{P}} X_0$ *as* $m \to \infty$.

Theorem 3.3 (Kukush and Van Huffel [7])**.** *Let conditions* (**G.1**)–(**G.3**) *hold. Assume further that*

$$\sup_{\substack{i\ge 1\\ j=1,\dots,n+d}} \mathsf{E}\,|\delta_{ij}|^{2r} < \infty;$$

$$\sum_{m=m_0}^{\infty} \left(\frac{\sqrt{m}}{\lambda_{\min}(A_0^\top A_0)}\right)^r < \infty;$$

$$\sum_{m=m_0}^{\infty} \left(\frac{\lambda_{\max}(A_0^\top A_0)}{\lambda_{\min}^2(A_0^\top A_0)}\right)^r < \infty$$

for some $r \ge 2$ *and* $m_0 \ge 1$. *Then the total least squares estimator is strongly consistent, that is,*

$$\widehat{X} \xrightarrow{\mathsf{P1}} X_0, \qquad m \to \infty.$$

Theorem 3.4 (Kukush and Van Huffel [7])**.** *Let conditions* (**G.1**)–(**G.3**) *hold. Assume further that*

$$\sup_{\substack{i\ge 1\\ j=1,\dots,n+d}} \mathsf{E}\,|\delta_{ij}|^{2r} < \infty;$$

$$m^{-1/r}\lambda_{\min}(A_0^\top A_0) \to \infty, \qquad m \to \infty;$$

$$\frac{\lambda_{\min}^2(A_0^\top A_0)}{\lambda_{\max}(A_0^\top A_0)} \to \infty, \qquad m \to \infty,$$

for some r *such that* $1 \le r < 2$. *Then* $\widehat{X} \xrightarrow{\mathsf{P}} X_0$ *as* $m \to \infty$.

4. Main results

Theorem 4.1. *Let conditions* (**G.1**)–(**G.3**) *hold. Assume that*

$$\sup_{\substack{i\ge 1\\ j=1,\dots,n+d}} \mathsf{E}\,|\delta_{ij}|^{2r} < \infty,$$

$$m^{-1/r}\lambda_{\min}(A_0^\top A_0) \to \infty, \qquad m \to \infty,$$

for some r *such that* $1 \le r \le 2$. *Then* $\widehat{X} \xrightarrow{\mathsf{P}} X_0$ *as* $m \to \infty$.

Theorem 4.2. *Let conditions* (**G.1**)–(**G.3**) *hold. Assume that*

$$\sup_{\substack{i\ge 1\\ j=1,\dots,n+d}} \mathsf{E}\,|\delta_{ij}|^{2r} < \infty;$$

$$\sum_{m=m_0}^{\infty} \left(\frac{\sqrt{m}}{\lambda_{\min}(A_0^\top A_0)}\right)^r < \infty$$

for some $r \ge 2$ *and* $m_0 \ge 1$. *Then* $\widehat{X} \xrightarrow{\mathsf{P1}} X_0$ *as* $m \to \infty$.

Theorem 4.3. *Let conditions* (**G.1**)–(**G.3**) *hold. Assume that*

$$\sup_{\substack{i\ge 1\\ j=1,\dots,n+d}} \mathsf{E}\,|\delta_{ij}|^{2r} < \infty;$$

$$\sum_{m=m_0}^{\infty} \frac{1}{\lambda_{\min}^r(A_0^\top A_0)} < \infty$$

for some r *such that* $1 \le r \le 2$ *and* $m_0 \ge 1$. *Then* $\widehat{X} \xrightarrow{\mathsf{P1}} X_0$ *as* $m \to \infty$.

Proof. This part of the proof is common for all Theorems 4.1, 4.2, and 4.3.

1. The $(d{+}1)$*-th eigenvalue of the matrix* $C_0^\top C_0$. First, the matrix $C_0^\top C_0$ is symmetric and nonnegative definite. Since $C_0 X_{\text{ext}}^0 = A_0 X_0 - B_0 = 0$, this matrix has at least a d-dimensional eigenspace corresponding to the eigenvalue 0.

Next we prove that

$$\lambda_{d+1}(C_0^\top C_0) \ge \lambda_{\min}(A_0^\top A_0). \tag{9}$$

Recall that the eigenvalues are ordered in ascending order. Note also a more general inequality

$$\lambda_{d+j}(C_0^\top C_0) \ge \lambda_j(A_0^\top A_0), \qquad j{=}1,\dots,n.$$

The latter inequality is a corollary of Theorem I.4.4 of [9], since

$$\lambda_{d+1}(C_0^\top C_0) = \sigma_n^2(C_0) \ge \sigma_n^2(A_0) = \lambda_{\min}(A_0^\top A_0).$$

The usual order is kept for singular values:

$$\sigma_1(A_0) \ge \sigma_2(A_0) \ge \cdots.$$

Inequality (9) implies that if the matrix $A_0^\top A_0$ is nonsingular, then $\lambda_{d+1}(C_0^\top C_0) > 0$, and thus $\operatorname{rank}(C_0^\top C_0) = d$. The conditions of any of Theorems 4.1, 4.2, or 4.3 imply that

$$\lambda_{\min}(A_0^\top A_0) \to \infty.$$

Thus if the assumptions of at least one of these theorems hold, then

$$\lambda_{d+1}(C_0^\top C_0) \ge \lambda_{\min}(A_0^\top A_0) > 0$$

for sufficiently large m.

2. Spectral decomposition of the matrices $C_0^\top C_0$, N, *and* $N^{-1/2} C_0^\top C_0 N^{-1/2}$. Below is the canonical decomposition of the matrix $C_0^\top C_0$:

$$\begin{aligned} C_0^\top C_0 &= U \operatorname{diag}\left(\lambda_{\min}(C_0^\top C_0), \lambda_2(C_0^\top C_0), \dots, \lambda_{\max}(C_0^\top C_0)\right) U^\top \\ &= U \operatorname{diag}\left(\left(\lambda_j(C_0^\top C_0)\right)_{j=1}^{n+d}\right) U^\top, \end{aligned}$$

where U is an orthogonal matrix ($U^{-1} = U^\top$) and

$$\lambda_j(C_0^\top C_0) = 0, \qquad j{=}1,\dots,d.$$

In particular, $\lambda_{\min}(C_0^\top C_0) = 0$. Put

$$N = C_0^\top C_0 + \lambda_{\min}(A_0^\top A_0) I.$$

Then the canonical decomposition of the matrix N is given by

$$N = U \operatorname{diag}\left(\left(\lambda_j(C_0^\top C_0) + \lambda_{\min}(A_0^\top A_0)\right)_{j=1}^{n+d}\right) U^\top.$$

Note that

$$\lambda_{\min}(N) = \cdots = \lambda_d(N) = \lambda_{\min}(A_0^\top A_0). \tag{10}$$

If the matrix $A_0^\top A_0$ is nonsingular, then the matrix N is nonsingular, too. Note that the matrix N is nonsingular for sufficiently large m provided the assumptions of one of Theorems 4.1, 4.2, or 4.3 hold.

Since $C_0 X_{\text{ext}}^0 = 0$, we get

$$N X_{\text{ext}}^0 = \lambda_{\min}(A_0^\top A_0) X_{\text{ext}}^0. \tag{11}$$

If the matrix N is nonsingular, then the canonical decomposition of the matrices $N^{-1/2}$ and $N^{-1/2} C_0^\top C_0 N^{-1/2}$ is given by

$$N^{-1/2} = U \operatorname{diag}\left(\left(\frac{1}{\sqrt{\lambda_j(C_0^\top C_0) + \lambda_{\min}(A_0^\top A_0)}}\right)_{j=1}^{n+d}\right) U^\top,$$
$$N^{-1/2} C_0^\top C_0 N^{-1/2} = U \operatorname{diag}\left(\left(\frac{\lambda_j(C_0^\top C_0)}{\lambda_j(C_0^\top C_0) + \lambda_{\min}(A_0^\top A_0)}\right)_{j=1}^{n+d}\right) U^\top.$$

The above decompositions together with (9) imply the following equalities and inequalities for the eigenvalues:

$$\left\|N^{-1/2}\right\| = \lambda_{\max}\left(N^{-1/2}\right) = \frac{1}{\sqrt{\lambda_{\min}(A_0^\top A_0)}}; \tag{12}$$

$$\lambda_j\left(N^{-1/2} C_0^\top C_0 N^{-1/2}\right) = 0, \qquad j = 1, \dots, d; \tag{13}$$

$$\frac{1}{2} \le \lambda_j\left(N^{-1/2} C_0^\top C_0 N^{-1/2}\right) \le 1, \qquad j = d+1, \dots, n+d. \tag{14}$$

3. An upper bound for $\left\|\sin\angle(\widehat{X}_{\text{ext}}, X_{\text{ext}}^0)\right\|$. Equality (10) implies that

$$\widehat{X}_{\text{ext}}^\top N \widehat{X}_{\text{ext}} \ge \lambda_{\min}(A_0^\top A_0) \widehat{X}_{\text{ext}}^\top \widehat{X}_{\text{ext}}$$

and

$$\frac{v^\top \widehat{X}_{\text{ext}}^\top X_{\text{ext}}^0 \left(X_{\text{ext}}^{0\top} X_{\text{ext}}^0\right)^{-1} X_{\text{ext}}^{0\top} \widehat{X}_{\text{ext}} v}{v^\top \widehat{X}_{\text{ext}}^\top \widehat{X}_{\text{ext}} v}$$
$$\ge \lambda_{\min}(A_0^\top A_0) \frac{v^\top \widehat{X}_{\text{ext}}^\top X_{\text{ext}}^0 \left(X_{\text{ext}}^{0\top} X_{\text{ext}}^0\right)^{-1} X_{\text{ext}}^{0\top} \widehat{X}_{\text{ext}} v}{v^\top \widehat{X}_{\text{ext}}^\top N \widehat{X}_{\text{ext}} v}$$

for all $v \in \mathbb{R}^d \setminus \{0\}$. Taking into account (11) we get

$$\frac{v^\top \widehat{X}_{\text{ext}}^\top X_{\text{ext}}^0 \left(X_{\text{ext}}^{0\top} X_{\text{ext}}^0\right)^{-1} X_{\text{ext}}^{0\top} \widehat{X}_{\text{ext}} v}{v^\top \widehat{X}_{\text{ext}}^\top \widehat{X}_{\text{ext}} v} \ge \frac{v^\top N \widehat{X}_{\text{ext}}^\top X_{\text{ext}}^0 \left(X_{\text{ext}}^{0\top} N X_{\text{ext}}^0\right)^{-1} X_{\text{ext}}^{0\top} N \widehat{X}_{\text{ext}} v}{v^\top \widehat{X}_{\text{ext}}^\top N \widehat{X}_{\text{ext}} v}.$$

Using (22) we prove the inequality

$$1-\left\|\sin\angle(\widehat{X}_{\mathrm{ext}},X_{\mathrm{ext}}^{0})\right\|^{2}\geq 1-\left\|\sin\angle(N^{1/2}\widehat{X}_{\mathrm{ext}},N^{1/2}X_{\mathrm{ext}}^{0})\right\|^{2},$$

$$\left\|\sin\angle(\widehat{X}_{\mathrm{ext}},X_{\mathrm{ext}}^{0})\right\|\leq\left\|\sin\angle(N^{1/2}\widehat{X}_{\mathrm{ext}},N^{1/2}X_{\mathrm{ext}}^{0})\right\|. \tag{15}$$

Put

$$\epsilon=\left\|N^{-1/2}\left(C^{\top}C-C_0^{\top}C_0-m\Sigma\right)N^{-1/2}\right\|.$$

Now we use Corollary 5.5 of Lemma 5.4 on the stability of the eigenspace. Let $N^{-1/2}C_0^{\top}C_0N^{-1/2}$ be the nonperturbed matrix whose null space is the column space of $N^{1/2}X_{\mathrm{ext}}^0$. Further let $N^{-1/2}(C^{\top}C-m\Sigma)N^{-1/2}$ be the perturbed matrix whose generalized invariant space with respect to the matrix $N^{-1/2}\Sigma N^{-1/2}$ corresponding to the d minimal generalized eigenvalues is the column space of the matrix $N^{1/2}\widehat{X}_{\mathrm{ext}}$. Using (13)–(14) we establish

$$\begin{aligned}&\left\|\sin\angle(N^{1/2}\widehat{X}_{\mathrm{ext}},N^{1/2}X_{\mathrm{ext}}^{0})\right\|^{2}\\&\qquad\leq\frac{\epsilon}{0.5}\left(1+\left\|N^{-1/2}\Sigma N^{-1/2}\right\|\lambda_{\max}\left((X_{\mathrm{ext}}^{0\top}\Sigma X_{\mathrm{ext}}^{0})^{-1}X_{\mathrm{ext}}^{0\top}NX_{\mathrm{ext}}^{0}\right)\right).\end{aligned}$$

Taking into account equalities (11) and (12) and inequality (15) we obtain

$$\begin{aligned}&\left\|\sin\angle(\widehat{X}_{\mathrm{ext}},X_{\mathrm{ext}}^{0})\right\|^{2}\leq\left\|\sin\angle(N^{1/2}\widehat{X}_{\mathrm{ext}},N^{1/2}X_{\mathrm{ext}}^{0})\right\|^{2}\\&\quad\leq 2\epsilon\left(1+\frac{\|\Sigma\|}{\lambda_{\min}(A_0^{\top}A_0)}\lambda_{\max}\left(\lambda_{\min}(A_0^{\top}A_0)(X_{\mathrm{ext}}^{0\top}\Sigma X_{\mathrm{ext}}^{0})^{-1}X_{\mathrm{ext}}^{0\top}X_{\mathrm{ext}}^{0}\right)\right)\\&\quad=2\epsilon\left(1+\|\Sigma\|\lambda_{\max}\left((X_{\mathrm{ext}}^{0\top}\Sigma X_{\mathrm{ext}}^{0})^{-1}X_{\mathrm{ext}}^{0\top}X_{\mathrm{ext}}^{0}\right)\right).\end{aligned}$$

To complete the proof, it remains to check that $\epsilon\to 0$ (this part of the proof is specific and is given separately for each of Theorems 4.1, 4.2, and 4.3). If $\epsilon\to 0$, then $\|\sin\angle(\widehat{X}_{\mathrm{ext}},X_{\mathrm{ext}}^0)\|\to 0$ and thus the matrix $\widehat{X}_{\mathrm{ext}}$ is transformed to the form $\left(\begin{smallmatrix}\widehat{X}\\-I\end{smallmatrix}\right)$ by transforming its columns for sufficiently large m (this is justified by Lemma 5.8). Moreover, $\widehat{X}\to X_0$ in this case. □

We introduce two $(n+d)\times(n+d)$ matrices:

$$M_1=N^{-1/2}C_0^{\top}\widetilde{C}N^{-1/2},$$
$$M_2=N^{-1/2}(\widetilde{C}^{\top}\widetilde{C}-m\Sigma)N^{-1/2}.$$

Since $\epsilon=\|M_1+M_1^{\top}+M_2\|$, we need to prove that $M_1\to 0$ and $M_2\to 0$ as $m\to\infty$ (in probability or almost surely depending on whether we prove the consistency or strong consistency of an estimator).

Rest of the proof of Theorem 4.1. We have

$$\begin{aligned}\|M_1\|_F^2&=\left\|N^{-1/2}C_0^{\top}\tilde{C}N^{-1/2}\right\|_F^2=\operatorname{tr}\left(N^{-1/2}C_0^{\top}\tilde{C}N^{-1}C_0\tilde{C}^{\top}N^{-1/2}\right)\\&=\operatorname{tr}\left(C_0N^{-1}C_0^{\top}\tilde{C}N^{-1}\tilde{C}^{\top}\right)=\sum_{i=1}^{m}\sum_{j=1}^{m}c_{0i}N^{-1}c_{0j}^{\top}\tilde{c}_jN^{-1}\tilde{c}_i^{\top}.\end{aligned}$$

Since $\mathsf{E}\,\tilde{c}_jN^{-1}\tilde{c}_i^{\top}=0$ for $i\neq j$ and $\mathsf{E}\,\tilde{c}_iN^{-1}\tilde{c}_i^{\top}=\operatorname{tr}(\Sigma N^{-1})$, we get

$$\mathsf{E}\,\|M_1\|_F^2=\sum_{i=1}^{m}c_{0i}N^{-1}c_{0i}^{\top}\operatorname{tr}\left(\Sigma N^{-1}\right)=\operatorname{tr}\left(C_0N^{-1}C_0^{\top}\right)\operatorname{tr}\left(\Sigma N^{-1}\right).$$

Now we estimate the factors on the right hand side of the latter equality:

$$\operatorname{tr}\left(C_0 N^{-1} C_0^\top\right) = \operatorname{tr}\left(C_0^\top C_0 N^{-1}\right) = \sum_{i=1}^{n+d} \lambda_i\left(C_0^\top C_0 N^{-1}\right).$$

Since $\lambda_i(C_0^\top C_0 N^{-1}) = 0$ for $i \le d$ and $\frac{1}{2} \le \lambda_i(C_0^\top C_0 N^{-1}) \le 1$ for $d < i \le d+n$, we conclude that

$$\frac{n}{2} \le \operatorname{tr}\left(C_0 N^{-1} C_0^\top\right) \le n. \tag{16}$$

Since both matrices N^{-1} and Σ are positive semi-definite,

$$\begin{aligned}\operatorname{tr}(\Sigma N^{-1}) &= \left\|N^{-1/2}\Sigma^{1/2}\right\|_F^2 \le \left\|N^{-1/2}\right\|^2 \left\|\Sigma^{1/2}\right\|_F^2 = \lambda_{\max}\left(N^{-1}\right) \operatorname{tr}\Sigma \\ &= \lambda_{\min}\left({}_0^\top A_0\right) \operatorname{tr}\Sigma.\end{aligned}$$

Finally,

$$\mathsf{E}\,\|M_1\|_F^2 \le \frac{n \operatorname{tr}\Sigma}{\lambda_{\min}(A_0^\top A_0)}.$$

The assumptions of Theorem 4.1 imply that $\lambda_{\max}(A_0^\top A_0) \to \infty$, whence $M_1 \xrightarrow{\mathsf{P}} 0$ as $m \to \infty$.

Now we prove the convergence $M_2 \xrightarrow{\mathsf{P}} 0$ as $m \to \infty$:

$$M_2 = N^{-1/2}(\tilde{C}^\top \tilde{C} - m\Sigma)N^{-1/2},$$

$$\|M_2\| \le \left\|N^{-1/2}\right\| \left\|\tilde{C}^\top \tilde{C} - m\Sigma\right\| \left\|N^{-1/2}\right\| = \frac{\left\|\sum_{i=1}^m (\tilde{c}_i^\top \tilde{c}_i - \Sigma)\right\|}{\lambda_{\min}(A_0^\top A_0)}. \tag{17}$$

To estimate $\|M_2\|^r$ we use the Rosenthal inequality for $1 \le \nu \le 2$ (see Theorem 5.7 below):

$$\mathsf{E}\,\|M_2\|^r \le \frac{\operatorname{const} \sum_{i=1}^m \mathsf{E}\left\|\tilde{c}_i^\top \tilde{c}_i - \Sigma\right\|^r}{\lambda_{\min}^r(A_0^\top A_0)}.$$

According to the assumptions of Theorem 4.1, the sequence $\left\{\mathsf{E}\,\|\tilde{c}_i^\top \tilde{c}_i - \Sigma\|^r, i = 1, 2, \ldots\right\}$ is bounded, whence

$$\mathsf{E}\,\|M_2\|^r \le \frac{O(m)}{\lambda_{\min}^r(A_0^\top A_0)}, \qquad m \to \infty.$$

Thus $\mathsf{E}\,\|M_2\|^r \to 0$ and $M_2 \xrightarrow{\mathsf{P}} 0$ as $m \to \infty$. □

Rest of the proof of Theorem 4.2. We have

$$M_1 = \sum_{i=1}^m N^{-1/2} c_{0i}^\top \tilde{c}_i N^{-1/2}.$$

By the Rosenthal inequality,

$$\begin{aligned}\mathsf{E}\,\|M_1\|^{2r} \le{}& \operatorname{const} \sum_{i=1}^m \mathsf{E}\left\|N^{-1/2} c_{0i}^\top \tilde{c}_i N^{-1/2}\right\|^{2r} \\ &+ \operatorname{const}\left(\sum_{i=1}^m \mathsf{E}\left\|N^{-1/2} c_{0i}^\top \tilde{c}_i N^{-1/2}\right\|^2\right)^r.\end{aligned}$$

The first term is estimated as follows:

$$\sum_{i=1}^m \mathsf{E}\left\|N^{-1/2} c_{0i}^\top \tilde{c}_i N^{-1/2}\right\|^{2r} \le \sum_{i=1}^m \left\|N^{-1/2} c_{0i}^\top\right\|^{2r} \max_{i=1,\ldots,m} \mathsf{E}\,\|\tilde{c}_i\|^{2r} \left\|N^{-1/2}\right\|^{2r}.$$

Thus

$$\sum_{i=1}^{m} \|N^{-1/2} c_{0i}^\top\|^{2r} \le \left(\sum_{i=1}^{m} \|N^{-1/2} c_{0i}^\top\|^2\right)^r$$
$$= \left(\sum_{i=1}^{m} c_{0i} N^{-1} c_{0i}^\top\right)^r = \left(\operatorname{tr}(C_0 N^{-1} C_0^\top)\right)^r \le n^r$$

by inequality (16). Note that inequality (16) holds if the matrix N is nonsingular. This is the case not only under the assumptions of Theorem 4.1, this also holds under the assumptions of Theorems 4.2 or 4.3. Also, the conditions of Theorem 4.2 imply that the sequence

$$\left\{\max_{i=1,\dots,m} \mathsf{E}\,\|\tilde c_i\|^{2r}, m = 1, 2, \dots\right\}$$

is bounded. We recall that $\|N^{-1/2}\| = \lambda_{\min}^{-1/2}(A_0^\top A_0)$. Thus

$$\sum_{i=1}^{m} \mathsf{E}\left\|N^{-1/2} c_{0i}^\top \tilde c_i N^{-1/2}\right\|^{2r} = \frac{O(1)}{\lambda_{\min}^r(A_0^\top A_0)}, \qquad m\to\infty.$$

The limit relation

$$\sum_{i=1}^{m} \mathsf{E}\left\|N^{-1/2} c_{0i}^\top \tilde c_i N^{-1/2}\right\|^{2} = \frac{O(1)}{\lambda_{\min}(A_0^\top A_0)}$$

is proved similarly. It is important for the proof of this relation that the sequence

$$\left\{\max_{i=1,\dots,m} \mathsf{E}\,\|\tilde c_i\|^{2}, m = 1, 2, \dots\right\}$$

is bounded by the conditions of Theorem 4.2.

Finally,

$$\mathsf{E}\,\|M_1\|^{2r} = \frac{O(1)}{\lambda_{\min}^r(A_0^\top A_0)}, \qquad m \to \infty.$$

Hence the conditions of Theorem 4.2 imply that $\sum_{m=m_0}^{\infty} \mathsf{E}\,\|M_1\|^{2r} < \infty$, whence $M_1 \xrightarrow{\mathsf{P1}} 0$ as $m \to \infty$.

Now we prove the convergence $M_2 \xrightarrow{\mathsf{P1}} 0$. To estimate $\mathsf{E}\,\|M_2\|^r$, we apply Rosenthal's inequality (also see (17)):

$$\mathsf{E}\,\|M_2\|^r \le \frac{\mathsf{E}\left\|\sum_{i=1}^m (\tilde c_i^\top \tilde c_i - \Sigma)\right\|^r}{\lambda_{\min}^r(A_0^\top A_0)}$$
$$\le \frac{\operatorname{const} \sum_{i=1}^m \mathsf{E}\,\|\tilde c_i^\top \tilde c_i - \Sigma\|^r}{\lambda_{\min}^r(A_0^\top A_0)} + \frac{\operatorname{const} \left(\sum_{i=1}^m \mathsf{E}\,\|\tilde c_i^\top \tilde c_i - \Sigma\|^2\right)^{r/2}}{\lambda_{\min}^r(A_0^\top A_0)}.$$

The sequences

$$\left\{\mathsf{E}\,\|\tilde c_i^\top \tilde c_i - \Sigma\|^r, i = 1, 2, \dots\right\} \quad \text{and} \quad \left\{\mathsf{E}\,\|\tilde c_i^\top \tilde c_i - \Sigma\|^2, i = 1, 2, \dots\right\}$$

are bounded by the assumptions of Theorem 4.2. Thus

$$\mathsf{E}\,\|M_2\|^r = \frac{O\left(m^{r/2}\right)}{\lambda_{\min}^r(A_0^\top A_0)}, \qquad m \to \infty;$$
$$\sum_{m=m_0}^{\infty} \mathsf{E}\,\|M_2\|^r < \infty$$

for some $m_0 \ge 1$, whence we conclude that $M_2 \xrightarrow{\mathsf{P1}} 0$ as $m \to \infty$. □

Rest of the proof of Theorem 4.3. The relation

$$\mathsf{E}\,\|M_1\|^{2r} = \frac{O(1)}{\lambda_{\min}^r(A_0^\top A_0)}, \qquad m \to \infty,$$

is proved in the same way as in the proof of Theorem 4.2, where we dealt with the case of $r \ge 1$. As in the proof of Theorem 4.2, we show that $M_1 \xrightarrow{\mathsf{P1}} 0$ as $m \to \infty$.

Now we prove the convergence $M_2 \xrightarrow{\mathsf{P1}} 0$ as $m \to \infty$. By the assumptions of Theorem 4.3,

$$\sum_{m=1}^{\infty} \frac{\mathsf{E}\,\|\tilde{c}_m^\top \tilde{c}_m - \Sigma\|^r}{\lambda_{\min}^r(A_0^\top A_0)} < \infty.$$

The random matrices $\tilde{c}_i^\top \tilde{c}_i - \Sigma$ have zero expectations. The sequence of nonnegative numbers $\{\lambda_{\min}(A_0^\top A_0), m = 1, 2, \ldots\}$ goes to $+\infty$ and is nondecreasing, since $A_0^\top A_0$ is nondecreasing in Loewner's order. By the strong law of large numbers [1, Theorem IX.12],

$$\frac{1}{\lambda_{\min}(A_0^\top A_0)} \sum_{i=1}^{m} (\tilde{c}_i^\top \tilde{c}_i - \Sigma) \xrightarrow{\mathsf{P1}} 0, \qquad m \to \infty.$$

Recalling (17), we get

$$\|M_2\| \le \frac{\left\|\sum_{i=1}^m (\tilde{c}_i^\top \tilde{c}_i - \Sigma)\right\|}{\lambda_{\min}(A_0^\top A_0)} \xrightarrow{\mathsf{P1}} 0, \qquad m \to \infty,$$

$$M_2 \xrightarrow{\mathsf{P1}} 0, \qquad m \to \infty. \qquad \square$$

5. Auxiliary results

5.1. Generalized eigenvectors.

Theorem 5.1 (A proper decomposition of a definite pair of matrices)**.** *Let two $n \times n$ matrices A and B be real symmetric or complex Hermitian. Let the matrix $\alpha A + \beta B$ be positive definite for some real numbers α and β. Then there exist a nonsingular matrix T_1 and diagonal matrices Λ and* M *such that*

$$A = \left(T_1^{-1}\right)^\top \Lambda T_1^{-1}, \qquad B = \left(T_1^{-1}\right)^\top \mathrm{M} T_1^{-1}.$$

Let

$$T_1 = [u_1, u_2, \ldots, u_n], \qquad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \qquad \mathrm{M} = \operatorname{diag}(\mu_1, \ldots, \mu_n)$$

in the above decomposition. Then the numbers $\lambda_i/\mu_i \in \mathbb{R} \cup \{\infty\}$ are called *generalized eigenvalues* and the columns u_i of the matrix T_1 are called *generalized right eigenvectors* of the matrix A with respect to B. Also,

$$\mu_i A u_i = \lambda_i B u_i.$$

Theorem 5.2 (Fischer)**.** *Let A and B be real symmetric matrices of the same sizes. Let there exist real numbers $\alpha > 0$ and $\beta > 0$ such that the matrix $\alpha A + \beta B$ is positive definite. Let the matrix B be positive semidefinite. We order the finite generalized eigenvalues in the ascending order, namely $\lambda_1/\mu_1 \le \lambda_2/\mu_2 \le \cdots \le \lambda_{\operatorname{rank} B}/\mu_{\operatorname{rank} B}$. Then*

$$\frac{\lambda_i}{\mu_i} = \min_{\substack{\dim V = i \\ B|_V > 0}} \max_{v \in V \setminus \{0\}} \frac{v^\top A v}{v^\top B v}, \qquad i = 1, 2, \ldots, \operatorname{rank} B.$$

The minimum in the latter relation is taken over i-dimensional subspaces of $\mathbb{R}^n$ that have the trivial intersection with the null subspace of the matrix B.

Theorem 5.2 is a corollary of Lemma VI.3.1 from [9].

5.2. **Perturbation of eigenvectors.**

Lemma 5.3. *Let A, B, and $\tilde{A}$ be symmetric matrices,*

$$\lambda_{\min}(A) = 0, \qquad \lambda_2(A) > 0, \qquad \lambda_{\min}(B) \ge 0.$$

Let $Ax_0 = 0$ and $Bx_0 \ne 0$. Let $x \ne 0$ be either a point of minimum of the function

$$f(x) := \frac{x^\top (A + \tilde{A})x}{x^\top Bx}$$

or a point such that

$$\liminf_{t \to x} f(t) = \inf_{t\colon t^\top Bt > 0} f(t). \tag{18}$$

Note that such a point x exists by assumptions of the lemma. Then

$$\sin^2 \angle(x, x_0) \le \frac{\|\tilde{A}\|}{\lambda_2(A)} \left(1 + \frac{\|x_0\|^2}{x_0^\top Bx_0} \frac{x^\top Bx}{\|x\|^2}\right).$$

The "angle" between two d-dimensional subspaces $V_1 \subset \mathbb{R}^{d+n}$ and $V_2 \subset \mathbb{R}^{d+n}$ is defined by $\min(d, n)$ canonical angles. The largest sinus of the canonical angles is denoted by

$$\|\sin \angle(V_1, V_2)\|.$$

The following is equality (1.5) of [10]:

$$\|\sin \angle(V_1, V_2)\| = \|P_{V_1}(I - P_{V_2})\|. \tag{19}$$

Recall that $\|\cdot\|$ is the operator norm of a matrix. Applying equality (19), one can prove that

$$\|\sin \angle(V_1, V_2)\|^2 = \min_{v \in V_1 \setminus \{0\}} \frac{v^\top P_{V_2}^\perp v}{\|v\|^2}. \tag{20}$$

If the columns of the matrix X form a basis of the subspace V_1, then

$$\|\sin \angle(X, V_2)\|^2 := \|\sin \angle(V_1, V_2)\|^2 = \lambda_{\max}\left((X^\top X)^{-1} X^\top P_{V_2}^\perp X\right), \tag{21}$$

$$1 - \|\sin \angle(X, V_2)\|^2 = \lambda_{\min}\left((X^\top X)^{-1} X^\top P_{V_2} X\right). \tag{22}$$

The following Lemma 5.4 is a multidimensional generalization of Lemma 5.3. The existence of the minimum is one of the assumptions of Lemma 5.4.

Lemma 5.4. *Let A, B, and $\tilde{A}$ be symmetric $n \times n$ matrices and let $\lambda_i(A) = 0$ for all i=$1, \ldots, d$ (in particular, $\lambda_{\min}(A) = 0$), $\lambda_{d+1}(A) > 0$, and $\lambda_{\min}(B) \ge 0$. Let X_0 be an $n \times d$ matrix such that $AX_0 = 0$ and the matrix $X_0^\top BX_0$ is nonsingular (this implies that $X_0^\top BX_0 > 0$ and* $\operatorname{rank} X_0 = d$*).*

Let the functional

$$f(X) = \lambda_{\max}\Big((X^\top BX)^{-1} X^\top (A + \tilde{A})X\Big), \qquad X \in \mathbb{R}^{n \times d},\ X^\top BX > 0,$$

attain its minimum. Then, for all points of minimum X,

$$\|\sin \angle(X, X_0)\|^2 \le \frac{\|\tilde{A}\|}{\lambda_{d+1}(A)} \left(1 + \|B\| \lambda_{\max}\left((X_0^\top BX_0)^{-1} X_0^\top X_0\right)\right).$$

Proof. If A and B are two symmetric matrices of the same sizes and if $B > 0$, then we write $\max \frac{A}{B}$ instead of $\lambda_{\max}(B^{-1}A)$. A motivation for such a change of notation is that

$$\lambda_{\max}\left(B^{-1}A\right) = \max_{v \ne 0} \frac{v^\top Av}{v^\top Bv} =: \max \frac{A}{B}.$$

Let X be a point of minimum of the functional $f(X)$. Since the functional $f(X)$ is defined at the point X_0, $f(X) \le f(X_0)$, that is,

$$\max \frac{X^\top (A+\tilde{A})X}{X^\top BX} \le \max \frac{X_0^\top (A+\tilde{A})X_0}{X_0^\top BX_0}.$$

Since

$$\begin{gathered} X^\top \tilde{A} X \ge -\|\tilde{A}\|\, X^\top X, \\ X^\top BX \le \|B\|\, X^\top X, \\ X_0^\top \tilde{A} X_0 \le \|\tilde{A}\|\, X_0^\top X_0, \\ AX_0 = 0, \end{gathered}$$

we obtain

$$\max \frac{X^\top AX - \|\tilde{A}\| X^\top X}{\|B\|\, X^\top X} \le \max \frac{\|\tilde{A}\|\, X_0^\top X_0}{X_0^\top BX_0},$$

$$\frac{1}{\|B\|} \cdot \left(\max \frac{X^\top AX}{X^\top X} - \|\tilde{A}\| \right) \le \|\tilde{A}\| \max \frac{X_0^\top X_0}{X_0^\top BX_0}. \tag{23}$$

Equality (21) implies that

$$\lambda_{d+1}(A)\, \|\sin \angle(X, X_0)\|^2 \le \max \frac{X^\top AX}{X^\top X},$$

since $A \ge \lambda_{d+1}(A) P_{X_0}^\perp$ in the sense of Loewner's order. Thus (23) implies that

$$\|\sin \angle(X, X_0)\|^2 \le \frac{\|\tilde{A}\|}{\lambda_{d+1}(A)} \left(1 + \|B\| \max \frac{X_0^\top X_0}{X_0^\top BX_0} \right)$$

and this is what was to be proved. □

Corollary 5.5. *Let A, B, and $\tilde{A}$ be symmetric $n \times n$ matrices and let $\lambda_i(A) = 0$ for all $i = 1, \dots, d$ (in particular, $\lambda_{\min}(A) = 0$), $\lambda_{d+1}(A) > 0$, and $\lambda_{\min}(B) \ge 0$. Let X_0 be an $n \times d$ matrix such that $AX_0 = 0$ and the matrix $X_0^\top BX_0$ is nonsingular (this implies that $X_0^\top BX_0 > 0$, $\operatorname{rank} X_0 = d$, and $\operatorname{rank} B \ge d$).*

Let there exist a scalar $k > 0$ such that the matrix $A + \tilde{A} + kB$ is positive definite (whence one concludes that $\langle A + \tilde{A}, B\rangle$ is a definite matrix pair). Let V_1 be a d-dimensional generalized invariant subspace corresponding to the minimal finite generalized eigenvalues of the matrix $A + \tilde{A}$ with respect to B.

Then

$$\|\sin \angle(V_1, X_0)\|^2 \le \frac{\|\tilde{A}\|}{\lambda_{d+1}(A)} \left(1 + \|B\| \lambda_{\max} \left((X_0^\top BX_0)^{-1} X_0^\top X_0 \right) \right).$$

5.3. **Rosenthal inequality.**

Theorem 5.6. *Let $\nu \ge 2$ be a nonrandom number. Then there are $\alpha \ge 0$ and $\beta \ge 0$ such that*

$$\mathsf{E} \left[\left| \sum_{i=1}^m \xi_i \right|^\nu \right] \le \alpha \sum_{i=1}^m \mathsf{E}\, [|\xi_i|^\nu] + \beta \left(\sum_{i=1}^m \mathsf{E}\, \xi_i^2 \right)^{\nu/2}$$

if the random variables $\{\xi_i, i = 1, \dots, m\}$, $m \ge 1$, are independent and have zero expectations, that is, $\mathsf{E}\, \xi_i = 0$ for all $i = 1, \dots, m$.

A proof of this result can be found in [6].

Theorem 5.7. *Let ν be a nonrandom number such that $1 \le \nu \le 2$. Then there exists a nonrandom number $\alpha \ge 0$ such that*

$$\mathsf{E}\left[\left|\sum_{i=1}^{m} \xi_i\right|^{\nu}\right] \le \alpha \sum_{i=1}^{m} \mathsf{E}\left[|\xi_i|^{\nu}\right]$$

if the random variables $\{\xi_i, i = 1, \dots, m\}$, $m \ge 1$, are independent and have zero expectations, that is, $\mathsf{E}\,\xi_i = 0$ for all $i = 1, \dots, m$.

Theorem 5.7 follows from Utev's interpolation lemma for moments of sums of weakly dependent random variables [2]. Perhaps, Theorem 5.7 is not new and has already been published elsewhere.

5.4. **An inequality between $\|\sin\angle(\widehat{X}_{\mathrm{ext}}, X^0_{\mathrm{ext}})\|$ and $\|\widehat{X} - X_0\|$.**

Lemma 5.8. *Let $\binom{X_0}{-I}$ be an $(n+d) \times d$ matrix and let $\left\{\binom{A_m}{B_m},\ m = 1, 2, \dots\right\}$ be a sequence of $(n+d) \times d$ matrices of rank d (the columns of every matrix are assumed to be linearly independent). If*

$$\left\|\sin\angle\left(\begin{pmatrix} A_m \\ B_m \end{pmatrix}, \begin{pmatrix} X_0 \\ -I \end{pmatrix}\right)\right\| \to 0 \quad \text{as } m \to \infty,$$

then

1) *the matrix B_m is nonsingular if m is sufficiently large;*
2) *$-A_m B_m^{-1} \to X_0$ as $m \to \infty$.*

Proof. Step 1. Throughout the proof, $\mathbf{P}_1$ denotes an $n \times n$ submatrix of the matrix $P^{\perp}_{\binom{X_0}{-I}}$:

$$I - P_{\binom{X_0}{-I}} = P^{\perp}_{\binom{X_0}{-I}} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_2^{\top} & \mathbf{P}_4 \end{pmatrix}.$$

Note that

$$\mathbf{P}_1 = I - X_0 (X_0^{\top} X_0 + I_d)^{-1} X_0^{\top}.$$

It is easy to show that

$$\mathbf{P}_1 = (X_0 X_0^{\top} + I_{n+d})^{-1}.$$

Thus

$$\lambda_{\min}(\mathbf{P}_1) = \frac{1}{\lambda_{\max}(X_0 X_0^{\top} + I)} = \frac{1}{1 + \|X_0\|^2}.$$

Step 2. If m is sufficiently large, then

$$\left\|\sin\angle\left(\begin{pmatrix} A_m \\ B_m \end{pmatrix}, \begin{pmatrix} X_0 \\ -I \end{pmatrix}\right)\right\| < \frac{1}{\sqrt{1 + \|X_0\|^2}}. \tag{24}$$

For such m, we show that the matrix B_m is nonsingular. Assume the converse. Then there exists $f \in \mathbb{R}^d \setminus \{0\}$ such that $B_m f = 0$. For this f and for $u = A_m f$, we have

$$\begin{pmatrix} u \\ 0_{d\times 1} \end{pmatrix} = \begin{pmatrix} A_m f \\ B_m f \end{pmatrix} \in \operatorname{span}\left\langle \begin{pmatrix} A_m \\ B_m \end{pmatrix} \right\rangle.$$

Since the columns of the matrix $\binom{A_m}{B_m}$ are linearly independent, we conclude that

$$\begin{pmatrix} u \\ 0 \end{pmatrix} \ne 0.$$

Then (20) implies that

$$\left\|\sin\angle\left(\begin{pmatrix} A_m \\ B_m \end{pmatrix}, \begin{pmatrix} X_0 \\ -I \end{pmatrix}\right)\right\|^2 \ge \frac{\binom{u}{0}^{\top} P^{\perp}_{\binom{X_0}{-I}} \binom{u}{0}}{\left\|\binom{u}{0}\right\|^2} = \frac{u^{\top} \mathbf{P}_1 u}{\|u\|^2} \ge \lambda_{\min}(\mathbf{P}_1) = \frac{1}{1 + \|X_0\|^2},$$

which contradicts inequality (24). Therefore, if m is such that inequality (24) holds, then the matrix B_m is nonsingular.

Step 3. Let $\delta > 0$. We show that $\|A_m B_m^{-1} + X_0\| < \delta$ for sufficiently large m.

Indeed, if m is sufficiently large, then

$$\left\|\sin\angle\left(\begin{pmatrix}A_m\\B_m\end{pmatrix},\begin{pmatrix}X_0\\I\end{pmatrix}\right)\right\| < \frac{\delta}{\sqrt{1+\|X_0\|^2}\sqrt{1+(\|X_0\|+\delta)^2}} < \frac{1}{\sqrt{1+\|X_0\|^2}}. \tag{25}$$

We already proved in Step 2 that the matrix B_m is nonsingular for such a number m. Now we prove that $\|A_m B_m^{-1} + X_0\| < \delta$ for such m.

Indeed, there exists $f \in \mathbb{R}^d \setminus \{0\}$ such that $\|(A_m B_m^{-1} + X_0) f\| = \|A_m B_m^{-1} + X_0\|\,\|f\|$. Put

$$u = (A_m B_m^{-1} + X_0) f,$$
$$z = \begin{pmatrix}A_m\\B_m\end{pmatrix} B_m^{-1} f = \begin{pmatrix}A_m B_m^{-1} f\\ f\end{pmatrix} = \begin{pmatrix}u\\0\end{pmatrix} - \begin{pmatrix}X_0\\-I\end{pmatrix} f \in \operatorname{span}\left\langle\begin{pmatrix}A_m\\B_m\end{pmatrix}\right\rangle.$$

Since $(X_0^\top, -I) P^\perp_{\left(\begin{smallmatrix}X_0\\-I\end{smallmatrix}\right)} = 0$ and $P^\perp_{\left(\begin{smallmatrix}X_0\\-I\end{smallmatrix}\right)} \left(\begin{smallmatrix}X_0\\-I\end{smallmatrix}\right) = 0$, we have

$$\begin{aligned} z^\top P^\perp_{\left(\begin{smallmatrix}X_0\\-I\end{smallmatrix}\right)} z &= \left(\begin{pmatrix}u\\0\end{pmatrix} - \begin{pmatrix}X_0\\-I\end{pmatrix} f\right)^\top P^\perp_{\left(\begin{smallmatrix}X_0\\-I\end{smallmatrix}\right)} \left(\begin{pmatrix}u\\0\end{pmatrix} - \begin{pmatrix}X_0\\-I\end{pmatrix} f\right) \\ &= \begin{pmatrix}u\\0\end{pmatrix}^\top P^\perp_{\left(\begin{smallmatrix}X_0\\-I\end{smallmatrix}\right)} \begin{pmatrix}u\\0\end{pmatrix} = u^\top \mathbf{P}_1 u \ge \|u\|^2 \lambda_{\min}(\mathbf{P}_1) = \frac{\|A_m B_m^{-1} + X_0\|^2\,\|f\|^2}{1+\|X_0\|^2}. \end{aligned}$$

Since the columns of the matrix $\left(\begin{smallmatrix}A_m\\B_m\end{smallmatrix}\right)$ are linearly independent, we conclude that $z \ne 0$. Further

$$0 < \|z\|^2 = \|A_m B_m^{-1} f\|^2 + \|f\|^2 \le \left(1 + \|A_m B_m^{-1}\|^2\right)\|f\|^2.$$

From (20) we get

$$\left\|\sin\angle\left(\begin{pmatrix}A_m\\B_m\end{pmatrix},\begin{pmatrix}X_0\\-I\end{pmatrix}\right)\right\|^2 \ge \frac{z^\top P^\perp_{\left(\begin{smallmatrix}X_0\\-I\end{smallmatrix}\right)} z}{\|z\|^2} \ge \frac{\|A_m B_m^{-1} + X_0\|^2}{(1+\|X_0\|^2)(1+\|A_m B_m^{-1}\|^2)},$$

$$\left\|\sin\angle\left(\begin{pmatrix}A_m\\B_m\end{pmatrix},\begin{pmatrix}X_0\\-I\end{pmatrix}\right)\right\| \ge \frac{\|A_m B_m^{-1} + X_0\|}{\sqrt{1+\|X_0\|^2}\sqrt{1+(\|X_0\| + \|A_m B_m^{-1} + X_0\|)^2}}. \tag{26}$$

Since the function

$$\delta \mapsto \frac{\delta}{\sqrt{1+\|X_0\|^2}\sqrt{1+(\|X_0\|+\delta)^2}}$$

increases in $(0, +\infty)$, bounds (25) and (26) imply that $\|A_m B_m^{-1} + X_0\| < \delta$. □

Concluding remarks

Sufficient conditions for the consistency and strong consistency of the total least squares estimator are given in the paper for the vector linear regression errors-in-variables model. These conditions are weaker than those given in the paper [7]. We are able to drop assumption (8) describing a bound for the growth of the condition number of the matrix $A_0^\top A_0$. If the errors have finite moments of order $2+\epsilon$ and, generally speaking, an infinite fourth moment, then we found conditions for the strong consistency of the estimator. A typical case where these conditions hold is presented by

$$\lambda_{\min}(A_0^\top A_0) = O(1)m.$$

It is quite possible that the assumption $\operatorname{rank}(\Sigma X^0_{\mathrm{ext}}) = d$ can also be dropped. If it does not hold, then one needs to apply the theory of degenerate definite matrix pairs which results in complications of numerical procedures for the evaluation of the estimator. However, there is a hope that the total least squares estimator defined as a solution of the problem (3), (5) remains consistent even in this case.

In some forthcoming publications, we plan to find new conditions for the consistency of the total least squares estimator in heteroscedastic regression (for the so-called Elementwise-Weighted TLS Estimator, which is a total least squares estimator with elementwise weighting).

Acknowledgment

The results presented in this paper and their proofs have been discussed at the seminar "Asymptotic Methods in Statistics" in Kyiv Taras Shevchenko University. The author is grateful to the participants of the seminar, especially to O. G. Kukush, for remarks and continuous help.

Bibliography

1. V. V. Petrov, *Sums of Independent Random Variables*, Nauka, Moscow, 1972; English transl., Springer-Verlag, Berlin–New York–Heidelberg, 1975. MR0322927 (48:1288)
2. S. V. Shklyar, *An interpolation inequality for moments of sums of random vectors*, Teor. Imovir. Mat. Stat. **63** (2000), 156–162; English transl. in Theory Probab. Math. Statist. **63** (2001), 171–177. MR1870786 (2002m:60041)
3. W. Fuller, *Measurement Error Models*, Wiley, New York, 1987. MR898653 (89a:62160)
4. P. Gallo, *Consistency of regression estimates when some variables are subject to error*, Commun. Statist. Theor. Meth. **11** (1982), no. 9, 973–983. MR655466 (83h:62106)
5. L. J. Gleser, *Estimation in a multivariate "errors in variables" regression model: Large sample results*, Ann. Statist. **9** (1981), 24–44. MR600530 (82c:62096)
6. W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov, *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statistics, vol. 129, Springer, New York, 1998. MR1618204 (99f:42065)
7. A. Kukush and S. Van Huffel, *Consistency of element-weighted total least squares estimator in multivariate errors-in-variables model $AX = B$*, Metrika **59** (2004), no. 1, 75–97. MR2043433 (2004m:62129)
8. A. Kukush, I. Markovsky, and S. Van Huffel, *Consistency of structured total least squares estimator in a multivariate errors-in-variables model*, J. Statist. Planning Inference **133** (2005), 315–358. MR2194481 (2007b:62030)
9. G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, San Diego, 1990. MR1061154 (92a:65017)
10. P.-Å. Wedin, *Perturbation bounds in connection with singular value decomposition*, BIT **12** (1972), 99–111. MR0309968 (46:9071)

Department of Probability Theory, Statistics, and Actuarial Mathematics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue 2, Kiev 03127, Ukraine

E-mail address: shklyar@mail.univ.kiev.ua

Received 8/NOV/2010

Translated by N. SEMENOV