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

\[
\begin{align*}
    b_i &= a_i^0 X_0 + \tilde{b}_i, \\
    a_i &= a_i^0 + \tilde{a}_i.
\end{align*}
\]

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:

\[
A_0 X_0 = B_0, \quad A = A_0 + \tilde{A}, \quad B = B_0 + \tilde{B},
\]

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

$$A_0 \in \mathbb{R}^{m \times n}, \quad B_0 \in \mathbb{R}^{m \times d}.$$ 

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

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

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

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

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

Let $X_{0 \text{ext}} = (X_0 - I_d)^T$. Then

$$C_0 \in \mathbb{R}^{m \times (n + d)}, \quad X_{0 \text{ext}} \in \mathbb{R}^{(n + d) \times d}.$$ 

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

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

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

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

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

The sequence $\{(x_i, y_i), y = 1, \ldots, m\}$ is observed. One needs to estimate the parameters $\beta_0$ and $\beta_1$ from the observations.

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

$$\begin{align*}
\| \text{pinv}(\Sigma^{1/2}) \Delta^T \|_F &\rightarrow \min; \\
(I - P_\Sigma)\Delta^T &\rightarrow 0; \\
\text{rank}(C - \Delta) &\leq n.
\end{align*}$$

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

$$\begin{align*}
(C - \Delta)\hat{X}_{\text{ext}} &= 0, \\
(C - \Delta)(\hat{X} \ - I) &= 0.
\end{align*}$$
The columns of $\hat{X}_{\text{ext}}$ have to constitute a basis of the invariant subspace of $C^\top C$ with respect to $\Sigma$ that corresponds to $d$ minimal generalized eigenvalues,

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

Two real symmetric matrices $A$ and $B$ are called a definite pair if there exist two real numbers $\alpha$ and $\beta$ 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 \textup{[3]}.

a) The pair of matrices $\langle C^\top C, \Sigma \rangle$ is degenerate (the matrices $C^\top C$ and $\Sigma$ 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) $\text{rank} C < n$. Then problem \textup{[3]} has a trivial solution $\Delta = 0$, but the eigenspace of the matrix $C - \Delta$ corresponding to the eigenvalue 0 in \textup{[4]} is $(n+d-\text{rank} C)$-dimensional (its dimension is greater than $n$).

2) $\text{rank} C > n$. Then there is an infinite number of points of minimum $\Delta$ of problem \textup{[3]}.

c) The system of linear equations \textup{[5]} may have no solution for such a number $\Delta$ that gives the minimum to problem \textup{[3]}.

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

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 \textup{[5]} is inconsistent we use the following idea. Given a point of minimum $\Delta$ we find $\hat{X}_{\text{ext}}$ from equation \textup{[1]} such that the columns of $\hat{X}_{\text{ext}}$ are linearly independent (of course, the matrices $\Delta$ and $\hat{X}_{\text{ext}}$ should be Borel functions of observations (note that measurable solutions of problem \textup{[3]} exist); or, if we allow the use of “randomized estimators”, then $\Delta$ and $\hat{X}_{\text{ext}}$ should be random matrices).

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

$$\left\| \sin \angle(\hat{X}_{\text{ext}}, X^0_{\text{ext}}) \right\| \to 0, \quad m \to \infty.$$ 

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

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

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 $\hat{X}_{\text{ext}}$ is a nonsingular matrix; that is, there is a linear transform of its columns such that the matrix $\hat{X}_{\text{ext}}$ has the form of $\left( \begin{smallmatrix} \hat{x} \\ -I_d \end{smallmatrix} \right)$ provided that
\[
\left\| \sin \angle (\hat{X}_{\text{ext}}, X^0_{\text{ext}}) \right\| < \frac{1}{\sqrt{1 + \left\| X_0 \right\|^2}}
\]
and that the columns of the matrix $\hat{X}_{\text{ext}}$ are linearly independent.

**Proposition 2.1.** Suppose that conditions (G.2) and (G.3) hold. If the matrix $A_0^T A_0$ is positive definite, then $\langle C^T C, \Sigma \rangle$ is a definite pair almost surely. In other words,
\[
P(C^T C + \Sigma > 0) = 1.
\]

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

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 $v \in \mathbb{R}^{n+d-\text{rank}(\Sigma)} \setminus \{0\}$ such that
\[
\begin{align*}
[I_n & \quad X_0] Fv = 0, \\
F_1 v & = -X_0 F_2 v, \\
-Fv & = (X_0) F_2 v = X^0_{\text{ext}} F_2 v, \\
0 & = -\Sigma Fv = \Sigma X^0_{\text{ext}} \cdot F_2 v.
\end{align*}
\]

Moreover, $Fv \neq 0$, since $v \neq 0$ and the columns of $F$ are linearly independent, whence $F_2 v \neq 0$ according to (6).

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

3. Condition (G.2) implies that $CF = 0$ almost surely. Indeed, $E \hat{c}_i = 0$ and $\text{Var}[c_i F] = 0$ for all $i = 1, 2, \ldots, m$.

4. It remains to prove the implication
\[
"A_0^T A_0 > 0" \text{ and } "\hat{C} F = 0" \implies "C^T C + \Sigma > 0".
\]

The matrices $C^T C$ and $\Sigma$ are positive semi-definite. Assume that
\[
x^T (C^T C + \Sigma) x = 0.
\]

We will show that $x = 0$. If this is the case, then $C^T C + \Sigma > 0$. Equality (7) holds only if $Cx = 0$ and $\Sigma x = 0$. Thus $x \in \text{Ker} \Sigma$ and hence there exists a vector $v \in \mathbb{R}^{n+d-\text{rank} \Sigma}$ such that $x = Fv$. This implies that
\[
0 = A_0^T Cx = A_0^T (C_0 + \hat{C})x = A_0^T C_0 Fv + A_0^T \hat{C} Fv = A_0^T A_0 [I_n \quad X_0] Fv + 0.
\]

The matrix $A_0^T A_0$ 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 $A_0^T A_0 [I_n X_0] F$ are linearly independent, as well. Therefore $v = 0$ and $x = Fv = 0$. $\square$

3. Some known results concerning the consistency

**Theorem 3.1** (Galoo [4]). Let $d = 1$. Suppose that conditions (G.1)–(G.3) hold. We further assume that
\[
m^{-1/2} \lambda_{\min}(A_0^T A_0) \to \infty, \quad m \to \infty;
\]
\[
\frac{\lambda_{\min}^2(A_0^T A_0)}{\lambda_{\max}(A_0^T A_0)} \to \infty, \quad m \to \infty.
\]
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,
\[ \hat{X} \xrightarrow{P} X_0, \quad m \to \infty. \]

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

\[
\sup_{i \geq 1} \mathbb{E} \delta^4_{ij} < \infty; \\
m^{-1/2} \lambda_{\min}(A_0^T A_0) \to \infty, \quad m \to \infty; \\
\frac{\lambda^2_{\min}(A_0^T A_0)}{\lambda_{\max}(A_0^T A_0)} \to \infty, \quad m \to \infty.
\]

(8)

Then \( \hat{X} \xrightarrow{P} X_0 \) as \( m \to \infty. \)

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

\[
\sup_{i \geq 1} \mathbb{E} |\delta_{ij}|^{2r} < \infty; \\
\sum_{m=m_0}^{\infty} \left( \frac{m^{1/r}}{\lambda_{\min}(A_0^T A_0)} \right)^r < \infty; \\
\sum_{m=m_0}^{\infty} \left( \frac{\lambda_{\max}(A_0^T A_0)}{\lambda^2_{\min}(A_0^T A_0)} \right)^r < \infty
\]

for some \( r \geq 2 \) and \( m_0 \geq 1. \) Then the total least squares estimator is strongly consistent, that is,
\[ \hat{X} \xrightarrow{p_1} X_0, \quad m \to \infty. \]

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

\[
\sup_{i \geq 1} \mathbb{E} |\delta_{ij}|^{2r} < \infty; \\
m^{-1/r} \lambda_{\min}(A_0^T A_0) \to \infty, \quad m \to \infty; \\
\frac{\lambda^2_{\min}(A_0^T A_0)}{\lambda_{\max}(A_0^T A_0)} \to \infty, \quad m \to \infty,
\]

for some \( r \) such that \( 1 \leq r < 2. \) Then \( \hat{X} \xrightarrow{P} X_0 \) as \( m \to \infty. \)

4. Main results

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

\[
\sup_{i \geq 1} \mathbb{E} |\delta_{ij}|^{2r} < \infty, \\
m^{-1/r} \lambda_{\min}(A_0^T A_0) \to \infty, \quad m \to \infty,
\]

for some \( r \) such that \( 1 \leq r \leq 2. \) Then \( \hat{X} \xrightarrow{P} X_0 \) as \( m \to \infty. \)
Theorem 4.2. Let conditions (G.1)–(G.3) hold. Assume that
\[
\sup_{i \geq 1} \sum_{j=1, \ldots, n+d} \mathbb{E} |\delta_{ij}|^{2r} < \infty;
\]
for some \( r \geq 2 \) and \( m_0 \geq 1 \). Then \( \hat{X} \xrightarrow{\mathbb{P}_1} X_0 \) as \( m \to \infty \).

Theorem 4.3. Let conditions (G.1)–(G.3) hold. Assume that
\[
\sup_{i \geq 1} \sum_{j=1, \ldots, n+d} \mathbb{E} |\delta_{ij}|^{2r} < \infty;
\]
for some \( r \) such that \( 1 \leq r < 2 \) and \( m_0 \geq 1 \). Then \( \hat{X} \xrightarrow{\mathbb{P}_1} X_0 \) 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 \( C_0^T C_0 \). First, the matrix \( C_0^T C_0 \) is symmetric and nonnegative definite. Since \( C_0 X_0^0 = A_0 X_0 - B_0 = 0 \), this matrix has at least a \( d \)-dimensional eigenspace corresponding to the eigenvalue 0.

Next we prove that
\[
\lambda_{d+1}(C_0^T C_0) \geq \lambda_{\min}(A_0^T A_0).
\]
Recall that the eigenvalues are ordered in ascending order. Note also a more general inequality
\[
\lambda_{d+j}(C_0^T C_0) \geq \lambda_j(A_0^T A_0), \quad j = 1, \ldots, n.
\]
The latter inequality is a corollary of Theorem I.4.4 of [9], since
\[
\lambda_{d+1}(C_0^T C_0) = \sigma_2^2(C_0) \geq \sigma_2^2(A_0) = \lambda_{\min}(A_0^T A_0).
\]
The usual order is kept for singular values:
\[
\sigma_1(A_0) \geq \sigma_2(A_0) \geq \cdots.
\]
Inequality [9] implies that if the matrix \( A_0^T A_0 \) is nonsingular, then \( \lambda_{d+1}(C_0^T C_0) > 0 \), and thus \( \text{rank}(C_0^T C_0) = d \). The conditions of any of Theorems 4.1, 4.2, or 4.3 imply that
\[
\lambda_{\min}(A_0^T A_0) \to \infty.
\]
Thus if the assumptions of at least one of these theorems hold, then
\[
\lambda_{d+1}(C_0^T C_0) \geq \lambda_{\min}(A_0^T A_0) > 0
\]
for sufficiently large \( m \).

2. Spectral decomposition of the matrices \( C_0^T C_0 \), \( N \), and \( N^{-1/2} C_0^T C_0 N^{-1/2} \). Below is the canonical decomposition of the matrix \( C_0^T C_0 \):
\[
C_0^T C_0 = U \text{ diag } (\lambda_{\min}(C_0^T C_0), \lambda_1(C_0^T C_0), \ldots, \lambda_\max(C_0^T C_0)) U^T
= U \text{ diag } (\lambda_j(C_0^T C_0))_{j=1}^{n+d} U^T,
\]
where \( U \) is an orthogonal matrix \((U^{-1} = U^T)\) and
\[
\lambda_j(C_0^T C_0) = 0, \quad j = 1, \ldots, d.
\]
In particular, $\lambda_{\min}(C_0^T C_0) = 0$. Put

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

Then the canonical decomposition of the matrix $N$ is given by

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

Note that

$$\lambda_{\min}(N) = \cdots = \lambda_d(N) = \lambda_{\min}(A_0^T A_0).$$  

If the matrix $A_0^T A_0$ is nonsingular, then the canonical decomposition of 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 $C_0 X_0^\text{ext} = 0$, we get

$$NX_0^\text{ext} = \lambda_{\min}(A_0^T A_0)X_0^\text{ext}.$$  

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

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

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

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

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

$$\lambda_j \left( N^{-1/2} C_0^T C_0 N^{-1/2} \right) = 0, \quad j = 1, \ldots, d;$$  

$$\frac{1}{2} \leq \lambda_j \left( N^{-1/2} C_0^T C_0 N^{-1/2} \right) \leq 1, \quad j = d + 1, \ldots, n + d.$$  

3. An upper bound for $\| \sin \angle(\hat{X}_\text{ext}, X_0^\text{ext}) \|$. Equality (10) implies that

$$\hat{X}_\text{ext}^T N \hat{X}_\text{ext} \geq \lambda_{\min}(A_0^T A_0) \hat{X}_\text{ext}^T \hat{X}_\text{ext}$$  

and

$$\frac{v^T \hat{X}_\text{ext}^T X_0^\text{ext} (X_0^\text{ext} X_0^\text{ext})^{-1} X_0^\text{ext}^T \hat{X}_\text{ext} v}{v^T \hat{X}_\text{ext}^T \hat{X}_\text{ext} v} \geq \lambda_{\min}(A_0^T A_0) \frac{v^T \hat{X}_\text{ext}^T X_0^\text{ext} (X_0^\text{ext} X_0^\text{ext})^{-1} X_0^\text{ext}^T \hat{X}_\text{ext} v}{v^T \hat{X}_\text{ext}^T N \hat{X}_\text{ext} v}$$  

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

$$\frac{v^T \hat{X}_\text{ext}^T X_0^\text{ext} (X_0^\text{ext} X_0^\text{ext})^{-1} X_0^\text{ext}^T \hat{X}_\text{ext} v}{v^T \hat{X}_\text{ext}^T \hat{X}_\text{ext} v} \geq \frac{v^T N \hat{X}_\text{ext}^T X_0^\text{ext} (X_0^\text{ext} N X_0^\text{ext})^{-1} X_0^\text{ext}^T N \hat{X}_\text{ext} v}{v^T \hat{X}_\text{ext}^T N \hat{X}_\text{ext} v}.$$
Using (22) we prove the inequality
\[
1 - \left\| \sin \angle (\tilde{X}_{\text{ext}}, X^0_{\text{ext}}) \right\|^2 \geq 1 - \left\| \sin \angle (N^{1/2} \tilde{X}_{\text{ext}}, N^{1/2} X^0_{\text{ext}}) \right\|^2 ,
\]
(15)
\[
\left\| \sin \angle (\tilde{X}_{\text{ext}}, X^0_{\text{ext}}) \right\| \leq \left\| \sin \angle (N^{1/2} \tilde{X}_{\text{ext}}, N^{1/2} X^0_{\text{ext}}) \right\| .
\]

Put
\[
\epsilon = \left\| N^{-1/2} (C^\top C - C^\top_0 C_0 - m \Sigma) N^{-1/2} \right\| .
\]

Now we use Corollary 5.5 of Lemma 5.4 on the stability of the eigenspace. Let \( N^{-1/2} C_0^\top C_0 N^{-1/2} \) be the nonperturbed matrix whose null space is the column space of \( N^{1/2} X^0_{\text{ext}} \). 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 \( N^{1/2} \tilde{X}_{\text{ext}} \). Using (13)–(14) we establish
\[
\left\| \sin \angle (N^{1/2} \tilde{X}_{\text{ext}}, N^{1/2} X^0_{\text{ext}}) \right\|^2 \leq \frac{\epsilon}{0.5} \left( 1 + \left\| N^{-1/2} \Sigma N^{-1/2} \right\| \lambda_{\text{max}} \left( (X^0_{\text{ext}} \Sigma X^0_{\text{ext}})^{-1} X^0_{\text{ext}} \top N X^0_{\text{ext}} \right) \right) .
\]
Taking into account equalities (11) and (12) and inequality (15) we obtain
\[
\left\| \sin \angle (\tilde{X}_{\text{ext}}, X^0_{\text{ext}}) \right\|^2 \leq \left\| \sin \angle (N^{1/2} \tilde{X}_{\text{ext}}, N^{1/2} X^0_{\text{ext}}) \right\|^2 \leq 2 \epsilon \left( 1 + \frac{\| \Sigma \|}{\lambda_{\text{min}} (A_0^\top A_0)} \lambda_{\text{max}} \left( \lambda_{\text{min}} (A_0^\top A_0) (X^0_{\text{ext}} \Sigma X^0_{\text{ext}})^{-1} X^0_{\text{ext}} \top X^0_{\text{ext}} \right) \right) = 2 \epsilon \left( 1 + \| \Sigma \| \lambda_{\text{max}} \left( (X^0_{\text{ext}} \Sigma X^0_{\text{ext}})^{-1} X^0_{\text{ext}} \top X^0_{\text{ext}} \right) \right) .
\]

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 \( \| \sin \angle (\tilde{X}_{\text{ext}}, X^0_{\text{ext}}) \| \to 0 \) and thus the matrix \( \tilde{X}_{\text{ext}} \) is transformed to the form \( (\tilde{X} - I) \) by transforming its columns for sufficiently large \( m \) (this is justified by Lemma 5.8). Moreover, \( \tilde{X} \to X_0 \) in this case.

We introduce two \((n + d) \times (n + d)\) matrices:
\[
M_1 = N^{-1/2} C_0^\top C N^{-1/2} ,
M_2 = N^{-1/2} (\tilde{C}^\top \tilde{C} - m \Sigma) N^{-1/2} .
\]

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
\[
\| M_1 \|_F^2 = \left\| N^{-1/2} C_0^\top C N^{-1/2} \right\|_F^2 = \text{tr} \left( N^{-1/2} C_0^\top C N^{-1/2} \right) = \text{tr} \left( C_0 N^{-1/2} C_0^\top C N^{-1/2} \right) = \sum_{i=1}^m \sum_{j=1}^m c_{ij} N^{-1/2} c_{ij}^\top N^{-1/2} c_{ij}^\top .
\]
Since \( \mathbb{E} c_i N^{-1} c_i^\top = 0 \) for \( i \neq j \) and \( \mathbb{E} c_i N^{-1} c_i^\top = \text{tr} (\Sigma N^{-1}) \), we get
\[
\mathbb{E} \| M_1 \|_F^2 = \sum_{i=1}^m c_{ii} N^{-1} c_{ii}^\top \text{tr} (\Sigma N^{-1}) = \text{tr} (C_0 N^{-1}) \text{tr} (\Sigma N^{-1}) .
\]
Now we estimate the factors on the right hand side of the latter equality:
\[
\text{tr} \left( C_0 N^{-1} C_0^\top \right) = \text{tr} \left( C_0^\top C_0 N^{-1} \right) = \sum_{i=1}^{n+d} \lambda_i \left( C_0^\top C_0 N^{-1} \right).
\]
Since \( \lambda_i(C_0^\top C_0 N^{-1}) = 0 \) for \( i \leq d \) and \( \frac{1}{2} \leq \lambda_i(C_0^\top C_0 N^{-1}) \leq 1 \) for \( d < i \leq d + n \), we conclude that
\[
\frac{n}{2} \leq \text{tr} \left( C_0 N^{-1} C_0^\top \right) \leq n.
\]
Since both matrices \( N^{-1} \) and \( \Sigma \) are positive semi-definite,
\[
\text{tr}(\Sigma N^{-1}) = \left\| N^{-1/2} \Sigma^{1/2} \right\|_F^2 \leq \left\| N^{-1/2} \right\|^2 \left\| \Sigma^{1/2} \right\|_F^2 = \lambda_{\max} \left( N^{-1} \right) \text{tr} \Sigma
\]
\[
= \lambda_{\min} \left( A_0^\top A_0 \right) \text{tr} \Sigma.
\]
Finally,
\[
\mathbb{E} \left\| M_1 \right\|^2_F \leq \frac{n \text{tr} \Sigma}{\lambda_{\min}(A_0^\top A_0)}.
\]

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

Now we prove the convergence \( M_2 \xrightarrow{P} 0 \) as \( m \to \infty \):
\[
M_2 = N^{-1/2}(\tilde{C}^\top \tilde{C} - m \Sigma)N^{-1/2},
\]
\[
\| M_2 \| \leq \| N^{-1/2} \| \left\| \tilde{C}^\top \tilde{C} - m \Sigma \right\| \| N^{-1/2} \| = \frac{\| \sum_{i=1}^{m} (\tilde{c}_i^\top \tilde{c}_i - \Sigma) \|}{\lambda_{\min}(A_0^\top A_0)}.
\]

To estimate \( \| M_2 \|^r \) we use the Rosenthal inequality for \( 1 \leq \nu \leq 2 \) (see Theorem 5.7 below):
\[
\mathbb{E} \left\| M_2 \right\|^r \leq \frac{\text{const} \sum_{i=1}^{m} \mathbb{E} \left\| \tilde{c}_i^\top \tilde{c}_i - \Sigma \right\|^r}{\lambda_{\min}(A_0^\top A_0)}.
\]

According to the assumptions of Theorem 4.1 the sequence \( \{ \mathbb{E} \left\| \tilde{c}_i^\top \tilde{c}_i - \Sigma \right\|^r, i = 1, 2, \ldots \} \) is bounded, whence
\[
\mathbb{E} \left\| M_2 \right\|^r \leq \frac{O(m)}{\lambda_{\min}(A_0^\top A_0)}, \quad m \to \infty.
\]

Thus \( \mathbb{E} \left\| M_2 \right\|^r \to 0 \) and \( M_2 \xrightarrow{P} 0 \) as \( m \to \infty \).

Rest of the proof of Theorem 4.2 We have
\[
M_1 = \sum_{i=1}^{m} N^{-1/2}c_0_i \tilde{c}_i N^{-1/2}.
\]
By the Rosenthal inequality,
\[
\mathbb{E} \left\| M_1 \right\|^{2r} \leq \text{const} \sum_{i=1}^{m} \mathbb{E} \left\| N^{-1/2}c_0_i \tilde{c}_i N^{-1/2} \right\|^{2r}
\]
\[
+ \text{const} \left( \sum_{i=1}^{m} \mathbb{E} \left\| N^{-1/2}c_0_i \tilde{c}_i N^{-1/2} \right\|^{2} \right)^{r}.
\]
The first term is estimated as follows:
\[
\sum_{i=1}^{m} \mathbb{E} \left\| N^{-1/2}c_0_i \tilde{c}_i N^{-1/2} \right\|^{2r} \leq \sum_{i=1}^{m} \left\| N^{-1/2}c_0_i \right\|^{2r} \max_{i=1, \ldots, m} \mathbb{E} \left\| \tilde{c}_i \right\|^{2r} \left\| N^{-1/2} \right\|^{2r}.
\]
Thus
\[
\sum_{i=1}^{m} \left\| N^{-1/2} c_{0i} \right\|^{2r} \leq \left( \sum_{i=1}^{m} \left\| N^{-1/2} c_{0i} \right\|^{2} \right)^{r} = \left( \sum_{i=1}^{m} c_{0i} N^{-1} c_{0i}^{\top} \right)^{r} = (\text{tr}(C_{0}N^{-1}C_{0}^{\top}))^{r} \leq n^{r}
\]
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
\[
\{ \text{max }_{i=1,\ldots,m} E \left\| \tilde{c}_{i} \right\|^{2r}, m = 1, 2, \ldots \}
\]
is bounded. We recall that \( \left\| N^{-1/2} \right\| = \lambda_{\text{min}}^{-1/2}(A_{0}^{\top} A_{0}) \). Thus
\[
\sum_{i=1}^{m} E \left\| N^{-1/2} c_{0i}^{\top} \tilde{c}_{i} N^{-1/2} \right\|^{2r} = \frac{O(1)}{\lambda_{\text{min}}^{r}(A_{0}^{\top} A_{0})}, \quad m \to \infty.
\]
The limit relation
\[
\sum_{i=1}^{m} E \left\| N^{-1/2} c_{0i}^{\top} \tilde{c}_{i} N^{-1/2} \right\|^{2} = \frac{O(1)}{\lambda_{\text{min}}(A_{0}^{\top} A_{0})}
\]
is proved similarly. It is important for the proof of this relation that the sequence
\[
\{ \text{max }_{i=1,\ldots,m} E \left\| \tilde{c}_{i} \right\|^{2}, m = 1, 2, \ldots \}
\]
is bounded by the conditions of Theorem 4.2.

Finally,
\[
E \left\| M_{1} \right\|^{2r} = \frac{O(1)}{\lambda_{\text{min}}^{r}(A_{0}^{\top} A_{0})}, \quad m \to \infty.
\]

Hence the conditions of Theorem 4.2 imply that \( \sum_{m=m_{0}}^{\infty} E \left\| M_{1} \right\|^{2r} < \infty \), whence \( M_{1} \overset{p^{1}}{\to} 0 \) as \( m \to \infty \).

Now we prove the convergence \( M_{2} \overset{p^{1}}{\to} 0 \). To estimate \( E \left\| M_{2} \right\|^{r} \), we apply Rosenthal’s inequality (also see (17)):
\[
E \left\| M_{2} \right\|^{r} \leq \frac{E \left\| \sum_{i=1}^{m} (\tilde{c}_{i}^{\top} \tilde{c}_{i} - \Sigma) \right\|^{r}}{\lambda_{\text{min}}^{r}(A_{0}^{\top} A_{0})} + \frac{\text{const} \sum_{i=1}^{m} E \left\| \tilde{c}_{i}^{\top} \tilde{c}_{i} - \Sigma \right\|^{r}}{\lambda_{\text{min}}(A_{0}^{\top} A_{0})} + \frac{\text{const} \left( \sum_{i=1}^{m} E \left\| \tilde{c}_{i}^{\top} \tilde{c}_{i} - \Sigma \right\|^{2} \right)^{r/2}}{\lambda_{\text{min}}^{r}(A_{0}^{\top} A_{0})}.
\]
The sequences
\[
\{ E \left\| \tilde{c}_{i}^{\top} \tilde{c}_{i} - \Sigma \right\|^{r}, i = 1, 2, \ldots \} \quad \text{and} \quad \{ E \left\| \tilde{c}_{i}^{\top} \tilde{c}_{i} - \Sigma \right\|^{2}, i = 1, 2, \ldots \}
\]
are bounded by the assumptions of Theorem 4.2. Thus
\[
E \left\| M_{2} \right\|^{r} = \frac{O(m^{r/2})}{\lambda_{\text{min}}(A_{0}^{\top} A_{0})}, \quad m \to \infty;
\]
\[
\sum_{m=m_{0}}^{\infty} E \left\| M_{2} \right\|^{r} < \infty
\]
for some \( m_{0} \geq 1 \), whence we conclude that \( M_{2} \overset{p^{1}}{\to} 0 \) as \( m \to \infty \). \( \square \)
Rest of the proof of Theorem 4.3] The relation
\[ E\|M_1\|^{2r} = \frac{O(1)}{\lambda_{\min}^r(A_0^t A_0)}, \quad m \to \infty, \]
is proved in the same way as in the proof of Theorem 4.2 where we dealt with the case of \( r \geq 1 \). As in the proof of Theorem 4.2 we show that \( M_1 \overset{P_1}{\longrightarrow} 0 \) as \( m \to \infty \).

Now we prove the convergence \( M_2 \overset{P_1}{\longrightarrow} 0 \) as \( m \to \infty \). By the assumptions of Theorem 4.3,
\[ \sum_{m=1}^{\infty} E\|\tilde{c}_m^T \tilde{c}_m - \Sigma\|^r \overset{\lambda_{\min}^r(A_0^t A_0)}{\lambda} < \infty. \]
The random matrices \( \tilde{c}_i^T \tilde{c}_i - \Sigma \) have zero expectations. The sequence of nonnegative numbers \( \{\lambda_{\min}(A_0^t A_0), m = 1, 2, \ldots\} \) goes to \( +\infty \) and is nondecreasing, since \( A_0^t A_0 \) is nondecreasing in Loewner’s order. By the strong law of large numbers [1, Theorem IX.12],
\[ \frac{1}{\lambda_{\min}(A_0^t A_0)} \sum_{i=1}^{m} (\tilde{c}_i^T \tilde{c}_i - \Sigma) \overset{P_1}{\longrightarrow} 0, \quad m \to \infty. \]
Recalling (17), we get
\[ \|M_2\| \leq \frac{\|\sum_{i=1}^{m} (\tilde{c}_i^T \tilde{c}_i - \Sigma)\|}{\lambda_{\min}(A_0^t A_0)} \overset{P_1}{\longrightarrow} 0, \quad m \to \infty, \]
\[ M_2 \overset{P_1}{\longrightarrow} 0, \quad m \to \infty. \]

□

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 \( \alpha \) and \( \beta \). Then there exist a nonsingular matrix \( T \) and diagonal matrices \( \Lambda \) and \( M \) such that
\[ A = (T_1^{-1})^T \Lambda T_1^{-1}, \quad B = (T_1^{-1})^T M T_1^{-1}. \]

Let
\[ T_1 = [u_1, u_2, \ldots, u_n], \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n), \quad M = \text{diag}(\mu_1, \ldots, \mu_n) \]
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 \( \lambda_1/\mu_1 \leq \lambda_2/\mu_2 \leq \cdots \leq \lambda_{\text{rank } B}/\mu_{\text{rank } B} \). Then
\[ \frac{\lambda_i}{\mu_i} = \min_{\dim V = i} \max_{v \in V \setminus \{0\}} \frac{v^T A v}{v^T B v}, \quad i = 1, 2, \ldots, \text{rank } B. \]
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,
\[ \lambda_{\min}(A) = 0, \quad \lambda_2(A) > 0, \quad \lambda_{\min}(B) \geq 0. \]
Let $Ax_0 = 0$ and $Bx_0 \neq 0$. Let $x \neq 0$ be either a point of minimum of the function
\[ f(x) := \frac{x^\top (A + \tilde{A}) x}{x^\top B x} \]
or a point such that
\[ \liminf_{t \to x} f(t) = \inf_{t: t^\top B t > 0} f(t). \]
(18)\] Note that such a point $x$ exists by assumptions of the lemma. Then
\[ \sin^2 \angle(x, x_0) \leq \frac{\|\tilde{A}\|}{\lambda_2(A)} \left(1 + \frac{\|x_0\|^2}{x_0^\top B x_0} \frac{x^\top B x}{\|x\|^2}\right). \]

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]:
\[ \|\sin \angle(V_1, V_2)\| = \|P_{V_1}(I - P_{V_2})\|. \]
(19)\] Recall that $\| \cdot \|$ is the operator norm of a matrix. Applying equality (19), one can prove that
\[ \|\sin \angle(V_1, V_2)\|_2^2 = \min_{v \in V_1 \setminus \{0\}} \frac{v^\top P_{V_2} v}{\|v\|^2}. \]
(20)\] If the columns of the matrix $X$ form a basis of the subspace $V_1$, then
\[ \|\sin \angle(X, V_2)\|_2^2 := \|\sin \angle(V_1, V_2)\|_2^2 = \lambda_{\max} \left((X^\top X)^{-1} X^\top P_{V_2} X\right), \]
(21)\]
\[ 1 - \|\sin \angle(X, V_2)\|_2^2 = \lambda_{\min} \left((X^\top X)^{-1} X^\top P_{V_2} X\right). \]
(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 $\lambda_{\min}(B) \geq 0$. 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 rank $X_0 = d$).

Let the functional
\[ f(X) = \lambda_{\max} \left((X^\top BX)^{-1} X^\top (A + \tilde{A}) X\right), \quad X \in \mathbb{R}^{n \times d}, \ X^\top BX > 0, \]
attain its minimum. Then, for all points of minimum $X$,
\[ \|\sin \angle(X, X_0)\|_2^2 \leq \frac{\|\tilde{A}\|}{\lambda_{d+1}(A)} \left(1 + \|B\| \lambda_{\max} \left((X_0^\top BX_0)^{-1} X_0^\top X_0\right)\right). \]

**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
\[ \lambda_{\max}(B^{-1}A) = \max_{v \neq 0} \frac{v^\top A v}{v^\top B v} =: \max \frac{A}{B}. \]
Let $X$ be a point of minimum of the functional $f(X)$. Since the functional $f(X)$ is defined at the point $X_0$, $f(X) \leq f(X_0)$, that is,
\[
\max \frac{X^\top (A + \tilde{A})X}{X^\top BX} \leq \max \frac{X_0^\top (A + \tilde{A})X_0}{X_0^\top BX_0}.
\]

Since
\[
X^\top \tilde{A}X \geq -\|\tilde{A}\|X^\top X, \\
X^\top BX \leq \|B\|X^\top X, \\
X_0^\top \tilde{A}X_0 \leq \|\tilde{A}\|X_0^\top X_0, \\
AX_0 = 0,
\]
we obtain
\[
\max \frac{X^\top AX - \|\tilde{A}\|X^\top X}{\|B\|X^\top X} \leq \max \frac{\|\tilde{A}\|X_0^\top X_0}{X_0^\top BX_0},
\]
\[
(23) \quad \frac{1}{\|B\|} \cdot \left( \max \frac{X^\top AX}{X^\top X} - \|\tilde{A}\| \right) \leq \|\tilde{A}\| \max \frac{X_0^\top X_0}{X_0^\top BX_0}.
\]

Equality (21) implies that
\[
\lambda_{d+1}(A) \|\sin \angle(X, X_0)\|^2 \leq \max \frac{X^\top AX}{X^\top X},
\]
since $A \geq \lambda_{d+1}(A)P_{X_0}^\perp$ in the sense of Loewner’s order. Thus (23) implies that
\[
\|\sin \angle(X, X_0)\|^2 \leq \frac{\|\tilde{A}\|}{\lambda_{d+1}(A)} \left( 1 + \|B\| \max \frac{X_0^\top X_0}{X_0^\top BX_0} \right)
\]
and this is what was to be proved. \(\square\)

**Corollary 5.5.** Let $A$, $B$, and $\tilde{A}$ be symmetric $n \times n$ matrices and let $\lambda_i(A) = 0$ for all $i = 1, \ldots, d$ (in particular, $\lambda_{\min}(A) = 0$), $\lambda_{d+1}(A) > 0$, and $\lambda_{\min}(B) \geq 0$. 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 $X_0^\top BX_0 > 0$, rank $X_0 = d$, and rank $B \geq d$).

Let there exist a scalar $k > 0$ such that the matrix $A + \tilde{A} + kB$ is positive definite (whence one concludes that $(A + \tilde{A}, B)$ 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
\[
\|\sin \angle(V_1, X_0)\|^2 \leq \frac{\|\tilde{A}\|}{\lambda_{d+1}(A)} \left( 1 + \|B\| \lambda_{\max}\left((X_0^\top BX_0)^{-1}X_0^\top X_0\right)\right).
\]

**5.3. Rosenthal inequality.**

**Theorem 5.6.** Let $\nu \geq 2$ be a nonrandom number. Then there are $\alpha \geq 0$ and $\beta \geq 0$ such that
\[
E \left[ \sum_{i=1}^m \xi_i \right] \leq \alpha \sum_{i=1}^m E [\xi_i]^{\nu} + \beta \left( \sum_{i=1}^m E \xi_i^2 \right)^{\nu/2}
\]
if the random variables $\{\xi_i, i = 1, \ldots, m\}$, $m \geq 1$, are independent and have zero expectations, that is, $E \xi_i = 0$ for all $i = 1, \ldots, m$.

A proof of this result can be found in [6].
Theorem 5.7. Let $\nu$ be a nonrandom number such that $1 \leq \nu \leq 2$. Then there exists a nonrandom number $\alpha \geq 0$ such that

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

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

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 $\| \sin \angle (\hat{X}_{\text{ext}}, X_0^{\text{ext}}) \|$ and $\| \hat{X} - X_0 \|$.

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

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

then

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

Proof. Step 1. Throughout the proof, $P_1$ denotes an $n \times n$ submatrix of the matrix $P_{(X_0)}^{\perp}$:

$$I - P_{(X_0)} = P_{(X_0)}^{\perp} = \begin{pmatrix} P_1 & P_2 \\ P_3 & P_4 \end{pmatrix}.$$

Note that

$$P_1 = I - X_0(X_0^\top X_0 + I_d)^{-1}X_0^\top.$$

It is easy to show that

$$P_1 = (X_0X_0^\top + I_{n+d})^{-1}.$$

Thus

$$\lambda_{\min}(P_1) = \frac{1}{\lambda_{\max}(X_0X_0^\top + I)} = \frac{1}{1 + \|X_0\|^2}.$$

Step 2. If $m$ is sufficiently large, then

$$\left\| \sin \angle \left( \left( \begin{array}{c} A_m \\ B_m \end{array} \right), \left( \begin{array}{c} X_0 \\ -I \end{array} \right) \right) \right\| \leq \frac{1}{\sqrt{1 + \|X_0\|^2}}.$$

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

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

Since the columns of the matrix $\left( \begin{array}{c} A_m \\ B_m \end{array} \right)$ are linearly independent, we conclude that

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

Then (20) implies that

$$\left\| \sin \angle \left( \left( \begin{array}{c} A_m \\ B_m \end{array} \right), \left( \begin{array}{c} X_0 \\ -I \end{array} \right) \right) \right\|^2 \geq \frac{\left( \begin{pmatrix} u \\ 0 \end{pmatrix} \right)^\top P_{(X_0)}^{\perp} \left( \begin{pmatrix} u \\ 0 \end{pmatrix} \right)}{\| \left( \begin{pmatrix} u \\ 0 \end{pmatrix} \right) \|^2} = \frac{u^\top P_1 u}{\|u\|^2} \geq \lambda_{\min}(P_1) = \frac{1}{1 + \|X_0\|^2},$$
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 \( \|A_mB_m^{-1} + X_0\| < \delta \) for sufficiently large \( m \).

Indeed, if \( m \) is sufficiently large, then

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

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

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

Put

\[
u = (A_mB_m^{-1} + X_0)f,\]

\[
z = \begin{pmatrix} A_m \\ B_m \end{pmatrix} B_m^{-1} f = \begin{pmatrix} A_m B_m^{-1} f \\ f \end{pmatrix} = \begin{pmatrix} u \\ 0 \end{pmatrix} - \begin{pmatrix} X_0 \\ -I \end{pmatrix} f \in \text{span} \left\{ \begin{pmatrix} A_m \\ B_m \end{pmatrix} \right\}.
\]

Since \( (X_0^T, -I)P_{(X_0)}^⊥ = 0 \) and \( P_{(X_0)}^⊥ (X_0^T - I) = 0 \), we have

\[
z^T P_{(X_0)}^⊥ z = \begin{pmatrix} u \\ 0 \end{pmatrix}^T \begin{pmatrix} u \\ 0 \end{pmatrix} - (X_0 - I) f \begin{pmatrix} u \\ 0 \end{pmatrix} = u^T P_{1} u \geq \|u\|^2 \lambda_{\min}(P_{1}) = \frac{\|A_mB_m^{-1} + X_0\|^2 \|f\|^2}{1 + \|X_0\|^2}.
\]

Since the columns of the matrix \( \begin{pmatrix} A_m \\ B_m \end{pmatrix} \) are linearly independent, we conclude that \( z \neq 0 \). Further

\[
0 < \|z\|^2 = \|A_mB_m^{-1} f\|^2 + \|f\|^2 \leq (1 + \|A_mB_m^{-1}\|^2) \|f\|^2.
\]

From (20) we get

\[
\left\| \sin \angle \left( \begin{pmatrix} A_m \\ B_m \end{pmatrix}, \begin{pmatrix} X_0 \\ -I \end{pmatrix} \right) \right\|^2 \geq \frac{z^T P_{(X_0)}^⊥ z}{\|z\|^2} \geq \frac{\|A_mB_m^{-1} + X_0\|^2}{(1 + \|X_0\|^2)(1 + \|A_mB_m^{-1}\|^2)},
\]

(26)  \[
\left\| \sin \angle \left( \begin{pmatrix} A_m \\ B_m \end{pmatrix}, \begin{pmatrix} X_0 \\ -I \end{pmatrix} \right) \right\| \geq \frac{\|A_mB_m^{-1} + X_0\|}{\sqrt{1 + \|X_0\|^2} \sqrt{1 + (\|X_0\| + \|A_mB_m^{-1} + X_0\|)^2}}.
\]

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 \( \|A_mB_m^{-1} + X_0\| < \delta \).

\[
\square
\]

**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 \( A_0^T A_0 \). 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

\[
\lambda_{\min}(A_0^T A_0) = O(1)m.
\]

It is quite possible that the assumption \( \text{rank}(\Sigma X_0^{\text{ext}}) = d \) 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 \([4, 5]\) 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


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