Yugoslav Journal of Operations Research
1 (1991), Number 2, 195-210

# APPROACHES TO IDENTIFICATION OF LINEAR RELATIONS FROM COMPOUND NOISY AND NOISE-FREE DATA

Dragan RADOJEVIĆ

*Knowledge Engineering Laboratory, Mihajlo Pupin Institute, P.O. Box 15, 11000 Belgrade, Yugoslavia*

**Abstract.** Frisch scheme defines the whole set of possible solutions of linear relations for numerical series in general case. This paper presents a modification to the Frisch scheme made for the following special case: some series of numbers are noiseless — independent variables; and the remaining are noisy — dependent variables. Special results from the set defined by modified Frisch scheme, which represent potential solution for prediction are analysed. The result obtained on unweighted principal components of the matrix of covariance of explained part of variance of dependent variables, represent the best solution for prediction.

*Key words and phrases*: Frisch scheme, estimation, singular value decomposition, principal components, canonical correlation

## 1. INTRODUCTION

The solution of the problem of identifying the linear relations on the basis of a finite data set in the presence of noise is not unique, which makes the problem conceptually very complex. The existing approaches (principal components analysis; factor analysis; linear regressions, etc.) lead to unique solutions based on the assumptions involved (regression analysis: noise is allocated to a single (dependent) variable; orthogonal regression: errors are evenly distributed to all variables, etc.). Determining the set of all possible solutions in both the solution space and error space has been in the focus of research attention and is mainly based on the so called Frisch scheme [**1**]. In this paper particular solutions in general Frisch scheme, which are of interest for prediction, are considered.

This paper provides an analysis of identifying the linear relations between series of numbers in the following special case: some series are exact — independent variables; and some are in the presence of noise — dependent variables. The set of all these possible solutions of interest is defined by modified Frisch scheme and is actually a subset determined by original Frisch scheme.

The stated modifications to the general Frisch scheme have been made in order to permit the implementation of a new procedure for modelling the dynamic processes described by time series in case when the time series are with no observation error and the complete error is taken to result from the model, so the model selection procedure reduces to the prediction error minimization. In this paper are considered several special solutions from the set defined by modified Frisch scheme, with intention to determine the best solution for prediction.

Section 2 of the paper gives a survey of known results relating to the Frisch scheme. In Section 3 particular solutions from general Frisch scheme are analyzed, which are of interest for estimation of random vectors. Section 4 defines the modified Frisch scheme. In Section 5 particular solutions from modified Frisch scheme of interest for prediction are analyzed.

## 2. FRISCH SCHEME

Determining the linear relations between series of numbers, in a general case, is not unique, because an approximation is involved. The solution to be obtained depends on a desired property of approximation (formalized through a criterion function accepted). The entire possible set of solutions is also of interest. We present here the known scheme that should be satisfied by all potential solutions [1].

### 2.1. Equivalence of Linear Relations

The problem of finding linear dependencies between observed series is equivalent to the problem of finding the linear dependence between the columns of the covariance matrix of analyzed series [1]. The following notation will be used here: $X$ an $N \times n$ matrix of $N$ observations of $n$ zero-mean series, $\Sigma$ an $n \times n$ matrix of covariance of series $X$.

### 2.2. Fundamental Assumptions (Frisch Scheme)

The basic assumptions to be satisfied by the analyzed sets $X$, i.e. by their covariance matrix, are:

Assumption 1
$$\operatorname{cor}(\Sigma) = 0. \tag{2.1}$$
(where:. cor = corang; $\operatorname{cor}(\Sigma) = n - \operatorname{rang}(\Sigma)$).

It says that, because of the presence of noise, there exist no trivial linear solutions.

Assumption 2: *Noise is additive.*

$$X = \hat{X} + \mathcal{E} \tag{2.2}$$

where: $\hat{X}$ is the matrix $X$ after noise elimination and $\mathcal{E}$ a noise matrix. All series of numbers are assumed to be noisy.

Assumption 3: *The columns of the noise matrix $\mathcal{E}$ are independent of the columns of the exact matrix $\hat{X}$.*

As a consequence of Assumptions 2 and 3:

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} \tag{2.3}$$

where $\hat{\Sigma}$ is a covariance matrix of series $\hat{X}$ and $\tilde{\Sigma}$ a covariance matrix of noise $\mathcal{E}$.

ASSUMPTION 4: *The noise covariance matrix, $\tilde{\Sigma}$, is a nonnegative definite (NND) diagonal matrix.*

White noise is assumed.

### 2.3. PROBLEM STATEMENT

The problem of finding all possible solutions that satisfy the general Frisch scheme can be defined in the noise space and solution space [**2**].

#### 2.3.1. *Problem Formulation in Noisy Space*

An $n \times n$ symmetrical, PD matrix $\Sigma$ with $\mathrm{cor}(\Sigma) = 0$ is given. Find all diagonal, NND matrices $\tilde{\Sigma}$ such that:

1. $\hat{\Sigma} = \Sigma - \tilde{\Sigma}$,
2. $\hat{\Sigma}$ is NND, and
3. $\mathrm{cor}(\Sigma)$ is maximum.

A vector $a$ is a solution vector if it satisfies the following relation:

$$\hat{\Sigma}\, a = (\Sigma - \tilde{\Sigma})\, a = 0. \tag{2.4}$$

#### 2.3.2. *Problem formulation in solution space*

An $n \times n$ symmetrical, PD matrix $\Sigma$ with $\mathrm{cor}(\Sigma) = 0$ is given. Find all vectors $a$ such that:

1. A NND, diagonal $\tilde{\Sigma}$ exists such that $\Sigma\, a = \tilde{\Sigma}\, a$ (2.5)
2. $\hat{\Sigma} = \Sigma - \tilde{\Sigma}$ is NND, and
3. $\mathrm{cor}(\hat{\Sigma})$ is maximum.

## 3. IMPORTANT SPECIAL SOLUTIONS WITHIN THE FRISCH SCHEME

This Section gives the analysis of special solutions in the set defined by the general Frisch scheme, which are of special importance for the prediction of dependent variables. As all the special solutions of interest reduce to projecting the analyzed random vector $\boldsymbol{X}$ onto principal components, we will first present a brief survey of known results.

### 3.1. The Principal Components of a Random Vector

For a random zero-mean vector $\boldsymbol{X} = (x_1, x_2, \dots, x_n)$, Span($\boldsymbol{X}$) denotes the Hilbert space of all random variables that are linear combinations of vectors from $\{x_1, x_2, \dots, x_n\}$.

$\boldsymbol{x}|\boldsymbol{X}$ denotes the linear estimator of the minimum variance of zero-mean random vector $\boldsymbol{x}$, on the basis of zero-mean random vector $\boldsymbol{X}$. This is also an orthogonal projection of $\boldsymbol{x}$ onto subspace Span($\boldsymbol{X}$). On the basis of elementary estimation theory, [3]:

$$\boldsymbol{x}|\boldsymbol{X} = \boldsymbol{X}[E\{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\}]^{-1}E\{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{x}\}. \tag{3.1}$$

#### 3.1.1. *Karhunen-Loeve Decomposition of Random Vector*

One of the orthonormal basis for space Span($\boldsymbol{X}$) is the vector:

$$Z = \boldsymbol{X}\boldsymbol{U}S^{-1} \tag{3.2a}$$

where:

$$\Sigma = \boldsymbol{U}S^2\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}_1S_1^2\boldsymbol{U}_1^{\mathrm{T}} + \boldsymbol{U}_2S_2^2\boldsymbol{U}_2^{\mathrm{T}} \tag{3.2b}$$

is the decomposition by the eigenvectors of covariance matrix $\Sigma = E(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$.

The presentation of the random vector in this basis is referred to as the Karhunen-Loeve decomposition of the random vector and reads:

$$\boldsymbol{X} = \boldsymbol{U}SZ = \sum_{k=1}^{n} \sigma_k \boldsymbol{u}_k z_k. \tag{3.3}$$

As the covariance matrix $\Sigma$ is symmetric and nonnegative definite, its eigen decomposition is identical to the singular value decomposition (SVD). Therefore, if the eigenvalues are arranged into a decreasing order $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$, then the principal components of $\Sigma$ are:

$$\boldsymbol{U}_1S_1^2\boldsymbol{U}_1^{\mathrm{T}} = \sum_{k=1}^{p} \sigma_k^2 \boldsymbol{u}_k \boldsymbol{u}_k^{\mathrm{T}}. \tag{3.4}$$

Principal components of matrix $\Sigma$ also represent the principal components of the random vector $\boldsymbol{X}$. The importance of the principal directions lies in that they provide optimum data compression of the random vector.

### 3.2. The Optimal Solution in Predictive Efficiency Means

That solution $a$ from the general Frisch scheme is sought for which the prediction error, expressed in terms of the following norm, is minimum:

$$\|E(X - X|Xa)\|_F^2$$

(where $\|\cdot\|_F$ is the Frobenius norm [4] defined as $\|A\|_F^2 = \operatorname{tr} A^T A$) or, in a more general case, that, matrix solution $A$ (matrix $n \times p$, $(p < n)$) is sought which minimizes the following norm:

$$\|E(X - X|XA)\|_F^2,$$

where, on the basis of the elementary estimation theory [3]:

$$\begin{aligned} X|XA &= XA[E\{A^T X^T XA\}]^{-1} E\{A^T X^T X\} \\ &= XA[A^T \Sigma A]^{-1} A^T \Sigma. \end{aligned} \tag{3.5}$$

The following result is known [5]:

$$\text{Minimum } \|E(X - X|XA)\|_F^2 = \sum_{k=p+1}^{n} \sigma_k^2 \tag{3.6}$$

and is achieved for:

$$A = U_1 \mathcal{A}$$

where: $U$ are the $p$ first principal components obtained by the SVD of covariance matrix $\Sigma$ and $\mathcal{A}$ — an arbitrary invertible $p \times p$ matrix.

Therefore, the optimal estimate of $X$ based on $XA$, $X|XA$, in the sense of a minimum prediction error, reads:

$$X|XA = \sum_{k=1}^{p} \sigma_k u_k z_k^T \tag{3.8}$$

i.e., the projection of the random vector $X$ onto the $p$ principal components.

The approximation of the random vector $X$, optimal in the sense of prediction efficiency maximization is achieved by principal components.

### 3.3. The Optimal Solution in Correlation Maximizes Means

The solution $a$, optimal with respect to maximizing the correlation from the set of solutions defined by the Frisch scheme, maximizes the following criterion function:

$$\|E\{X^T X a\}\|_F^2$$

subject to the constraint:

$$a^T a = 1. \tag{3.9}$$

In the matrix case: the optimal solution $A = \|a_1 a_2 \ldots a_p\|$ is obtained by maximizing the following criterion function:

$$\|E\{X^T X A\}\|_F^2$$

subject to the constraint:

$$\text{(a)} \quad A^T A = I_p. \tag{3.10a}$$

the components of matrix $A$ are orthogonal, or:

$$\text{(b)} \quad A^{\mathrm{T}}\Sigma A = I_p. \tag{3.10b}$$

The components of $XA$ are uncorrelated and have a unit variance; where: $\Sigma$ is the covariance matrix of random vector $X$.

The solution of the problem stated in this way is [5]:

$$A = U_1\mathcal{A}, \tag{3.11}$$

where: $U_1$ are the $p$ first principal components obtained by the SVD of covariance matrix $\Sigma$.

In case (a):

$$\mathcal{A} = I_p \tag{3.12a}$$

$$XA = (\sigma_1 z_1, \sigma_2 z_2, \dots, \sigma_p z_p). \tag{3.13a}$$

In case (b):

$$\mathcal{A} = S_1^{-1} \tag{3.12b}$$

$$XA = (z_1, z_2, \dots, z_p). \tag{3.13b}$$

The approximation of random vector $X$, optimal in the sense of correlation maximization, is also achieved by the principal components.

### 3.4. The Optimal Solution in Information Maximizes Means

We analyze here the case when $X$ is a Gaussian random vector. The problem is to find the solution from the set defined by the Frisch scheme having the property of self-information maximization. For the Gaussian random vector $X$, self-information is defined as:

$$0.5\log(\det(\Sigma))$$

where:

$$\Sigma = E(X^{\mathrm{T}}X).$$

Thus, the maximization of the information contained in $Xa$ is formally achieved by maximizing the following criterion function:

$$0.5 * \log(a^{\mathrm{T}}\Sigma a)$$

subject to constraint:

$$a^{\mathrm{T}}a = 1.$$

In the matrix case, when $A = \|a_1 a_2 \dots a_p\|$ and $(p < n)$, the optimization problem reduces to:

$$\max\ 0.5\log(\det(A^{\mathrm{T}}\Sigma A))$$

subject to constraint:

$$A^{\mathrm{T}}A = I_p. \tag{3.14}$$

The solution of the problem stated in this way is [6]:

$$A = U_1 \tag{3.15}$$

where: $U_1$ are the $p$ first principal components obtained by SVD of covariance matrix $\Sigma$.

The optimal approximation of the random vector $X$ reads:

$$XA = (\sigma_1 z_1, \sigma_2 z_2, \ldots, \sigma_p z_p). \tag{3.16}$$

The maximum self-information in approximation of the Gaussian random vector $X$ is:

$$0.5 \log(\det(A^{\mathrm{T}} \Sigma A)) = \sum_{k=1}^{p} \log \sigma_k. \tag{3.17}$$

The approximation of the random Gaussian vector $X$, optimal in the sense of information contents maximization, is also achieved by the principal components.

### 3.5. Comment

The Frisch scheme defines all the possible solutions of interest for approximating the random vector $X$. By seeking for the solutions with the following properties that are of special interest for the modelling of random vectors:

(a) prediction error minimization,

(b) correlation maximization and

(c) information contents maximization,

one obtains the same result. All the three stated properties are achieved through approximation of the random vector by the principal components of the covariance matrix $\Sigma$ of random vector $X$.

## 4. MODIFIED FRISCH SCHEME

We analyzed here the case when we distinguish the series of numbers between each other with respect to noise allocation. Some series are in the presence of noise and are treated as dependent variables, whereas the remaining series are accurate and are treated as independent variables in potential relations, [7].

Grouping of the series into noiseless and noisy series is performed as follows:

$$X = \|X_1 \; X_2\|, \tag{4.1}$$

where: $X$ is an $N \times m$ matrix of $N$ observations of $m$ quantities, $X_1$ is an $N \times m_1$ matrix of $N$ observations of $m_1$ noiseless series, and $X_2$ is an $N \times m_2$ matrix $N$ observations of $m_2$ noisy series.

The series after noise elimination are denoted as follows:

$$\hat{X} = \|X_1 \; \hat{X}_2\|, \tag{4.2}$$

where: $\hat{X}$ is an $N \times m$ matrix $X$ after noise elimination, and $\hat{X}_2$ is an $N \times m_2$ matrix $X_2$ after noise elimination.

Matrix of noise is denoted as follows:

$$\mathcal{E} = \|0 \ \mathcal{E}_2\|, \tag{4.3}$$

where: $\mathcal{E}$ is a $N \times m$ noise matrix of matrix $X$ and $\mathcal{E}_2$ is a $N \times m_2$ noise matrix of matrix $X_2$.

The observed matrix $X$ can be decomposed into an "accurate" part and noise:

$$X = \hat{X} + \mathcal{E} = \|X_1 \ \hat{X}_2\| + \|0 \ \mathcal{E}_2\|. \tag{4.4}$$

This decomposition, just as the one in the general case considered in Section 2, is not unique and only the decompositions satisfying the Frisch scheme are of interest.

A vector $a_1$ is a vector solution if it satisfies:

$$\hat{X} a_i = 0 \qquad \text{or} \tag{4.5a}$$

$$(X - \mathcal{E}) a_i = 0. \tag{4.5b}$$

If there are $k$ different solutions, the following can be written:

$$\hat{X} A = 0, \qquad \text{or} \tag{4.6a}$$

$$(X - \mathcal{E}) A = 0, \tag{4.6b}$$

where:

$A$ is an $m \times k$ matrix of $k$ solutions of the potential linear dependencies between the series analyzed,

$$A = \left\| \begin{matrix} A_1 \\ \bar{A}_2 \end{matrix} \right\|, \tag{4.7}$$

$A_1$ is an $m_1 \times k$ matrix of the parameters of $k$ solutions that correspond to the noiseless series, $(k < m_1)$;

$\bar{A}_2$ is an $m_2 \times k$ matrix of the parameters of $k$ solutions that correspond to the noisy series, $(k < m_2)$.

The following can be written for the product $XA$ in the case analyzed:

$$XA = \hat{X}A + \mathcal{E}A = \|X_1 \ \hat{X}_2\| \left\| \begin{matrix} A_1 \\ \bar{A}_2 \end{matrix} \right\| + \|0 \ \mathcal{E}_2\| \left\| \begin{matrix} A_1 \\ \bar{A}_2 \end{matrix} \right\|. \tag{4.8}$$

By using the covariance matrix of the analyzed observed series, $\Sigma$, which is defined by the following relation:

$$\Sigma := E(X^{\mathrm{T}} X) = E\left\{ \left\| \begin{matrix} X_1^{\mathrm{T}} \\ X_2^{\mathrm{T}} \end{matrix} \right\| \|X_1 \ X_2\| \right\} = \left\| \begin{matrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{matrix} \right\| \tag{4.9}$$

and the covariance matrix of noise, $\tilde{\Sigma}$:

$$\tilde{\Sigma} := E(\mathcal{E}^{\mathrm{T}} \mathcal{E}) = E\left\{ \left\| \begin{matrix} 0 \\ \mathcal{E}_2^{\mathrm{T}} \end{matrix} \right\| \|0 \ \mathcal{E}_2\| \right\} = \left\| \begin{matrix} 0 & 0 \\ 0 & \tilde{C}_{22} \end{matrix} \right\|. \tag{4.10}$$

it is possible to define the modified Frisch scheme.

### 4.1. Basic Assumptions — the Modified Frisch Scheme

The basic assumptions to be satisfied by the series analyzed

$$X = \|X_1\ X_2\|,$$

and by their covariance matrix

$$\Sigma = \begin{Vmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{Vmatrix} \tag{4.11}$$

are:

Assumption 1:
$$\operatorname{cor}(\Sigma) = 0. \tag{4.12}$$

Assumption 2: *Noise is additive.*

$$X = \hat{X} + \mathcal{E} = \|X_1\ \ \hat{X}_2\| + \|0\ \ \mathcal{E}_2\|. \tag{4.13}$$

where: $\hat{X}$ is the matrix $X$ after noise elimination, and $\mathcal{E}$ the noise matrix.

Only the data in $X_2$ are assumed to be noisy.

Assumption 3: *The columns of the pure noise matrix, $\mathcal{E}$, or more precisely the columns of submatrix $\mathcal{E}_2$, are independent of (uncorrelated with) the columns of the matrix of accurate data $\hat{X}$.*

As a consequence of assumptions 2 and 3 we have:

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma},$$

where: $\hat{\Sigma}$ is the matrix of covariance of series $\hat{X}$ and $\tilde{\Sigma}$ the noise covariance matrix $\mathcal{E}$:

$$\tilde{\Sigma} = \begin{Vmatrix} 0 & 0 \\ 0 & \tilde{C}_{22} \end{Vmatrix}. \tag{4.14}$$

Assumption 4: *The noise covariance matrix, $\tilde{\Sigma}$, is a nonnegative definite (NND) diagonal matrix.*

White noise is assumed.

The problem of finding all possible solutions that satisfy the modified Frisch scheme I can be defined in the noise space and solution space, analogously to the general case considered in Section 2.

#### 4.1.1. *Problem Formulation in Noise Space*

An $n \times n$ symmetrical, PD matrix is given:

$$\Sigma = \begin{Vmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{Vmatrix}$$

whose:

$$\mathrm{cor}(\Sigma) = 0.$$

The problem is to find all diagonal, NND matrices

$$\tilde{\Sigma} = \begin{Vmatrix} 0 & 0 \\ 0 & \tilde{C}_{22} \end{Vmatrix}.$$

such that:

1. $\hat{\Sigma} = \Sigma - \tilde{\Sigma}$,
2. $\hat{\Sigma}$ is NND, and
3. $\mathrm{cor}(\hat{\Sigma})$ is maximum.

A vector $a$ is a solution vector if it satisfies the following relation:

$$\hat{\Sigma}\ a = (\Sigma - \tilde{\Sigma})\ a = 0.$$

### 4.1.2. *Problem formulation in solution space*

An $n \times n$ symmetrical, PD matrix is given:

$$\Sigma = \begin{Vmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{Vmatrix}$$

where:

$$\mathrm{cor}(\Sigma) = 0.$$

Find all vectors $a$ such that:

1. There exists a NND diagonal matrix

$$\tilde{\Sigma} = \begin{Vmatrix} 0 & 0 \\ 0 & \tilde{C}_{22} \end{Vmatrix},$$

for which:

$$\Sigma\ a = \tilde{\Sigma}\ a,$$

2. $\hat{\Sigma} = \Sigma - \tilde{\Sigma}$ is NND, and
3. $\mathrm{cor}(\tilde{\Sigma})$ is maximum.

The original problem of finding the linear relation can be generalized for the case of $k$ linear relations in the following way:

$$\Sigma A = (\hat{\Sigma} - \tilde{\Sigma})A, \tag{4.15}$$

where: $A$ is an $n \times k$ matrix of parameters of $k$ $(k < n)$ linear relations.

As the following holds for solutions $A$:

$$\hat{\Sigma}A = 0. \tag{4.15a}$$

It follows that

$$\Sigma A = \tilde{\Sigma} A \tag{4.15b}$$

or, written in a different form:

$$\begin{Vmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{Vmatrix} \begin{Vmatrix} A_1 \\ \bar{A}_2 \end{Vmatrix} = \begin{Vmatrix} 0 & 0 \\ 0 & \tilde{C}_{22} \end{Vmatrix} \begin{Vmatrix} A_1 \\ \bar{A}_2 \end{Vmatrix} . \tag{4.16}$$

The relation:

$$C_{11} A_1 + C_{12} \bar{A}_2 = 0$$

gives:

$$A_1 = -C_{11}^{-1} C_{12} \bar{A}_2. \tag{4.17}$$

By including this equation in the following

$$C_{21} A_1 + C_{22} \bar{A}_2 = \tilde{C}_{22} \bar{A}_2 \tag{4.18}$$

we obtain:

$$(C_{22} - C_{21} C_{11}^{-1} C_{12}) \bar{A}_2 = \tilde{C}_{22} \bar{A}_2. \tag{4.19}$$

The potential solutions of the problem analyzed, finding the linear dependence between the series, some of which are noisy and the remaining are noiseless, must satisfy the above-given system of equations.

## 5. IMPORTANT SPECIAL SOLUTIONS WITHIN THE MODIFIED FRISCH SCHEME

This Section gives the analysis of special solutions from the set defined by the modified Frisch scheme which are of importance for the simultaneous estimation of several dependent variables as a function of multiple independent variables. Attention is focussed on the solutions that:

(a) maximize the correlation measure without normalization

(b) minimize the prediction error of unnormalized variables

(c) maximize the normalized correlation measure (canonical correlation)

(d) maximize the explained variance.

The results of this Section are supposed to provide a background for developing a noniterative algorithm for modeling of both scalar and multivariable time series.

### 5.1. The Optimal Solution With Respect to Maximizing the Unweighted Correlation Measure (Principal Components $C_{12}$)

The principal components $C_{12}$ represent the solution, within the modified Frisch scheme, which maximizes the unnormalized measure of correlation between the estimates of dependent and independent variables.

Determining the parameters of the relaxation between the dependent and independent variables by using the method of principal components $C_{12}$ reduces to solving the following optimization problem:

$$\text{maximize } \|E(A_1^T X_1^T X_2)\|_F^2$$

subject to the constraints from the modified Frisch scheme and:

$$A_1^{\mathrm{T}} A_1 = I_k. \tag{5.1}$$

Since:

$$\|E(A_1^{\mathrm{T}} X_1^{\mathrm{T}} X_2)\|_F^2 = \|E(A_1^{\mathrm{T}} C_{12})\|_F^2 = \operatorname{tr} A_1^{\mathrm{T}} C_{12} C_{21} A_1. \tag{5.2}$$

The solution of the optimization problem is obtained by applying the singular value decomposition off crosscovariance matrix $C_{12}$ and determining the principal components $U_1$, [8].

$$C_{12} = USV^{\mathrm{T}} = U_1 S_1 V_1^{\mathrm{T}} + U_2 S_2 V_1^{\mathrm{T}}, \tag{5.3}$$

where the following relations hold:

$$U_1^{\mathrm{T}} U_1 = I_k \tag{5.4a}$$

$$V_1^{\mathrm{T}} V_1 = I_k. \tag{5.4b}$$

The solution of the optimization problem is:

$$A_1 = U_1 \tag{5.5}$$

$$A_2 = -(C_{21} C_{12})^{-1} C_{21} C_{11} U_1. \tag{5.6}$$

The linear relation based on the principal components $C_{12}$ reads.

$$\hat{X}_2 (C_{21} C_{12})^{-1} C_{21} C_{11} U_1 = X_1 U_1.$$

The solution obtained in this way is not optimal with respect to the prediction, because the dependent variable has not been normalized and the result is affected by the relative differences in the variances of dependent variables.

### 5.2. The Optimal Solution With Respect to Minimizing the Prediction Error for Unweighted Variables

The directions of the minimum prediction error are the solution of the problem of minimizing the total prediction error for the unnormalized independent variables.

$$\text{minimize } \|E(X_1 A_1 + X_2 A_2)\|_F^2$$

subject to the constraints from the Frisch scheme and:

$$A_2^{\mathrm{T}} A_2 = I_k. \tag{5.7}$$

Since:

$$\begin{aligned} \|E(X_1 A_1 + X_2 A_2)\|_F^2 &= \|E(\mathcal{E}_2 A_2)\|_F^2 \\ &= \operatorname{tr} A_2^{\mathrm{T}} \tilde{C}_{22} A_2 \\ &= \operatorname{tr} A_2^{\mathrm{T}} (C_{22} - C_{21} C_{11}^{-1} C_{12}) A_2 \end{aligned} \tag{5.8}$$

the solution of the optimization problem is obtained by applying the singular value decomposition of $(C_{22} - C_{21} C_{11} C_{12})^{-1}$ and determining the principal components $U_1$:

$$(C_{22} - C_{21} C_{11} C_{12})^{-1} = US^2 U^{\mathrm{T}} = U_1 S_1^2 U_1^{\mathrm{T}} + U_2 S_2^2 U_2^{\mathrm{T}}, \tag{5.9}$$

where the following relations hold:

$$U_1^{\mathrm{T}} U_1 = I_k \tag{5.10}$$

The solution of the optimization problem is:

$$A_2 = U_1 \tag{5.11}$$

$$A_1 = -C_{11}^{-1} C_{12} U_1. \tag{5.12}$$

The linear relation based on the principal components $(C_{22} - C_{21}C_{11}^{-1}C_{12})^{-1}$, i.e. on the components of the minimum prediction error reads:

$$\hat{X}_2 U_1 = X_1 C_{11}^{-1} C_{12} U_1.$$

The solution obtained in this way is not optimal with respect to the prediction because the variables have not been normalized by the constraints introduced.

### 5.3. The Optimal Solution With Respect to Maximizing the Unweighted Correlation Measure

The linear relations obtained by applying the canonical correlation analysis **[9]**, **[10]**, represent the solution within the Frisch scheme which has the property of maximizing the information contained in the estimate of dependent variables taken from the set of independent variables.

Determining the parameters of the linear relation by using the method of canonical correlation analysis reduces formally to the following optimization problem:

$$\text{maximize } \|E(A_1^{\mathrm{T}} X_1^{\mathrm{T}} X_2 C_{22}^{-1/2})\|_F$$

subject to the constraints from the Frisch scheme and:

$$A_1^{\mathrm{T}} C_{11} A_1 = I_k. \tag{5.13}$$

Since:

$$\|E(A_1^{\mathrm{T}} X_1^{\mathrm{T}} X_2 C_{22}^{-1/2})\|_F = \|A_1^{\mathrm{T}} C_{12} C_{22}^{-1/2}\|_F = \operatorname{tr} A_1^{\mathrm{T}} C_{12} C_{22}^{-1} C_{21} A_1 \tag{5.14}$$

the solution of the optimization problem is obtained by determining the principal components of expression $C_{11}^{-1/2} C_{12} C_{22}^{-1/2}$:

$$C_{11}^{-1/2} C_{12} C_{22}^{-1/2} = USV^{\mathrm{T}} = U_1 S_1 V_1^{\mathrm{T}} + U_2 S_2 V_2^{\mathrm{T}}, \tag{5.15}$$

where the following relations hold:

$$U_1^{\mathrm{T}} U_1 = I_k \tag{5.16a}$$

$$V_1^{\mathrm{T}} V_1 = I_k. \tag{5.16b}$$

The solution of the optimization problem is:

$$A_1 = C_{11}^{-1/2} U_1 \tag{5.17}$$

$$A_2 = -(C_{21}C_{12})^{-1}C_{21}C_{11}^{1/2}U_1. \tag{5.18}$$

The linear relation based on the canonical correlation analysis reads:

$$\hat{X}_2(C_{21}C_{12})^{-1}C_{21}C_{11}^{1/2}U_1 = X_1U_1.$$

The solution obtained in this way is not optimal from the prediction standpoint because the treatment of dependent and independent variables is symmetrical.

NOTE: the same result is obtained with the following problem statement minimize

$$\text{minimize } \|E(X_1A_1 + X_2A_2)\|_F^2$$

subject to the constraints from the Frisch scheme and:

$$A_2^T C_{22} A_2 = I_k; \qquad \text{or} \tag{5.19}$$

$$A_{11}^T C_{11} A_1 = I_k. \tag{5.20}$$

## 5.4. The Optimal Solution With Respect to Maximizing the Explained Part of the Variance of Dependent Variables (the Unweighted Principal Components)

The unweighted principal components are the solutions within the modified Frisch scheme having the property of maximizing the explained part of the variance in the set of dependent variables on the basis of normalized independent variables.

Determining the linear relation by using the method of unweighted principal components reduces formally to solving the following optimization problem, [**10**] and [**11**]:

$$\text{maximize } \|E(A_1^T X_1^T X_2)\|_F^2$$

subject to the constraints from the Frisch scheme and:

$$A_1^T C_{11} A_1 = I_k. \tag{5.21}$$

Since:

$$\|E(A_1^T X_1^T X_2)\|_F^2 = \|E(A_1^T C_{12})\|_F^2 = \operatorname{tr} A_1^T C_{12} C_{21} A_1 \tag{5.22}$$

the solution of the optimization problem is obtained by determining the principal components of expression $C_{21}C_{11}^{-1}C_{12}$, by applying the SVD:

$$C_{21}C_{11}^{-1}C_{12} = US^2U^T = U_1S_1^2U_1^T + U_2S_2^2U_2^T. \tag{5.23}$$

where the following relation holds:

$$U_1^T U_1 = I_k. \tag{5.24}$$

The solution of the optimization problem is:

$$A_1 = -C_{11}^{-1}C_{12}U_1 \tag{5.25}$$

$$A_2 = U_1. \tag{5.26}$$

The linear relation based on the principal components of the explained part of variance reads:

$$\hat{X}_2 U_1 = X_1 C_{11}^{-1} C_{12} U_1 .$$

The solution obtained by employing the method of the unweighted principal components of the matrix of explained part of variance is optimal from the prediction standpoint. This solution represents a basis of a future algorithm for time-series modelling by a noniterative method.

### 5.5. Comments

We have analyzed the solutions of interest for the purposes of prediction from a set of solutions defined by the modified Frisch scheme which are obtained by optimization with respect to the above-stated criteria.

It can be seen that the solution for the analyzed criterion functions and constraints differ from one another. The method of the unweighted principal components (the principal components of matrix of the explained variance of dependent variables) is the solution representing a background for implementing a noniterative algorithm for time-series modelling.

## 6. CONCLUSION

This paper presents a modification to the Frisch scheme made for the following special case: some series of numbers are noiseless — independent variables; and the remaining are noisy — dependent variables. Special results from the set defined by modified Frisch scheme, which represent potential solution for prediction are analysed. The result obtained on unweighted principal components of the matrix of covariance of explained part of variance of dependent variables, represent the best solution for prediction.

The results obtained represent a basis of a new method for identifying simultaneously the structure and parameter values of a model intended for the prediction of vectorial stationary time series. The new method will be described in a forthcoming paper.

## REFERENCES

[1] Kalman, R. E., *Identification of noisy systems*, Uspekhi Matematicheskih Nauk **40** (1985), 27–41.

[2] De Moor, B. and Vandwvalle, *The uniqueness versus the non-uniqueness principle in the identification of linear relations from noisy data*, Proc of 25th Conf. Decision and Control, Athens, Greece, 1986.

[3] Dillon, R. W. and Goldstein M., *Multivariate Analysis — Methods and Applications*, John Wiley & Sons, New York, 1984.

[4] Golub, H. G. and Van Loan C. F., *Matrix Computations*, second edition, The Johns Hopkins University Press, Baltimore and London, 1989.

[5] Rao C. R., *The use and interpretation of principal component analysis in applied research*, Sankhya Series A, Vol. **26** (1964), 329–358.

[6] Haykin, S. S., *Nonlinear Methods of Spectral Analysis*, Springer-Verlag, New York, 1979.

[7] Radojević, D. and Krtolica R., *Frisch scheme modified for a dynamic case*, (in Serbian), Proc. of SYM-OP-IS'90, Dubrovnik, Kupari, Yugoslavia, 1990, 47–50.

[8] Cadzov, J. A., *Spectral estimation: an overdetermined rational model equation approach*, Proceeding of the IEEE **70**, No. 9 (1982), 907–939.

[9] Akaike, H., *Markovian representation of stochastic processes by canonical variables*, SIAM Journal on Control **13**, No. 1 (1975), 162–173.

[10] Aoki, M., *State Space Modeling of Time Series*, second edition. Springer-Verlag, Berlin-Heidelberg, 1991.

[11] Arun, K. S., B. Rao, and S. Y. Kung, *A new predictive efficiency criterion for approximate stochastic realization*, 1983 Conf. on Decision and Control, (1983), 1353–1355.