Spectra and Dynamic PCs via Block Matrix Decompositions

Marianna Bolla

Math. Inst., Budapest University of Technology and Economics

marib@math.bme.hu

High Dimensional Time Series Workshop, Vienna

June 10, 2022.

Outline

- Asymptotic relation between the eigenvalues of the block Toeplitz matrix of autocovariances in the time domain and the union of spectra of the spectral density matrices at the Fourier frequencies in the frequency domain.
- As a consequence, the principal component (PC) transformation of the multivariate real time series results in the complex increment process in the frequency domain.
- Low rank approximation of high dimensional time series.
- Dimension reduction and dynamic PC algorithm for real life data.
- Block LDL decomposition for innovations.

Outline

- Asymptotic relation between the eigenvalues of the block Toeplitz matrix of autocovariances in the time domain and the union of spectra of the spectral density matrices at the Fourier frequencies in the frequency domain.
- As a consequence, the principal component (PC) transformation of the multivariate real time series results in the complex increment process in the frequency domain.
- Low rank approximation of high dimensional time series.
- Dimension reduction and dynamic PC algorithm for real life data.
- Block LDL decomposition for innovations.

Outline

- Asymptotic relation between the eigenvalues of the block Toeplitz matrix of autocovariances in the time domain and the union of spectra of the spectral density matrices at the Fourier frequencies in the frequency domain.
- As a consequence, the principal component (PC) transformation of the multivariate real time series results in the complex increment process in the frequency domain.
- Low rank approximation of high dimensional time series.
- Dimension reduction and dynamic PC algorithm for real life data.
- Block LDL decomposition for innovations.

Outline

- Asymptotic relation between the eigenvalues of the block Toeplitz matrix of autocovariances in the time domain and the union of spectra of the spectral density matrices at the Fourier frequencies in the frequency domain.
- As a consequence, the principal component (PC) transformation of the multivariate real time series results in the complex increment process in the frequency domain.
- Low rank approximation of high dimensional time series.
- Dimension reduction and dynamic PC algorithm for real life data.
- Block LDL decomposition for innovations.

Outline

- Asymptotic relation between the eigenvalues of the block Toeplitz matrix of autocovariances in the time domain and the union of spectra of the spectral density matrices at the Fourier frequencies in the frequency domain.
- As a consequence, the principal component (PC) transformation of the multivariate real time series results in the complex increment process in the frequency domain.
- Low rank approximation of high dimensional time series.
- Dimension reduction and dynamic PC algorithm for real life data.
- Block LDL decomposition for innovations.

Time domain

$\{\mathbf{X}_t\}$: d-dimensional, weakly stationary time series with real components and autocovariance matrices $\mathbf{C}(h)$, $\mathbf{C}(-h) = \mathbf{C}^T(h)$, $h \in \mathbb{Z}$.
$\mathfrak{C}_n$: covariance matrix of $[\mathbf{X}_1^T, \dots, \mathbf{X}_n^T]^T \in \mathbb{R}^{nd}$:

$$\mathfrak{C}_n := \begin{bmatrix} \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \cdots & \mathbf{C}(n-1) \\ \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \cdots & \mathbf{C}(n-2) \\ \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \cdots & \mathbf{C}(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}^T(n-1) & \mathbf{C}^T(n-2) & \mathbf{C}^T(n-3) & \cdots & \mathbf{C}(0) \end{bmatrix}.$$

This is a symmetric, positive semidefinite block Toeplitz matrix, the (i, j) block of which is $\mathbf{C}(j - i)$.

Eigenvalues of the block circulant matrix

To characterize the eigenvalues of $\mathfrak{C}_n$, first we need the symmetric block circulant matrix $\mathfrak{C}_n^{(s)}$ that we consider now for odd n, say $n = 2k + 1$.
The (i, j) block of $\mathfrak{C}_n^{(s)}$ for $1 \le i \le j \le n$ is

$$\mathfrak{C}_n^{(s)}(\mathrm{block}_i, \mathrm{block}_j) = \begin{cases} \mathbf{C}(j-i) & j-i \le k \\ \mathbf{C}(n-(j-i)), & j-i > k. \end{cases}$$

For $i > j$, it is

$$\mathfrak{C}_n^{(s)}(\mathrm{block}_i, \mathrm{block}_j) = \begin{cases} \mathbf{C}^T(i-j) & i-j \le k \\ \mathbf{C}^T(n-(i-j)), & i-j > k. \end{cases}$$

$\mathfrak{C}_n^{(s)}$ is a symmetric block Toeplitz matrix again, and it is the same as $\mathfrak{C}_n$ within the blocks (i,j)s for which $|j-i| \le k$ holds. For example, if $n = 7$ and $k = 3$, then we have

$$\mathfrak{C}_7^{(s)} := \begin{bmatrix} \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) & \mathbf{C}(3) & \mathbf{C}(2) & \mathbf{C}(1) \\ \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) & \mathbf{C}(3) & \mathbf{C}(2) \\ \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) & \mathbf{C}(3) \\ \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) & \mathbf{C}(3) \\ \mathbf{C}^T(3) & \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) & \mathbf{C}(2) \\ \mathbf{C}^T(2) & \mathbf{C}^T(3) & \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) & \mathbf{C}(1) \\ \mathbf{C}^T(1) & \mathbf{C}^T(2) & \mathbf{C}^T(3) & \mathbf{C}^T(3) & \mathbf{C}^T(2) & \mathbf{C}^T(1) & \mathbf{C}(0) \end{bmatrix}.$$

1D

In 1D, by Kronecker products (with permutation matrices), it is well known that the jth eigenvalue of $\mathbf{C}_n^{(s)}$ is $\sum_{h=0}^{n-1} c(h)\rho_j^h$, where $\rho_j = e^{i\omega_j}$ is the jth primitive (complex) nth root of 1 and $\omega_j = \frac{2\pi j}{n}$ is the jth Fourier frequency ($j = 0, 1, \ldots, n-1$).
The eigenvector corresponding to the jth eigenvalue is $(1, \rho_j, \ldots, \rho_j^{n-1})^T$; it has norm $\sqrt{n}$.
After normalizing with $\frac{1}{\sqrt{n}}$, we get a complete orthonormal set of eigenvectors (of complex coordinates).

multi-D

When $\mathbf{C}(h)$s are $d \times d$ matrices, by inflation techniques and applying Kronecker products, we use blocks instead of entries, and the eigenvectors also follow a block structure.
Friedman (1961) and Tee (2007) characterize the eigenvalues and eigenvectors of a general symmetric block circulant matrix. We apply this result in our situation, when $n = 2k + 1$ is odd (for even n similar results hold).
The spectrum of $\mathfrak{C}_n^{(s)}$ is the union of spectra of the matrices

$$\mathbf{M}_j = \mathbf{C}(0) + \sum_{h=1}^{k} [\mathbf{C}(h)\rho_j^h + \mathbf{C}^T(h)\rho_j^{-h}]$$

for $j = 0, 2, \ldots, n-1$, whereas the eigenvectors are obtained by compounding the eigenvectors of these $d \times d$ matrices.

$\mathbf{M}_{n-j} = \overline{\mathbf{M}_j}$ (entrywise conjugate), therefore, it has the same eigenvalues as $\mathbf{M}_j$, but the eigenvectors are the (componentwise) complex conjugates of the eigenvectors of $\mathbf{M}_j$.

Summarizing, for odd $n = 2k + 1$, the *nd* eigenvalues of $\mathfrak{C}_n^{(s)}$ are obtained as the union of the eigenvalues of $\mathbf{M}_0$ and those of $\mathbf{M}_j$ $(j = 1, \dots, k)$ duplicated. Note that for even n, similar arguments hold with the difference that there the spectrum of $\mathfrak{C}_n^{(s)}$ is the union of the eigenvalues of $\mathbf{M}_0$ and $\mathbf{M}_{n-1}$, whereas the eigenvalues of $\mathbf{M}_1, \dots, \mathbf{M}_{\frac{n}{2}-1}$ are duplicated.

The eigenvectors of $\mathfrak{C}_n^{(s)}$ are obtainable by compounding the d orthonormal eigenvectors of the $d \times d$ self-adjoint matrices $\mathbf{M}_0, \mathbf{M}_1, \ldots, \mathbf{M}_{n-1}$ as follows.
For $j = 1, \ldots, k$: if $\mathbf{v}$ is an eigenvector of $\mathbf{M}_j$ with eigenvalue λ, then the compounded (stacked) vector

$$\mathbf{w} = (\mathbf{v}^T, \rho_j \mathbf{v}^T, \rho_j^2 \mathbf{v}^T, \ldots, \rho_j^{n-1} \mathbf{v}^T)^T \in \mathbb{C}^{nd}$$

is an eigenvector of $\mathfrak{C}_n^{(s)}$ with the same eigenvalue λ. Further, if

$$\mathbf{z} = (\mathbf{t}^T, \rho_\ell \mathbf{t}^T, \rho_\ell^2 \mathbf{t}^T, \ldots, \rho_\ell^{n-1} \mathbf{t}^T)^T \in \mathbb{C}^{nd}$$

is another eigenvector of $\mathfrak{C}_n^{(s)}$ compounded from an eigenvector $\mathbf{t}$ of another $\mathbf{M}_\ell$ ($\ell \neq j$), then $\mathbf{w}$ and $\mathbf{z}$ are orthogonal, irrespective whether $\mathbf{M}_\ell$ has the same eigenvalue λ as $\mathbf{M}_j$ or not. Similar construction holds starting with the eigenvectors of $\mathbf{M}_0$.

For each $j = 0, 1, \dots, n-1$, there are d pairwise orthogonal eigenvectors (potential $\mathbf{v}$s) of $\mathbf{M}_j$, and the so obtained $\mathbf{w}$s are also pairwise orthogonal. Assume that the eigenvectors of $\mathbf{M}_j$ are enumerated in non-increasing order of its eigenvalues, and the inflated $\mathbf{w}$s also follow this ordering, for $j = 0, 1, \dots, n-1$.
If $\mathbf{v}$ is an eigenvector of $\mathbf{M}_j$ with real eigenvalue λ, then $\overline{\mathbf{v}}$ is the corresponding eigenvector of $\mathbf{M}_{n-j}$ with the same eigenvalue λ; further, the compounded (stacked) $\mathbf{w}$ and $\overline{\mathbf{w}} \in \mathbb{C}^{nd}$ are orthogonal eigenvectors of $\mathfrak{C}_n^{(s)}$ corresponding to the eigenvalue λ with multiplicity (at least) two; $\mathbf{w}$ and $\overline{\mathbf{w}}$ have the same norm.
From them, corresponding to this double eigenvalue λ, the new orthogonal pair of eigenvectors

$$\frac{\mathbf{w} + \overline{\mathbf{w}}}{\sqrt{2}} \quad \text{and} \quad i\frac{\mathbf{w} - \overline{\mathbf{w}}}{\sqrt{2}}.$$

Note that it is necessary to have an orthogonal system of eigenvectors with real coordinates whenever the underlying time series is real, and so, $\mathfrak{C}_n^{(s)}$ is a real symmetric matrix.

After normalization, denote by $\mathbf{u}_1, \ldots, \mathbf{u}_{nd}$ the so obtained orthonormal set of eigenvectors (of real coordinates) of $\mathfrak{C}_n^{(s)}$ (in the above ordering) and by $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_{nd})$ the $nd \times nd$ orthogonal matrix containing them columnwise; further, let

$$\mathfrak{C}_n^{(s)} = \mathbf{U} \mathbf{\Lambda}^{(s)} \mathbf{U}^T$$

be the corresponding spectral decomposition. After this preparation, we are able to prove the main theorem.

Frequency domain

Denoting by $\mathbf{C}(h) = [c_{ij}(h)]$ the $d \times d$ autocovariance matrices $(\mathbf{C}(-h) = \mathbf{C}^T(h),\ h \in \mathbb{Z})$ in the time domain, assume that their entries are absolutely summable, i.e., $\sum_{h=0}^{\infty} |c_{pq}(h)| < \infty$ for $p, q = 1, \dots, d$.
Then, the self-adjoint, positive semidefinite spectral density matrix $\mathbf{f}(\omega)$ exists in the frequency domain, and it is defined by

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathbf{C}(h) e^{-ih\omega}, \quad \omega \in [0, 2\pi].$$

At the Fourier frequencies, it resembles the formula for $\mathbf{M}_j$s.

The Main Theorem

Let $\{\mathbf{X}_t\}$ be d-dimensional weakly stationary time series of real components. For odd $n = 2k+1$, consider $\mathbf{X}_1, \dots, \mathbf{X}_n$ with the block Toeplitz matrix $\mathfrak{C}_n$; further, the Fourier frequencies $\omega_j = \frac{2\pi j}{n}$ for $j = 0, \dots, n-1$. Let

$$\mathbf{D}_n = \mathtt{diag}\,(\mathtt{Spec}\,\mathbf{f}(0), \mathtt{Spec}\,\mathbf{f}(\omega_1), \dots, \mathtt{Spec}\,\mathbf{f}(\omega_k), \\ \mathtt{Spec}\,\mathbf{f}(\omega_k), \dots, \mathtt{Spec}\,\mathbf{f}(\omega_1)).$$

Here $\mathtt{Spec}$ contains the eigenvalues of the affected matrix in non-increasing order if not otherwise stated. (The duplication is due to the fact that $\mathbf{f}(\omega_j) = \overline{\mathbf{f}(\omega_{n-j})}$, $j = 1, \dots, k$, for real time series). Then, with the above spectral decomposition:

$$\mathbf{U}^T \mathfrak{C}_n \mathbf{U} - 2\pi \mathbf{D}_n \to \mathbf{O}, \quad n \to \infty,$$

i.e., the entries of the matrix $\mathbf{U}^T \mathfrak{C}_n \mathbf{U} - 2\pi \mathbf{D}_n$ tend to 0 uniformly as $n \to \infty$.

Consequence I: Real PCA

$\mathbf{U}$: $nd \times nd$ matrix containing the orthonormal eigenvectors $\mathbf{u}_j$ (of real coordinates) of the block circulant matrix $\mathfrak{C}_n^{(s)}$ in its columns. This is a real orthogonal matrix, and so, the „principal component" transformation

$$\mathbf{Y} = \mathbf{U}^T \mathbf{X}$$

of the nd-dimensional random vector $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$ results in an nd-dimensional random vector $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_n^T)^T$ (of real coordinates), the components of which are approximately uncorrelated, since

$$\mathbb{E}\mathbf{Y}\mathbf{Y}^T = \mathbf{U}^T \mathfrak{C}_n \mathbf{U} \sim 2\pi \mathbf{D}_n$$

for „large" n. Therefore, the „principal components", i.e., the random variables $Y_j = \mathbf{u}_j^T \mathbf{X}$ ($j = 1, \dots, nd$) are asymptotically uncorrelated.

Consequently, the d-dimensional random vectors $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are both **cross-sectionally and longitudinally uncorrelated**, approximately.

Consequence II: Complex PCA

$\mathbf{W}$: $nd \times nd$ matrix containing the orthonormal eigenvectors $\mathbf{w}_j$ (of complex coordinates) of the block circulant matrix $\mathfrak{C}_n^{(s)}$ in its columns.
Let $\mathbf{Z} = (\mathbf{Z}_1^T, \dots, \mathbf{Z}_n^T)^T$ denote the random vector obtained by

$$\mathbf{Z} = \mathbf{W}^* \mathbf{X}.$$

Its (complex) components are also uncorrelated and $\mathbb{E}\mathbf{Z}\mathbf{Z}^* \sim 2\pi \mathbf{D}_n$ again. Instead, we consider the blocks $\mathbf{Z}_j$s of it, and perform a „partial principal component transformation" (in d-dimension) of them. Let $\mathbf{w}_{1j}, \dots, \mathbf{w}_{dj}$ be the columns of $\mathbf{W}$ corresponding to the coordinates of $\mathbf{Z}_j$. Then by the block nature of the eigenvectors:

$$\mathbf{Z}_j = \frac{1}{\sqrt{n}} (\mathbf{V}_j^* \otimes \mathbf{r}^*) \mathbf{X},$$

where $\mathbf{r}^* = (1, \rho_j^{-1}, \rho_j^{-2}, \dots, \rho_j^{-(n-1)})$ and $\mathbf{V}_j$ is the $d \times d$ unitary matrix in the spectral decomposition $\mathbf{M}_j = \mathbf{V}_j \Lambda_j \mathbf{V}_j^*$.

Consequence III: Inverse Discrete Fourier Transform

The main Theorem implies that

$$\mathbb{E}(\mathbf{V}_j\mathbf{Z}_j)(\mathbf{V}_j\mathbf{Z}_j)^* = \mathbf{V}_j\Lambda_j\mathbf{V}_j^* = \mathbf{M}_j.$$

At the same time,

$$\mathbf{V}_j\mathbf{Z}_j = \frac{1}{\sqrt{n}}\mathbf{V}_j(\mathbf{V}_j^* \otimes \mathbf{r}^*)\mathbf{X} = \frac{1}{\sqrt{n}}(\mathbf{I}_d \otimes \mathbf{r}^*)\mathbf{X}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{X}_t e^{-it\omega_j}, \quad j = 1, \ldots, n.$$

This is the finite DFT of $\mathbf{X}_1, \ldots, \mathbf{X}_n$. It is also in accord with the definition of **the orthogonal increment process** $\{\mathbf{Z}_\omega\}$ of which $\mathbf{V}_j\mathbf{Z}_j \sim \mathbf{Z}_{\omega_j}$ is the discrete analogue. Also, $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are asymptotically pairwise orthogonal akin to $\mathbf{V}_1\mathbf{Z}_1, \ldots, \mathbf{V}_n\mathbf{Z}_n$. Further,

$$\mathbb{E}(\mathbf{V}_j\mathbf{Z}_j)(\mathbf{V}_j\mathbf{Z}_j)^* \sim 2\pi\mathbf{f}(\omega_j),$$

and it is in accord with the fact that

$$\mathbb{E}\mathbf{Z}_j\mathbf{Z}_j^* \sim 2\pi\,\mathtt{diag}\ \mathrm{spec}\ \mathbf{f}(\omega_j), \quad j = 1, \ldots, n.$$

Consequence IV: Bounds for the eigenvalues of $\mathfrak{C}_n$

$$m := \inf_{\omega \in [0,2\pi], q \in \{1,\ldots,d\}} \lambda_q(\mathbf{f}(\omega)) > 0,$$
$$M := \sup_{\omega \in [0,2\pi], q \in \{1,\ldots,d\}} \lambda_q(\mathbf{f}(\omega)) < \infty.$$

(Note that under the conditions of the Main Theorem, $\mathbf{f}(\omega) > 0$ and it is continuous almost everywhere on $[0, 2\pi]$, so the above conditions are readily satisfied.)
Then for the eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{nd}$ of the block Toeplitz matrix $\mathfrak{C}_n$ the following holds:

$$2\pi m \le \lambda_1 \le \lambda_{nd} \le 2\pi M.$$

Low rank approximation of the process

To find the best k-rank approximation of the weakly stationary, regular d-dimensional process, the d-dimensional vectors $\mathbf{V}_j\mathbf{Z}_j$s, obtained by DFT, should be projected onto the subspace spanned by the k leading eigenvectors of $\mathbf{V}_j$, $k \le d$, denoted by $\mathbf{V}_j^{(k)}$. Assume that the eigenvalues in $\mathbf{\Lambda}_j$ are in non-increasing order. Let us denote the k leading eigenvectors by $\mathbf{v}_{j1}, \dots, \mathbf{v}_{jk}$. Then the best rank k approximation of $\mathbf{V}_j\mathbf{Z}_j$:

$$(\mathbf{V}_j\mathbf{Z}_j)^{(k)} = \mathtt{Proj}_{\,\mathtt{Span}\,\{\mathbf{v}_{j1},\dots,\mathbf{v}_{jk}\}}\mathbf{V}_j\mathbf{Z}_j = \left[\mathbf{V}_j^{(k)}(\mathbf{V}_j^{(k)})^*\right]\mathbf{V}_j\mathbf{Z}_j$$
$$= \sum_{\ell=1}^{k}(\mathbf{v}_{j\ell}^*\mathbf{V}_j\mathbf{Z}_j)\mathbf{v}_{j\ell} = \sum_{\ell=1}^{k} Z_{j\ell}\mathbf{v}_{j\ell},$$

where $Z_{j\ell}$ denotes the ℓth coordinate of $\mathbf{Z}_j$.

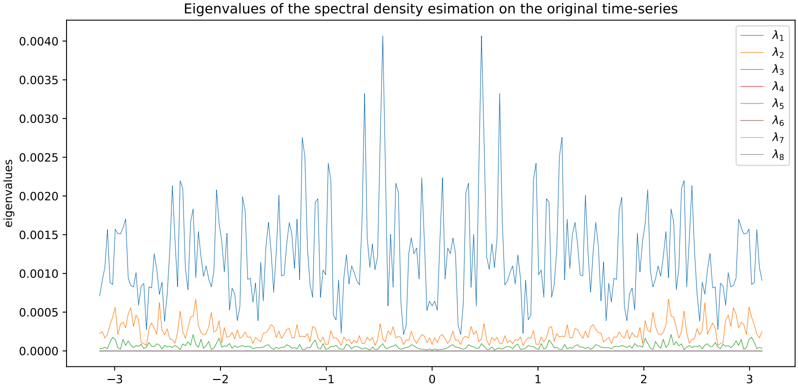
Frequency domain to time domain

This transformation gives rise to **rank** reduction in the frequency domain, then via DFT (due to $\mathbf{X} = \mathbf{WZ}$), in the time domain too. The best rank k approximation of $\mathbf{X}_t$:

$$(\mathbf{X}_t)^{(k)} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\mathbf{V}_j \mathbf{Z}_j)^{(k)} e^{it\omega_j}$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\sum_{\ell=1}^{k} Z_{j\ell} \mathbf{v}_{j\ell} \right) e^{it\omega_j},$$

for $t = 1, \dots, n$.
We can show that is is also a d-dimensional real time series, but its spectral density matrix is of rank $k \le d$.
This can be the common component for DFA.

Application to stock returns (Akbilgic, O. et al.)



The raw data were used.
The spectra shows 3 leading eigenvalues, the size of the gap after the leading eigenvalues depends on the spectral density estimation method.

Singular Autoregression

In a d-dimensional, weakly stationary time series with zero expectation, we linearly predict $\mathbf{X}_n$ based on past values $\mathbf{X}_1, \dots, \mathbf{X}_{n-1}$. Let $\hat{\mathbf{X}}_1 := 0$, and denote by $\hat{\mathbf{X}}_n$ the best one-step ahead linear prediction (based on $(n-1)$-long past) that minimizes the mean square error

$$\mathbb{E}(\mathbf{X}_n - \hat{\mathbf{X}}_n)^2 = \|\mathbf{X}_n - \hat{\mathbf{X}}_n\|^2, \quad n = 1, 2, \dots$$

which is the instance of simultaneous linear regressions. $\mathbf{X}_t$ can be expanded in terms of the now d-dimensional innovations, i.e. the prediction error terms

$$\boldsymbol{\eta}_n := \mathbf{X}_n - \hat{\mathbf{X}}_n,$$

with error covariance matrix $\mathbf{E}_n = \mathbb{E}\boldsymbol{\eta}_n\boldsymbol{\eta}_n^T$.

Block LDL decomposition with pseudo-inverses

Consider the first n steps, i.e. the recursive equations

$$\mathbf{X}_j = \sum_{k=1}^{j-1} \mathbf{B}_{jk} \boldsymbol{\eta}_k + \boldsymbol{\eta}_j, \quad j = 1, 2, \dots, n$$

in the case when the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ are available. If our process is stationary, the coefficient matrices are irrespective of the choice of the starting time, and in the regular case, they approach the one-step ahead projection based on the infinite past. It can be that the error covariance matrices are not zeros, but they are of reduced rank or better and better approach a rank r innovation covariance matrix, with decreasing ranks $(r \le d)$.
Multiplying the above equations by $\mathbf{X}_j^T$ from the right, and taking expectation, the solution for the matrices $\mathbf{B}_{jk}$ and $\mathbf{E}_j$ $(j = 1, \dots, n;\ k = 1, \dots, j-1)$ can be obtained via the block LDL (variant of the block Cholesky) decomposition:

$$\mathfrak{C}_n = \mathbf{L}_n \mathbf{D}_n \mathbf{L}_n^T .$$

Hhere $\mathfrak{C}_n$ is $nd \times nd$ positive semidefinite block Toeplitz matrix of general block entry $\mathbf{C}(i-j)$,

$$\mathbf{L}_n = \begin{bmatrix} \mathbf{I} & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} \\ \mathbf{B}_{21} & \mathbf{I} & \dots & \mathbf{O} & \mathbf{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{B}_{n1} & \mathbf{B}_{n2} & \dots & \mathbf{B}_{n,n-1} & \mathbf{I} \end{bmatrix}, \quad \mathbf{D}_n = \begin{bmatrix} \mathbf{E}_1 & \mathbf{O} & \dots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{E}_2 & \dots & \mathbf{O} & \mathbf{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{O} & \mathbf{O} & \dots & \mathbf{O} & \mathbf{E}_n \end{bmatrix}.$$

To find the block LDL (Cholesky) decomposition, the following recursion is used: for $j = 1, \dots, n$

$$\mathbf{E}_j := \mathbf{C}(0) - \sum_{k=1}^{j-1} \mathbf{B}_{jk} \mathbf{E}_k \mathbf{B}_{jk}^T, \quad j = 1, \dots, n$$

and for $i = j+1, \dots, n$

$$\mathbf{B}_{ij} := \left(\mathbf{C}(i-j) - \sum_{k=1}^{j-1} \mathbf{B}_{ik} \mathbf{E}_k \mathbf{B}_{ik}^T \right) \mathbf{E}_j^+,$$

where we take the Moore–Penrose inverse (denoted by $+$ in the superscript) if necessary and we do not enter into the summation if $j = 1$.

Remarks

- The innovation algorithm (variant of the Durbin–Levinson) also does it, provided $\mathfrak{C}_n$ is non-singular.
- Because of
$$|\mathfrak{C}_n| = |\mathbf{D}_n| = \prod_{j=1}^{n} |\mathbf{E}_j|,$$
if $|\mathfrak{C}_n| = 0$, then $|\mathbf{E}_j|$ becomes 0 (at least from a certain index j), but we can treat this situation with the pseudoinverse.
- Since $|\mathfrak{C}_n|$ is the product of the eigenvalues of $\mathfrak{C}_n$, which asymptotically comprise the union of the spectra of $\mathbf{f}(\omega)$ ($d \times d$ spectral density matrix) at the Fourier frequencies, singular prediction error matrices indicate reduced rank spectral density.
- $\mathbf{E}_1 = \mathbf{C}(0)$, $\mathtt{rank}\,\mathbf{C}(0) = r$, and $\mathbf{E}_j$s are the one-step ahead prediction (based on $j-1$ long past) error covariance matrices with non-increasing ranks.

By the multi-dimensional Wold decomposition, $\mathbf{E}_n \to \mathbf{\Sigma}$ in L^2-norm, where $\mathbf{\Sigma}$ is the error covariance matrix of the one-step ahead prediction based on the infinite past, $\texttt{rank}\,\mathbf{\Sigma} = q \le r$.
If the prediction is based on the infinite past, then with $n \to \infty$ this procedure (which is a nested one) extends to the multidimensional Wold decomposition.
If $n \to \infty$, the matrix $\mathbf{L}_n$ better and better approaches a block Toeplitz one, and the matrices $\mathbf{E}_1, \dots, \mathbf{E}_n$ are closer and closer to $\mathbf{\Sigma}$, the covariance matrix of the innovation process. Since $\|\mathbf{E}_n - \mathbf{\Sigma}\|^2 \to 0$ as $n \to \infty$, $\mathbf{B}_{nj} \to \mathbf{B}_j$ as $n \to \infty$ too, as it continuously depends on $\mathbf{E}_j$s.
As $\mathbf{E}_j$ is Cauchy sequence and we stop at a j ($j < n$, n is „large") where it does not change „much", then the jth block-row of $\mathbf{L}_n$ can be considered that it contains the effective coefficient matrices $\mathbf{B}_{jk}$s ($k = 1, \dots, j-1$) in a finite segment of the Wold decomposition. So a singular VAR(j) process is obtained if $q < r$.

Perturbation of eigenvalues

If there is a gap in the spectrum of $\mathbf{\Sigma}$, like

$$\lambda_1 \geq \cdots \geq \lambda_k \geq \Delta \gg \varepsilon \geq \lambda_{k+1} \geq \cdots \geq \lambda_d ,$$

then there is a gap in the spectrum of $\mathbf{E}_n$ too. Indeed, to any $\delta > 0$ there is an N such that for $n \geq N$: $\|\mathbf{E}_n - \mathbf{\Sigma}\| < \delta$. Then for the eigenvalues of $\mathbf{E}_n$,

$$\lambda_1^{(n)} \geq \cdots \geq \lambda_k^{(n)} \geq \Delta - \delta \gg \varepsilon + \delta \geq \lambda_{k+1}^{(n)} \geq \cdots \geq \lambda_d^{(n)} .$$

Consequently, for the best rank k approximations (with Gram-decompositions):

$$\|\mathbf{\Sigma} - \mathbf{\Sigma}^k\| \leq \varepsilon \quad \text{and} \quad \|\mathbf{E}_n - \mathbf{E}_n^k\| \leq \delta + \varepsilon$$

holds by the Weyl perturbation theorem. Therefore,

$$\|\mathbf{\Sigma}^k - \mathbf{E}_n^k\| \leq \|\mathbf{\Sigma}^k - \mathbf{\Sigma}\| + \|\mathbf{\Sigma} - \mathbf{E}_n\| + \|\mathbf{E}_n - \mathbf{E}_n^k\| \leq \varepsilon + \delta + (\delta + \varepsilon) = 2(\delta + \varepsilon)$$

that can be arbitrarily close to 2ε.

Perturbation of spectral subspaces

At the same time, the projections onto the subspaces spanned by the eigenvectors of the r structural eigenvalues of these matrices are close to each other, in the sense of the Davis–Kahan theorem.
Let $S_1 := [\Delta - \delta, \lambda_1 + \delta]$ and $S_2 := [\lambda_d + \delta, \varepsilon + \delta]$. Then for $n > N$:

$$\begin{aligned}
\|\mathbf{P}_{\boldsymbol{\Sigma}}(S_1) - \mathbf{P}_{\mathbf{E}_n}(S_1)\|_F^2 &= \|\mathbf{P}_{\boldsymbol{\Sigma}}(S_1)\|_F^2 + \|\mathbf{P}_{\mathbf{E}_n}(S_1)\|_F^2 - 2\mathtt{tr}\,[\mathbf{P}_{\boldsymbol{\Sigma}}(S_1)\mathbf{P}_{\mathbf{E}_n}^T(S \\
&= 2r - 2\mathtt{tr}\,[\mathbf{P}_{\boldsymbol{\Sigma}}(S_1)(\mathbf{I}_d - \mathbf{P}_{\mathbf{E}_n}^T(S_2))] \\
&= 2r - 2\mathtt{tr}\,[\mathbf{P}_{\boldsymbol{\Sigma}}(S_1) - \mathbf{P}_{\boldsymbol{\Sigma}}(S_1)\mathbf{P}_{\mathbf{E}_n}^T(S_2)] \\
&= 2r - 2r + 2\mathtt{tr}\,[\mathbf{P}_{\boldsymbol{\Sigma}}(S_1)\mathbf{P}_{\mathbf{E}_n}^T(S_2)] \\
&\le 2d\|\mathbf{P}_{\boldsymbol{\Sigma}}(S_1)\mathbf{P}_{\mathbf{E}_n}^T(S_2)\| \\
&\le 2d\frac{c}{\Delta - \delta - \varepsilon}\|\boldsymbol{\Sigma} - \mathbf{E}_n\| \le 2d\frac{c\delta}{\Delta - \delta - \varepsilon}
\end{aligned}$$

that can be arbitrarily „small" is δ is arbitrarily „small". Here $\mathbf{P}_{\boldsymbol{\Sigma}}(S)$ denotes the projection onto the subspace spanned by the eigenvectors of $\boldsymbol{\Sigma}$ corresponding to its eigenvalues in S.
Note that Davis–Kahan type theorems are effective is there is indeed a „large" gap in the spectrum (like $O(n)$ and $o(n)$).

Partly inspired by
Lippi, M., Deistler, M., Anderson, B., High-dimensional dynamic factor models . . . (arXiv: 2202.07745).

Happy birthday and many happy returns,
Manfred.