In *Proc. 2nd IMA Intl. Conf. on Mathematics in Communications*, Lancaster Univ., UK, Dec. 16-18, 2002.

Blind Identification and Equalization of MIMO FIR Channels Based on Subspace Decomposition and Independent Component Analysis

V. Zarzoso*, A. K. Nandi*, J. Igual-García† and L. Vergara-Domínguez†

*Department of Electrical Engineering and Electronics, University of Liverpool, UK
†Departamento de Comunicaciones, Universidad Politécnica de Valencia, Spain

Abstract

The identification and equalization of single-input digital communication channels can be accomplished blindly (without training sequences) using only second-order statistics (SOS) and parameter estimation algorithms based on matrix algebra tools. In this contribution, a blind identification and equalization (BIE) method based on the spectral decomposition of the sensor covariance matrix is extended to the multiuser scenario. This extended SOS-based BIE procedure leads to a co-channel interference (CCI) cancellation problem free of intersymbol interference. The statistical independence between the users' signals allows the application of independent component analysis, which arises as a strong alternative to CCI-cancellation techniques exploiting other spatio-temporal properties.

1 Introduction

In point-to-point digital communications, linear channel distortion (primarily caused by limited bandwidth and multipath propagation) introduces intersymbol interference (ISI) in the received signal, producing errors in symbol detection. A variety of equalizer designs can be employed to compensate for the channel effects [1]. *Blind* channel identification and equalization (BIE) methods present the benefit of not requiring training sequences. Original blind equalizers were based on higher-order statistics (HOS), which are computationally demanding and yield a slow convergence, not always to the global solution [1].

In single-input multiple-output (SIMO) signal models, non-minimum phase channels can be identified using only second-order statistics (SOS) and estimation procedures based on matrix algebra theory [2, 3]. Due to the cyclostationarity of the emitted digital signal, SIMO models originate when sampling the received signal faster than the baud rate (time oversampling) and/or multiple spatially-separated sensors exist (spatial oversampling).

V. Zarzoso is supported through a Postdoctoral Research Fellowship awarded by the Royal Academy of Engineering.

Due to their foreseen relevance in future wireless communication networks, air interface solutions based on non-orthogonal multiple access strategies have become an important research area [4]. In this type of communication environments, co-channel interference (CCI) caused by other users transmitting at the same time-frequency slot adds to multipath-induced ISI, hindering the reception of the signal(s) of interest. To mitigate the channel effects and ensure reliable detection, space-time equalization must be carried out. Time equalization aims at ISI removal, whereas space equalization involves CCI cancellation. The diversity arising from spatio-temporal oversampling results in multi-input multi-output (MIMO) signal models.

The present contribution extends the SOS-based SIMO-BIE method of [3] to the multiuser case. It is proved that the application of this extended method yields an ISI-free linear mixture of the users' signals, i.e., a CCI-only cancellation problem. We discuss the use of independent component analysis in this CCI-removal stage.

Notations: $\mathbf{I}_n$ refers to the $n \times n$ identity matrix and $\mathbf{0}_{m\times n}$ to the $m \times n$ zero matrix. Superindices $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{-1}$ indicate the transpose, hermitian (conjugate-transpose) and inverse matrix operators, respectively. $\mathrm{E}[\cdot]$ stands for mathematical expectation, symbol $\otimes$ represents the Kronecker product, and $\mathcal{R}(\mathbf{A})$ denotes the range (or column) space of matrix $\mathbf{A}$.

2 Signal Model

Consider a digital communication system where: i) K users simultaneously transmit zero-mean unit-power mutually-independent data symbols $\{s_{k,n}\}_{k=1}^{K} \in \mathbb{C}$, at a known constant rate; ii) the finite impulse responses $h_k^{(i)}$ representing the propagation (including the effects of the transmitter and receiver filters) between the kth source and the ith sensor (perhaps 'virtual', if induced by oversampling) span at most $M+1$ data symbols; iii) the additive measurement noise $w_n^{(i)}$ is zero-mean and uncorrelated with the data sequences. The discrete-time complex baseband received signal can then be expressed as:

$$x_n^{(i)} = \sum_{k=1}^{K} \sum_{m=0}^{M} h_{k,m}^{(i)} s_{k,n-m} + w_n^{(i)}, \qquad i = 1, \ldots, L. \tag{2.1}$$

N consecutive samples of the L (virtual) channel outputs are stored in vector $\mathbf{x}_n = [\mathbf{x}_n^{(1)\mathrm{T}}, \ldots, \mathbf{x}_n^{(L)\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{x}_n^{(i)} = [x_n^{(i)}, \ldots, x_{n-N+1}^{(i)}]^{\mathrm{T}}$, with analogous notation for the noise vector $\mathbf{w}_n$. Similarly, define the source symbol vector $\mathbf{s}_n = [\mathbf{s}_{1,n}^{\mathrm{T}}, \ldots, \mathbf{s}_{K,n}^{\mathrm{T}}]^{\mathrm{T}}$, with $\mathbf{s}_{k,n} = [s_{k,n}, \ldots, s_{k,n-M-N+1}]^{\mathrm{T}}$. The MIMO signal model is given by:

$$\mathbf{x}_n = \mathbf{H}_N \mathbf{s}_n + \mathbf{w}_n. \tag{2.2}$$

Defining $p \triangleq LN$, $c \triangleq M+N$ and $d \triangleq Kc$, in the above model $\mathbf{H}_N = [\mathbf{H}_{1,N}, \ldots, \mathbf{H}_{K,N}] \in \mathbb{C}^{p\times d}$, $\mathbf{H}_{k,N} = [\mathbf{H}_{k,N}^{(1)\mathrm{T}}, \ldots, \mathbf{H}_{k,N}^{(L)\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{p\times c}$ and $\mathbf{H}_{k,N}^{(i)} \in$

$\mathbb{C}^{N\times c}$ represents the Toeplitz *filtering matrix* associated with the linear filter $\mathbf{h}_k^{(i)} = [h_{k,0}^{(i)}, \ldots, h_{k,M}^{(i)}]^{\mathrm{T}}$ [3]. We assume that the channel matrix $\mathbf{H}_N$ is full column rank. The objective of BIE is to estimate $\mathbf{H}_N$ (channel identification) and $\mathbf{s}_n$ (space-time equalization) from the only observation of the received vector $\mathbf{x}_n$. These tasks are tantamount to recovering the channel coefficient vector $\mathbf{h} = [\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{KL(M+1)}$, with $\mathbf{h}_k = [\mathbf{h}_k^{(1)^{\mathrm{T}}}, \ldots, \mathbf{h}_k^{(L)^{\mathrm{T}}}]^{\mathrm{T}}$, and the source vector $\mathbf{s} = [s_{1,n}, \ldots, s_{K,n}]^{\mathrm{T}} = [\mathbf{s}_n(1), \mathbf{s}_n(c+1), \ldots, \mathbf{s}_n\big((K-1)c+1\big)]^{\mathrm{T}}$, where $\mathbf{s}_n(i)$ denotes the ith element of vector $\mathbf{s}_n$.

3 Subspace Approach

For $K = 1$, a second-order BIE approach is proposed in [3], based on the subspace decomposition of the sensor-output covariance matrix and benefiting from the channel-matrix Toeplitz structure. In this section we generalize the method to the multiuser case ($K > 1$). The identification of the filtering-matrix range space leads to the following indeterminacy result, which is proved in the appendix.

Theorem 1. *Assume that* (**A1**) *matrix* $\mathbf{H}_{N-1}$ *is full column rank;* (**A2**) $N > M$. *Let* $\tilde{\mathbf{H}}_N$ *be a nonzero filtering matrix with the same dimensions as* $\mathbf{H}_N$. *Then,* $\mathcal{R}(\tilde{\mathbf{H}}_N) \subset \mathcal{R}(\mathbf{H}_N)$ *if and only if* $\tilde{\mathbf{H}}_N = \mathbf{H}_N(\mathbf{A} \otimes \mathbf{I}_c)$, *where* $\mathbf{A}$ *is a regular* $K \times K$ *matrix with elements in* $\mathbb{C}$.

Now, from matrix model (2.2): $\mathbf{R}_x = \mathrm{E}[\mathbf{x}_n\mathbf{x}_n^{\mathrm{H}}] = \mathbf{H}_N\mathbf{R}_s\mathbf{H}_N^{\mathrm{H}} + \mathbf{R}_w$, where $\mathbf{R}_x$, $\mathbf{R}_s$ and $\mathbf{R}_w$ denote, respectively, the covariance matrices of the sensor output, sources and noise, defined accordingly. For simplicity, let us assume that $\mathbf{R}_w = \sigma^2\mathbf{I}_p$. Let $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_p$ denote the eigenvalues of $\mathbf{R}_x$. Since $\mathbf{R}_s$ is full rank, the *signal part* of $\mathbf{R}_x$, i.e., $\mathbf{H}_N\mathbf{R}_s\mathbf{H}_N^{\mathrm{H}}$, has rank d. Hence $\lambda_i > \sigma^2$, for $1 \leqslant i \leqslant d$ and $\lambda_i = \sigma^2$, for $d < i \leqslant p$. The sensor covariance matrix can then be expressed as $\mathbf{R}_x = \mathbf{U}\mathrm{diag}(\lambda_1, \ldots, \lambda_d)\mathbf{U}^{\mathrm{H}} + \sigma^2\mathbf{V}\mathbf{V}^{\mathrm{H}}$. The columns of matrix $\mathbf{U}$, which are the eigenvectors associated with $\{\lambda_i\}_{i=1}^d$, span the *signal subspace*. Its orthoghonal complement, $\mathcal{R}(\mathbf{U})^{\perp}$, is called the *noise subspace*, and is spanned by the columns of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_{p-d}]$, which are the eigenvectors associated with $\{\lambda_i\}_{i=d+1}^p$. The signal subspace is also spanned by the columns of the filtering matrix $\mathbf{H}_N$, i.e., $\mathcal{R}(\mathbf{H}_N) = \mathcal{R}(\mathbf{U})$, and hence, $\mathcal{R}(\mathbf{H}_N) = \mathcal{R}(\mathbf{V})^{\perp}$. In particular, $\mathbf{v}_i^{\mathrm{H}}\mathbf{H}_N = \mathbf{0}_{1\times d}$, $1 \leqslant i \leqslant p-d$. These orthogonality relations constitute a set of $d(p-d)$ linear equations in the $KL(M+1)$ unknown channel coefficients. To avoid indeterminacy, we must have $c(p-d) \geqslant L(M+1)$. This leads to the constraint $(L-K)N^2 + M(L-2K)N \geqslant KM^2 + L(M+1)$, which sets lower bounds on the smoothing factor N given the other system parameters (cf. [5]).

In practice, only sample estimates $\hat{\mathbf{v}}_i$ of the noise eigenvectors are available, and a solution in the least squares (LS) sense is sought, leading to the minimization of the quadratic form:

$$\xi(\mathbf{h}) = \sum_{i=1}^{p-d} \|\hat{\mathbf{v}}_i^{\mathrm{H}}\mathbf{H}_N\|^2. \tag{3.1}$$

The dependence of cost function ξ on the channel coefficient vector $\mathbf{h}$ can be made explicit with the help of the lemma below (proved in the appendix).

Lemma 2. If $\mathbf{V}_{i,M+1}$ represents the $L(M+1) \times c$ filtering matrix linked to the $LN \times 1$ vector $\hat{\mathbf{v}}_i$, then $\hat{\mathbf{v}}_i^{\mathrm{T}} \mathbf{H}_N = \mathbf{h}^{\mathrm{T}} \tilde{\mathbf{V}}_{i,M+1}$, where $\tilde{\mathbf{V}}_{i,M+1} = \mathbf{I}_K \otimes \mathbf{V}_{i,M+1}$.

For convenience, subindices $(\cdot)_N$ and $(\cdot)_{M+1}$ are dropped in the sequel. By virtue of Lemma 2: $\|\hat{\mathbf{v}}_i^{\mathrm{H}} \mathbf{H}\|^2 = \hat{\mathbf{v}}_i^{\mathrm{H}} \mathbf{H}\mathbf{H}^{\mathrm{H}} \hat{\mathbf{v}}_i = \mathbf{h}^{\mathrm{H}} \tilde{\mathbf{V}}_i \tilde{\mathbf{V}}_i^{\mathrm{H}} \mathbf{h} = \mathrm{trace}(\bar{\mathbf{H}}^{\mathrm{H}} \mathbf{V}_i \mathbf{V}_i^{\mathrm{H}} \bar{\mathbf{H}})$, where $\bar{\mathbf{H}} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{L(M+1)\times K}$. Cost function (3.1) finally becomes:

$$\xi(\bar{\mathbf{H}}) = \mathrm{trace}(\bar{\mathbf{H}}^{\mathrm{H}} \mathbf{Q} \bar{\mathbf{H}}), \quad \text{with } \mathbf{Q} = \sum_{i=1}^{p-d} \mathbf{V}_i \mathbf{V}_i^{\mathrm{H}}. \tag{3.2}$$

In order to avoid the trivial solution $\bar{\mathbf{H}} = \mathbf{0}_{L(M+1)\times K}$, criterion (3.2) must be minimized subject to certain constraint. For instance: if $\mathrm{trace}(\bar{\mathbf{H}}^{\mathrm{H}} \bar{\mathbf{H}}) = 1$, then the columns of estimate $\hat{\bar{\mathbf{H}}}$ are proportional to the eigenvectors associated with the smallest eigenvalue of matrix $\mathbf{Q}$; if $\mathrm{trace}(\mathbf{C}^{\mathrm{H}} \bar{\mathbf{H}}) = 1$, where $\mathbf{C}$ is a $L(M+1) \times K$ constraint matrix, then $\hat{\bar{\mathbf{H}}} = (\mathbf{Q}^{-1}\mathbf{C})/\mathrm{trace}(\mathbf{C}^{\mathrm{H}} \mathbf{Q}^{-1} \mathbf{C})$.

The above procedure identifies the channel matrix up to the indeterminacy shown in Theorem 1: $\hat{\mathbf{H}} = \mathbf{H}(\mathbf{A} \otimes \mathbf{I}_c)$, with $\mathbf{A}$ an unknown $K \times K$ regular matrix. In the absence of noise, the zero-forcing equalizer output is given by $\mathbf{z}_n = \hat{\mathbf{H}}^{\mathrm{H}} (\hat{\mathbf{H}} \hat{\mathbf{H}}^{\mathrm{H}})^{-1} \mathbf{x}_n = (\mathbf{A}^{-1} \otimes \mathbf{I}_c)\, \mathbf{s}_n$. Defining $\mathbf{z} = [\mathbf{z}_n(1), \mathbf{z}_n(c+1), \ldots, \mathbf{z}_n\big((K-1)c+1\big)]^{\mathrm{T}}$, where $\mathbf{z}_n(i)$ denotes the ith element of vector $\mathbf{z}_n$, the equalized system is equivalent to

$$\mathbf{z} = \mathbf{A}^{-1} \mathbf{s}. \tag{3.3}$$

4 ICA-Based CCI Cancellation and Symbol Detection

Eqn. (3.3) represents an ISI-free CCI-cancellation problem. To achieve space equalization, most proposed methods exploit properties inherent to digital signals, such as their constant modulus or finite alphabet (FA) [6]. The global convergence of some of these methods (e.g., the family of iterative LS algorithms [6]) is not generally guaranteed, even in the absence of noise [5].

The very plausible property of the mutual independence between the users' signals can also be exploited through the statistical tool of independent component analysis (ICA), which requires HOS if the source symbols are i.i.d. Benefits are exhibited in two aspects of the BIE problem [7]: CCI cancellation and symbol detection. Firstly, ICA leads to dramatic improvements in CCI-cancellation performance, compared to an FA-based method. Secondly, an ICA-based refinement yields remarkable performance gains relative to the conventional minimum mean square error (MMSE) receiver. Furthermore, the particular structure of the FIR-MIMO model enables a simplification of the ICA-aided MMSE detector, with improved performance and a lower computational cost.

Illustrative simulation results will be presented at the conference.

5 Conclusions

The subspace-based blind identification method of [3] relies on the channel-matrix structure and, as opposed to [2], is able to operate regardless of the source spectra. The extension of this method to the multiuser scenario performs time equalization only. Various spatio-temporal properties of the communication system can be exploited to achieve space equalization in a second processing stage. The versatility and improved CCI-cancellation and symbol detection performance of independence-exploiting ICA-based techniques render them very promising in next-generation cellular and ad-hoc wireless commercial networks as well as non-cooperative military scenarios.

References

1. J. K. Tugnait, L. Tong, and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Processing Magazine*, Vol. 17, No. 3, pp. 16–28, May 2000.
2. L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Transactions on Information Theory*, Vol. 40, No. 2, pp. 340–349, Mar. 1994.
3. E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, Vol. 43, No. 2, pp. 516–525, Feb. 1995.
4. T. Ojanperä and R. Prasad (Eds.), *Wideband CDMA for Third Generation Mobile Communications*, Artech House, Boston, MA, 1998.
5. X. Wang and H. V. Poor, "Blind equalization and multiuser detection in dispersive CDMA channels," *IEEE Transactions on Communications*, Vol. 46, No. 1, pp. 91–103, Jan. 1998.
6. A.-J. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Transactions on Signal Processing*, Vol. 45, No. 1, pp. 173–190, Jan. 1997.
7. V. Zarzoso and A. K. Nandi, "Exploiting independence for co-channel interference cancellation and symbol detection in multi-user digital communications," in *ICASSP-2003, 28th International Conference on Acoustics, Speech and Signal Processing*, to be held in Hong Kong, Apr. 6–10, 2003, submitted.

Appendix: Proofs

Proof of Theorem 1
Define $\underline{\mathbf{H}}_{(k)} \in \mathbb{C}^{L\times K}$ as $\left(\underline{\mathbf{H}}_{(k)}\right)_{ij} = h_{j,k}^{(i)}$, i.e., the matrix composed of all channel coefficients associated with delay k. Then, matrix

$$\underline{\mathbf{H}}_N = \begin{bmatrix} \underline{\mathbf{H}}_{(0)} & \underline{\mathbf{H}}_{(1)} & \cdots & \underline{\mathbf{H}}_{(M)} & \mathbf{0}_{L\times K} & \cdots & \mathbf{0}_{L\times K} \\ \mathbf{0}_{L\times K} & \underline{\mathbf{H}}_{(0)} & \underline{\mathbf{H}}_{(1)} & \cdots & \underline{\mathbf{H}}_{(M)} & \mathbf{0}_{L\times K} & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{L\times K} & \cdots & \mathbf{0}_{L\times K} & \underline{\mathbf{H}}_{(0)} & \underline{\mathbf{H}}_{(1)} & \cdots & \underline{\mathbf{H}}_{(M)} \end{bmatrix} \tag{A.1}$$

corresponds to a rearrangement of the rows and columns of matrix $\mathbf{H}_N$, and hence their column spaces are canonically equivalent (remark that a subsequent rearrangement of the source vector would leave the observed sensor output unaltered). Due to the block-Toeplitz structure of $\underline{\mathbf{H}}_N$ and the shift property of Toeplitz matrices, we have the two equivalent representations:

$$\underline{\mathbf{H}}_N = \begin{bmatrix} \underline{\mathbf{H}}_{(0)} & \mathbf{P}_{N-1} \\ \mathbf{0}_{L(N-1)\times K} & \underline{\mathbf{H}}_{N-1} \end{bmatrix} \quad \text{and} \quad \underline{\mathbf{H}}_N = \begin{bmatrix} \underline{\mathbf{H}}_{N-1} & \mathbf{0}_{L(N-1)\times K} \\ \mathbf{Q}_{N-1} & \underline{\mathbf{H}}_{(M)} \end{bmatrix} \tag{A.2}$$

with $\begin{cases} \mathbf{P}_{N-1} = [\underline{\mathbf{H}}_{(1)}, \ldots, \underline{\mathbf{H}}_{(M)}, \mathbf{0}_{L\times K(N-1)}] \\ \mathbf{Q}_{N-1} = [\mathbf{0}_{L\times K(N-1)}, \underline{\mathbf{H}}_{(0)}, \ldots, \underline{\mathbf{H}}_{(M-1)}]. \end{cases}$

The $p \times d$ block-Toeplitz matrix $\underline{\tilde{\mathbf{H}}}_N$ can be constructed accordingly.

The sufficiency part of the theorem is evident; the only difficulty lies in the necessity part. Assume that $\mathcal{R}(\underline{\tilde{\mathbf{H}}}_N) \subset \mathcal{R}(\underline{\mathbf{H}}_N)$. In particular, the columns of $[\underline{\tilde{\mathbf{H}}}_{(0)}^{\mathrm{T}}, \mathbf{0}_{L(N-1)\times K}^{\mathrm{T}}]^{\mathrm{T}}$, which belong to $\mathcal{R}(\underline{\tilde{\mathbf{H}}}_N)$, also belong to $\mathcal{R}(\underline{\mathbf{H}}_N)$. Taking into account the first representation in (A.2), there exist matrices $\mathbf{A}_0 \in \mathbb{C}^{K\times K}$ and $\mathbf{X}_0 \in \mathbb{C}^{K(c-1)\times K}$ such that

$$\begin{bmatrix} \underline{\tilde{\mathbf{H}}}_{(0)} \\ \mathbf{0}_{L(N-1)\times K} \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{H}}_{(0)} & \mathbf{P}_{N-1} \\ \mathbf{0}_{L(N-1)\times K} & \underline{\mathbf{H}}_{N-1} \end{bmatrix} \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{X}_0 \end{bmatrix}. \tag{A.3}$$

By virtue of assumption **A1**, matrix $\underline{\mathbf{H}}_{N-1}$ is full column rank, and then the above linear system implies that $\mathbf{X}_0 = \mathbf{0}_{K(c-1)\times K}$, and hence $\underline{\tilde{\mathbf{H}}}_{(0)} = \underline{\mathbf{H}}_{(0)}\mathbf{A}_0$. Similarly, the relation $[\underline{\tilde{\mathbf{H}}}_{(1)}^{\mathrm{T}}, \underline{\tilde{\mathbf{H}}}_{(0)}^{\mathrm{T}}, \mathbf{0}_{L(N-2)\times K}^{\mathrm{T}}]^{\mathrm{T}} \in \mathcal{R}(\underline{\mathbf{H}}_N)$ shows the existence of $\mathbf{A}_1 \in \mathbb{C}^{K\times K}$ such that $\underline{\tilde{\mathbf{H}}}_{(1)} = \underline{\mathbf{H}}_{(0)}\mathbf{A}_1 + \underline{\mathbf{H}}_{(1)}\mathbf{A}_0$. Analogous reasoning proves the existence of matrices $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_M \in \mathbb{C}^{K\times K}$ verifying:

$$\underline{\tilde{\mathbf{H}}}_{(k)} = \sum_{n=0}^{k} \underline{\mathbf{H}}_{(n)}\mathbf{A}_{k-n}, \qquad k = 0, \ldots, M. \tag{A.4}$$

Proceeding in a similar manner with the second decomposition in (A.2), one can show the existence of matrices $\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_M \in \mathbb{C}^{K\times K}$ fulfilling:

$$\underline{\tilde{\mathbf{H}}}_{(k)} = \sum_{n=k}^{M} \underline{\mathbf{H}}_{(n)}\mathbf{B}_{M+k-n}, \qquad k = 0, \ldots, M. \tag{A.5}$$

Both set of equations (A.4)–(A.5) can be compactly expressed as:

$$\underline{\mathbf{H}}_{M+1}\mathbf{A} = \underline{\mathbf{H}}_{M+1}\mathbf{B}, \quad \text{with} \begin{cases} \mathbf{A} = [\mathbf{A}_M^{\mathrm{T}}, \mathbf{A}_{M-1}^{\mathrm{T}}, \ldots, \mathbf{A}_0^{\mathrm{T}}, \mathbf{0}_{KM\times K}^{\mathrm{T}}]^{\mathrm{T}} \\ \mathbf{B} = [\mathbf{0}_{KM\times K}^{\mathrm{T}}, \mathbf{B}_M^{\mathrm{T}}, \mathbf{B}_{M-1}^{\mathrm{T}}, \ldots, \mathbf{B}_0^{\mathrm{T}}]^{\mathrm{T}}. \end{cases} \tag{A.6}$$

From hypothesis **A2**, $\underline{\mathbf{H}}_{M+1}$ is full rank, because $M + 1 \leqslant N$. Hence, the unique solution of the linear system (A.6) is $\mathbf{A} = \mathbf{B}$. This implies: $\mathbf{A}_0 = \mathbf{B}_M$; $\mathbf{A}_i = \mathbf{0}_{K\times K}$, $i = 1, \ldots, M$; and $\mathbf{B}_i = \mathbf{0}_{K\times K}$, $i = 0, \ldots, M-1$. Therefore, $\underline{\tilde{\mathbf{H}}}_{(k)} = \underline{\mathbf{H}}_{(k)}\mathbf{A}_0$, $k = 0, \ldots, M$, and then $\underline{\tilde{\mathbf{H}}}_N = \underline{\mathbf{H}}_N(\mathbf{I}_{M+N} \otimes \mathbf{A}_0)$. Due to the relationship between the structures of $\mathbf{H}_N$ and $\underline{\mathbf{H}}_N$, the latter equality leads to $\tilde{\mathbf{H}}_N = \mathbf{H}_N(\mathbf{A}_0 \otimes \mathbf{I}_c)$. Finally, matrix $\mathbf{A}_0$ must be regular to guarantee the full column rank of $\tilde{\mathbf{H}}_N$. □

Proof of Lemma 2

The Toeplitz structure of $\mathbf{H}_{k,N}^{(i)}$ leads to $\hat{\mathbf{v}}_i^{\mathrm{T}}\mathbf{H}_{k,N} = \mathbf{h}_k^{\mathrm{T}}\mathbf{V}_{i,M+1}$ [3]. Hence, $\hat{\mathbf{v}}_i^{\mathrm{T}}\mathbf{H}_N = \hat{\mathbf{v}}_i^{\mathrm{T}}[\mathbf{H}_{1,N}, \ldots, \mathbf{H}_{K,N}] = [\hat{\mathbf{v}}_i^{\mathrm{T}}\mathbf{H}_{1,N}, \ldots, \hat{\mathbf{v}}_i^{\mathrm{T}}\mathbf{H}_{K,N}] = [\mathbf{h}_1^{\mathrm{T}}\mathbf{V}_{i,M+1}, \ldots, \mathbf{h}_K^{\mathrm{T}}\mathbf{V}_{i,M+1}]$. The last term can be expressed as $[\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_K^{\mathrm{T}}]\,\mathrm{diag}(\underbrace{\mathbf{V}_{i,M+1}, \ldots, \mathbf{V}_{i,M+1}}_{K})$, which is equivalent to $\mathbf{h}^{\mathrm{T}}\tilde{\mathbf{V}}_{i,M+1}$. □