1 CONVERGENCE OF LINEAR SUCCESSIVE INTERFERENCE CANCELLATION IN CDMA

Lars K. Rasmussen[1], Teng J. Lim[1] and Ann-Louise Johansson[2]

[1] Centre for Wireless Communications
20 Science Park Road, #02-34/37 Teletech Park
Singapore 117674
[2] Department of Information Theory
Chalmers University of Technology, S-412 96 Gothenburg
Sweden

cwclkr@leonis.nus.edu.sg
cwclimtj@leonis.nus.edu.sg
anne@it.chalmers.se

In the proceedings of the IEEE Workshop on Multiple Access, Mobility and Teletraffic for Personal Communications (MMT), Melbourne, Australia, December 1997, pages 13–28.

Abstract: In this paper we consider a matrix-algebraic approach to linear successive interference cancellation (SIC). It has been shown that both single- and multi-stage linear SIC schemes correspond to a one-shot linear matrix filtering. Eigenvalue conditions on the resulting linear filter for convergence of the multi-stage scheme are considered and the concept of ϵ-convergence is introduced for determining the number of stages necessary for practical convergence. It is observed that the BER for the users does not generally reach its minimum when the scheme converges, hence, optimal performance is achieved after a limited number of cancellation stages, which does not have to be the same for different users.

1.1 INTRODUCTION

For practical implementation of multiuser detection in CDMA, parallel and successive interference cancellation (SIC) schemes have been subject to most attention. The first structure based on parallel cancellation was the multi-stage detector in [1]. In [2], Dent et al. proposed a serial approach that was extended to a multi-stage structure in [3]. In [4] it was shown that linear SIC corresponds to a one-shot linear matrix filtering and expressions for the exact BER was derived. In this paper we derive some quantitative statements concerning convergence and convergence rate for the linear multi-stage SIC. Furthermore, a technique is devised for determining the number of stages required for practical convergence.

Throughout this paper scalars are lower-case, vectors are bold face lower-case, and matrices are bold face upper-case. Subscripting is dropped where no ambiguities arise. The symbols $(\cdot)^\top$, $(\cdot)^{-1}$ and $\|\cdot\|$ are the transposition, inversion and Euclidean vector-norm operators respectively, and the delimiter $\{\cdot\}^y$ defines a space of dimension y. All vectors are defined as column vectors with row vectors represented by transposition. $\mathbb{R}$ denotes the set of real numbers, and the following notation is used for the product of matrices,

$$\begin{aligned}
\mathbf{X}_1\mathbf{X}_2\cdots\mathbf{X}_{n-1}\mathbf{X}_n &= \prod_{i=1}^{n}\mathbf{X}_i, \quad \prod_{i=1}^{0}\mathbf{X}_i = \mathbf{I} \\
\mathbf{X}_n\mathbf{X}_{n-1}\cdots\mathbf{X}_2\mathbf{X}_1 &= \coprod_{i=n}^{1}\mathbf{X}_i, \quad \coprod_{i=0}^{1}\mathbf{X}_i = \mathbf{I}.
\end{aligned}$$

1.2 SYSTEM MODELS

In this section, the model for the uplink of the CDMA communication system considered throughout this paper is briefly described. The uplink model is based on a discrete-time symbol-synchronous CDMA system assuming single-path channels and the presence of stationary additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2 = N_0/2$. A specific user in this K-user communication system transmits a binary information-symbol $d_k \in \{-1, 1\}$, by multiplying with a spreading code $\mathbf{s}_k \in \left\{\frac{-1}{\sqrt{N}}, \frac{1}{\sqrt{N}}\right\}^N$, with a length of N chips and then transmitting over an AWGN channel using BPSK. Each user is received at a user-specific energy level w_k that is assumed constant over one bit-interval. Note that we have assumed that $\mathbf{s}_k^\top\mathbf{s}_k = 1$. The output of a chip-matched filter is then expressed as a linear combination of spreading codes, specifically, the chip matched filtered received vector, $\mathbf{r}$, is a column vector of length N, encompassing the transmissions for all users. The received vector $\mathbf{r}$ is hence described through matrix-algebra as

$$\mathbf{r} = \mathbf{AWd} + \mathbf{n} \in \mathbb{R}^N, \tag{1.1}$$

where

$$\begin{aligned}\mathbf{A} &= (\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K) \in \left\{\frac{-1}{\sqrt{N}}, \frac{1}{\sqrt{N}}\right\}^{N,K}, \\ \mathbf{W} &= \operatorname{diag}(\sqrt{w_1}, \sqrt{w_2}, \cdots, \sqrt{w_K}) \in \mathbb{R}^{K,K}, \\ \mathbf{d} &= (d_1, d_2, \cdots, d_K)^\top \in \{-1, 1\}^K.\end{aligned}$$

The sampled noise corrupting the output of the chip-matched filter is independent in each sample since the channel noise is assumed to be white and the chip waveforms are assumed to fulfill the Nyquist criterion (e.g. rectangular chip pulses).

The received vector is contained in a vector space of dimension N, $\mathbf{r} \in \mathbb{R}^N$. It is, however, only the part of $\mathbf{r}$ residing in the signal space that is affecting the detector decision. The signal space is determined by $\operatorname{span}\{\mathbf{A}\}$. If $N = K$, $\mathbf{r} \in \operatorname{span}\{\mathbf{A}\}$. In general however, $\mathbf{r} = \mathbf{r}_s + \mathbf{r}_{s^\perp}$ with $\mathbf{r}_s \in \operatorname{span}\{\mathbf{A}\}$ and $\mathbf{r}_{s^\perp} \in \operatorname{null}(\mathbf{A})$ where $\operatorname{null}(\mathbf{A})$ denotes the nullspace of $\mathbf{A}$. We can determine $\mathbf{r}_s$ and $\mathbf{r}_{s^\perp}$ by orthogonal projections:

$$\begin{aligned}\mathbf{r}_s &= \mathbf{A}\left(\mathbf{A}^\top\mathbf{A}\right)^{\perp 1}\mathbf{A}^\top\mathbf{r} = \mathbf{P}_\mathbf{A}\mathbf{r} \\ \mathbf{r}_{s^\perp} &= (\mathbf{I} - \mathbf{P}_\mathbf{A})\,\mathbf{r} = \mathbf{P}_{\mathbf{A}^\perp}\mathbf{r}.\end{aligned}$$

This will be of significance for convergence rate considerations.

Successive interference cancellation schemes are best described by defining an interference cancelling unit (ICU) as shown in Figure 1.1. This unit is then

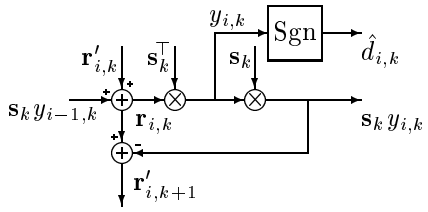

Figure 1.1 Linear successive interference cancellation unit.

used as a building-block in the multi-stage SIC scheme shown in Figure 1.2. The single-stage scheme is obtained by omitting all but the first stage. It is assumed that the K users are ordered according to their received signal power. The input signal vector to an ICU of user k at stage i is $\mathbf{r}'_{i,k}$. For the first user in the first stage $\mathbf{r}'_{1,1} = \mathbf{r}$. The contribution to be cancelled in the ICU of user k at stage i is $\mathbf{s}_k y_{i,k}$, where $y_{i,k} = \mathbf{s}_k^\top \mathbf{r}_{i,k} = \mathbf{s}_k^\top(\mathbf{r}'_{i,k} + \mathbf{s}_k y_{i-1,k})$. In geometrical terms this is a projection of the current received vector $\mathbf{r}_{i,k}$ onto the relevant spreading code $\mathbf{s}_k$. For the first stage $\mathbf{r}_{1,k} = \mathbf{r}'_{1,k}$ since the input, which corresponds to the cancelled contribution of the previous stage, is $\mathbf{s}_k y_{0,k} = 0$ for all ICU blocks of the first stage.

1.3 CONVERGENCE CONSIDERATIONS

The first user in the first stage of any of the SIC schemes is operating directly on the received signal vector $\mathbf{r}_{1,1} = \mathbf{r}$. The first-stage filter output is $y_{1,1} =$

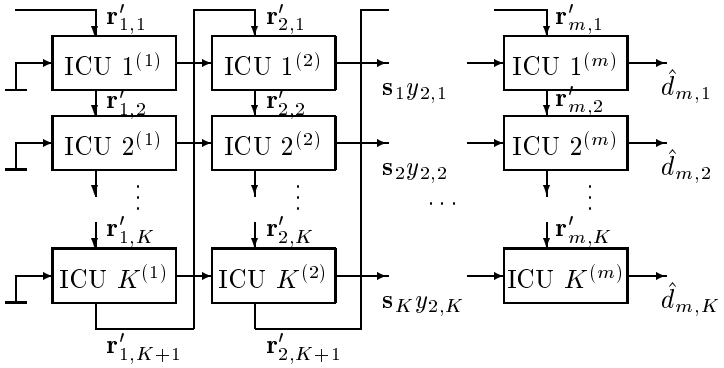

Figure 1.2 Multi-stage linear successive interference cancellation structure.

$\mathbf{s}_1^\top \mathbf{r}_{1,1} = \mathbf{s}_1^\top \mathbf{r}$, leading to the resulting input vector for the next unit as

$$\mathbf{r}_{1,2} = \mathbf{r}_{1,1} - \mathbf{s}_1 y_{1,1} = \mathbf{r}_{1,1} - \mathbf{s}_1 \mathbf{s}_1^\top \mathbf{r}_{1,1} = \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r},$$

where $\mathbf{I}$ is an $N \times N$ identity matrix. The next step is then $y_{1,2} = \mathbf{s}_2^\top \mathbf{r}_{1,2} = \mathbf{s}_2^\top \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r}$, and

$$\mathbf{r}_{1,3} = \mathbf{r}_{1,2} - \mathbf{s}_2 y_{1,2} = \mathbf{r}_{1,2} - \mathbf{s}_2 \mathbf{s}_2^\top \mathbf{r}_{1,2} = \left(\mathbf{I} - \mathbf{s}_2 \mathbf{s}_2^\top\right) \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r}.$$

In general we get

$$\mathbf{r}_{1,k} = \prod_{j=k-1}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \mathbf{r},$$

and

$$y_{1,k} = \mathbf{s}_k^\top \prod_{j=k-1}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \mathbf{r}.$$

A single-stage linear SIC scheme can therefore be represented as a linear matrix-filtering of the received signal vector $\mathbf{r}$, i.e., collecting the decision variables in a vector $\mathbf{y}_1 = (y_{1,1}, y_{1,2}, ..., y_{1,K})^\top$, we get $\mathbf{y}_1 = \mathbf{G}_1^\top \mathbf{r}$, where the subscript on $\mathbf{y}_1$ refers to the first stage and $\mathbf{G}_1$ is an $N \times K$ matrix filter with columns determined by

$$\mathbf{g}_{1,k} = \left[\mathbf{s}_k^\top \prod_{j=k-1}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right)\right]^\top = \prod_{j=1}^{k-1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \mathbf{s}_k,$$

where we get the last equality since $\mathbf{s}_j \mathbf{s}_j^\top$, and thereby also $\left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right)$, are symmetrical.

The output vector $\mathbf{r}'_{1,K+1}$ (see Figure 1.1) represents the residual received vector at the end of the first stage after all users have been processed. It is determined as

$$\mathbf{r}'_{1,K+1} = \coprod_{j=K}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \mathbf{r}. \tag{1.2}$$

For notational purposes in what is to follow we define

$$\begin{aligned} \mathbf{\Phi}_k &= \left(\mathbf{I} - \mathbf{s}_k \mathbf{s}_k^\top\right)\left(\mathbf{I} - \mathbf{s}_{k-1}\mathbf{s}_{k-1}^\top\right) \cdots \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \\ &\qquad \left(\mathbf{I} - \mathbf{s}_K \mathbf{s}_K^\top\right) \cdots \left(\mathbf{I} - \mathbf{s}_{k+1}\mathbf{s}_{k+1}^\top\right), \qquad \text{for } k \in \{1, 2, \ldots K\}. \\ &= \coprod_{j=k}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \coprod_{m=K}^{k+1} \left(\mathbf{I} - \mathbf{s}_m \mathbf{s}_m^\top\right). \end{aligned} \tag{1.3}$$

Eqn. (1.2) then becomes $\mathbf{r}'_{1,K+1} = \mathbf{\Phi}_K \mathbf{r}$. According to the scheme suggested in [3] we reintroduce the cancelled contribution of the relevant user into the residual received vector from the first stage, i.e.,

$$\mathbf{r}_{2,1} = \mathbf{r}'_{1,K+1} + \mathbf{s}_1 y_{1,1} = \mathbf{\Phi}_K \mathbf{r} + \mathbf{s}_1 \mathbf{s}_1^\top \mathbf{r} = \left(\mathbf{\Phi}_K + \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r}.$$

The matched filter output is then

$$y_{2,1} = \mathbf{s}_1^\top \mathbf{r}_{2,1} = \mathbf{s}_1^\top \left(\mathbf{\Phi}_K + \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r} = \mathbf{s}_1^\top \left(\mathbf{\Phi}_K + \mathbf{I}\right) \mathbf{r}.$$

In the next step we must both cancel the signal from the current step as well as add in the relevant signal from the previous stage. This gives us the following current received vector

$$\begin{aligned} \mathbf{r}_{2,2} &= \mathbf{r}_{2,1} - \mathbf{s}_1 y_{2,1} + \mathbf{s}_2 y_{1,2} \\ &= \left(\mathbf{\Phi}_K + \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r} - \mathbf{s}_1 \mathbf{s}_1^\top \left(\mathbf{\Phi}_K + \mathbf{I}\right) \mathbf{r} + \mathbf{s}_2 \mathbf{s}_2^\top \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r} \\ &= \left(\left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{\Phi}_K + \mathbf{s}_2 \mathbf{s}_2^\top \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right)\right) \mathbf{r} \\ &= \left(\mathbf{\Phi}_1 + \mathbf{s}_2 \mathbf{s}_2^\top\right) \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r}, \end{aligned}$$

because $\left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{\Phi}_K = \mathbf{\Phi}_1 \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right)$, and the following filter output

$$y_{2,2} = \mathbf{s}_2^\top \left(\mathbf{\Phi}_1 + \mathbf{s}_2 \mathbf{s}_2^\top\right) \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r} = \mathbf{s}_2^\top \left(\mathbf{\Phi}_1 + \mathbf{I}\right) \left(\mathbf{I} - \mathbf{s}_1 \mathbf{s}_1^\top\right) \mathbf{r}.$$

In general the terms in the second stage are described by

$$\mathbf{r}_{2,k} = \left(\mathbf{\Phi}_{k-1} + \mathbf{s}_k \mathbf{s}_k^\top\right) \coprod_{j=k-1}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \mathbf{r},$$

and

$$y_{2,k} = \mathbf{s}_k^\top \left(\mathbf{\Phi}_{k-1} + \mathbf{I}\right) \coprod_{j=k-1}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \mathbf{r}.$$

We recognise the last term of each equation as $\mathbf{r}_{1,k}$, so the following description is equivalent. $\mathbf{r}_{2,k} = \left(\mathbf{\Phi}_{k-1} + \mathbf{s}_k\mathbf{s}_k^\top\right)\mathbf{r}_{1,k}$, and $y_{2,k} = \mathbf{s}_k^\top\left(\mathbf{\Phi}_{k-1} + \mathbf{I}\right)\mathbf{r}_{1,k} = \mathbf{s}_k^\top\mathbf{\Phi}_{k-1}\mathbf{r}_{1,k} + y_{1,k}$. A two-stage linear SIC scheme is thus represented as $\mathbf{y}_2 = \mathbf{G}_2^\top\mathbf{r}$, where $\mathbf{G}_2$ is determined as

$$\mathbf{g}_{2,k} = \left[\mathbf{s}_k^\top\left(\mathbf{\Phi}_{k-1} + \mathbf{I}\right)\coprod_{j=k-1}^{1}\left(\mathbf{I} - \mathbf{s}_j\mathbf{s}_j^\top\right)\right]^\top = \prod_{j=1}^{k-1}\left(\mathbf{I} - \mathbf{s}_j\mathbf{s}_j^\top\right)\left(\mathbf{\Phi}_{k-1}^\top + \mathbf{I}\right)\mathbf{s}_k.$$

Continuing this way we obtain the general expressions for the k^{th} user in the i^{th} stage as $\mathbf{y}_i = \mathbf{G}_i^\top\mathbf{r}$, where $\mathbf{G}_i$ is determined as

$$\mathbf{g}_{i,k} = \prod_{j=1}^{k-1}\left(\mathbf{I} - \mathbf{s}_j\mathbf{s}_j^\top\right)\sum_{p=0}^{i-1}(\mathbf{\Phi}_{k-1}^\top)^p\mathbf{s}_k.$$

It follows then that the BER for the k^{th} user after i stages is

$$P_b(i,k) = \frac{1}{2^{K-1}}\sum_{\substack{\text{all } \mathbf{d}\\ d_k=1}} Q\left(\frac{\sqrt{2}\mathbf{g}_{i,k}^\top\mathbf{A}\mathbf{W}\mathbf{d}}{\sqrt{N_0\mathbf{g}_{i,k}^\top\mathbf{g}_{i,k}}}\right).$$

Based on this, it is possible to describe the optimal number of stages to be used for each user, i.e.,

$$i_{\text{opt}} = \min_i P_b(i,k).$$

This is however a very difficult analytical problem. Based on a series of numerical examples, it is observed that for low SNR the minimum BER is in fact achieved for a relatively low intermediate number of stages. An increasing number of stages does not necessarily improve performance. As the noise diminishes, the observed minimum flattens out and eventually disappears so that for $N_0 \to 0$, the BER performance will monotonically improve for an increasing number of stages. This behavior is similar to the MMSE detector for CDMA and it follows that the optimal number of stages will change according to noise level as well as the correlation among users. The minimum overall BER is achieved by letting each user have the corresponding optimal number of stages, which is not necessarily the same for all users.

In [5] it was pointed out that in case the linear SIC converges, the performance will converge towards the performance of the decorrelator. Here we consider whether the linear SIC will converge at all, and if so, at what rate. When considering convergence rate, it is convenient to have the received vector $\mathbf{r}$ completely contained in the signal space. In case $K < N$ we will therefore consider $\mathbf{r}_s$ rather than $\mathbf{r}$. The starting vector is therefore $\mathbf{r}_{1,1} = \mathbf{r}_s = \mathbf{P}_\mathbf{A}\mathbf{r}$.

In order to investigate the convergence rate, consider the current received vector for user k at stage i, $\mathbf{r}_{i,k}$, and the corresponding decision variable, $y_{i,k}$. From the above we have

$$\mathbf{r}_{i,k} = \mathbf{\Phi}_{k-1}\mathbf{r}_{i-1,k} + \mathbf{s}_k\mathbf{s}_k^\top\mathbf{r}_{i-1,k} = \mathbf{\Phi}_{k-1}\mathbf{r}_{i-1,k} + \mathbf{s}_k y_{i-1,k},$$

where the first term is the residual received vector forwarded from the previous ICU, and the second term is the reintroduction of the cancelled contribution for user k from the previous stage. If we avoid reintroducing $y_{i-1,k}$ and only process the residual received vector $\mathbf{r}'_{i,k}$, then we have the following equivalent representation,

$$\begin{aligned} \mathbf{r}'_{i,k} &= \mathbf{\Phi}_{k-1}\mathbf{r}'_{i-1,k} = \left(\mathbf{\Phi}_{k-1}\right)^{i-1}\mathbf{r}'_{1,k} \\ y'_{i,k} &= \mathbf{s}_k^{\top}\mathbf{\Phi}_{k-1}\mathbf{r}'_{i-1,k} = \mathbf{s}_k^{\top}\left(\mathbf{\Phi}_{k-1}\right)^{i-1}\mathbf{r}'_{1,k}, \end{aligned}$$

and then $y_{i,k} = \sum_{j=1}^{i} y'_{j,k}$. This is merely an alternative description of the linear SIC scheme which will prove useful for convergence rate considerations.

For convergence we have

$$y_{i,k} - y_{i-1,k} = y'_{i,k} = \mathbf{s}_k^{\top}\mathbf{\Phi}_{k-1}\mathbf{r}'_{i-1,k} \to 0 \quad \text{for } i \to \infty \quad \text{for } \forall\, k.$$

That obviously corresponds to

$$\mathbf{r}'_{i,k} = \mathbf{\Phi}_{k-1}\mathbf{r}'_{i-1,k} = \left(\mathbf{\Phi}_{k-1}\right)^{i-1}\mathbf{r}'_{1,k} \to \mathbf{0} \quad \text{for } i \to \infty \quad \text{for } \forall\, k. \tag{1.4}$$

We will concentrate on (1.4) for convergence rate estimation. Following the approach for the derivation of the steepest descent algorithm [6], a similarity transformation is applied to $\mathbf{\Phi}_{k-1}$,

$$\mathbf{\Phi}_{k-1} = \mathbf{U}_{k-1}\mathbf{\Lambda}_{k-1}\mathbf{U}_{k-1}^{\perp 1}, \tag{1.5}$$

where $\mathbf{\Lambda}_{k-1}$ is a diagonal matrix of eigenvalues, $\lambda_{k-1}(j)$, of $\mathbf{\Phi}_{k-1}$ and $\mathbf{U}_{k-1}$ is a matrix of the corresponding eigenvectors. Since $\mathbf{\Phi}_{k-1}$ is not symmetric, a unitary similarity transformation cannot be applied. Furthermore, the transformation is only valid if $\mathbf{\Phi}_{k-1}$ is non-defective [7]. Under certain circumstances when successive users are orthogonal, the corresponding $\mathbf{\Phi}_{k-1}$ is in fact defective. This is discussed in greater detail in Appendix A where it is concluded based on a combination of analysis and simulation observations that given no users are orthogonal the geometrical multiplicity equals the algebraic multiplicity for all the eigenvalues of $\mathbf{\Phi}_{k-1}$ and hence, the matrix is non-defective. It should be emphasised however that orthogonal users improve the convergence rate of the SIC. It is just that this particular approach to convergence rate considerations does not apply in case of a defective $\mathbf{\Phi}_{k-1}$. Systems including orthogonal users converge faster than systems without. Assuming that $\mathbf{\Phi}_{k-1}$ is non-defective, we then get for $i > 1$

$$\begin{aligned} & \mathbf{r}'_{i,k} = \mathbf{U}_{k-1}\mathbf{\Lambda}_{k-1}\mathbf{U}_{k-1}^{\perp 1}\mathbf{r}'_{i-1,k} \\ \Rightarrow\quad & \mathbf{U}_{k-1}^{\perp 1}\mathbf{r}'_{i,k} = \mathbf{\Lambda}_{k-1}\mathbf{U}_{k-1}^{\perp 1}\mathbf{r}'_{i-1,k} \\ \Rightarrow\quad & \mathbf{v}_{i,k} = \mathbf{\Lambda}_{k-1}\mathbf{v}_{i-1,k} = \mathbf{\Lambda}_{k-1}^{i-1}\mathbf{v}_{1,k}, \end{aligned}$$

where for all i

$$\begin{aligned} & \mathbf{v}_{i,k} = \mathbf{U}_{k-1}^{\perp 1}\mathbf{r}'_{i,k} \\ \Rightarrow\quad & \mathbf{r}'_{i,k} = \mathbf{U}_{k-1}\mathbf{v}_{i,k} = \sum_{n=1}^{N}\mathbf{u}_n v_{i,k}(n) = \sum_{n=1}^{N}\mathbf{u}_n \lambda_{k-1}^{i-1}(n) v_{1,k}(n), \end{aligned} \tag{1.6}$$

$\mathbf{U}_{k-1} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_N]$ and $v_{i,k}(n)$ denotes individual elements of $\mathbf{v}_{i,k}$.

Based on our assumptions, $\mathbf{r}_s$ is a vector in a K-dimensional subspace constituting the signal space in $\mathbb{R}^{N\times N}$. $\mathbf{r}'_{i,k}$ is a projection onto the $(K-1)$-dimensional subspace of span $\left\{\mathbf{I} - \mathbf{s}_k\mathbf{s}_k^\top\right\}$ or equivalently null $\left(\mathbf{s}_k\mathbf{s}_k^\top\right)$. Since $\mathbf{U}_{k-1}$ consists of N linearly independent eigenvectors, it is clearly a basis for $\mathbb{C}^{N\times N}$. As $\mathbf{r}'_{i,k}$ is confined to a $(K-1)$-dimensional subspace, only $(K-1)$ elements of $\mathbf{v}_{i,k}$ are non-zero.

$\mathbf{\Phi}_{k-1}$ can be expressed as

$$\mathbf{\Phi}_{k-1} = \mathbf{I} - \mathbf{\Psi}_{k-1}, \tag{1.7}$$

where $\mathbf{\Psi}_{k-1}$ is a matrix of rank K, consisting of a weighted linear combination of projectors based on the spreading codes of $\mathbf{A}$ (see Appendix A for details). The $(N-K)$ zero eigenvalues of $\mathbf{\Psi}_{k-1}$ and the related eigenvectors correspond to the noise space of $\mathbf{A}$, in fact the eigenvectors span null $(\mathbf{A})$. However, $\mathbf{r}_s$ has no components in null$(\mathbf{A})$ so the corresponding elements of $\mathbf{v}_{i,k}$ are zero. $\mathbf{v}_{i,k}$ has therefore maximally $(K-1)$ non-zero elemements since $\mathbf{\Phi}_{k-1}$ has at least one eigenvalue equal to zero. These $(K-1)$ eigenvalues determine the convergence characteristics. We therefore have

$$v_{i,k}(n) = \lambda_{k-1}(n)v_{i-1,k}(n) = \lambda_{k-1}^{i-1}(n)v_{1,k}(n), \quad \text{for } n = 1, 2..., K-1,$$

where we have ordered the eigenvectors such that the first $(K-1)$ elements correspond to the relevant $(K-1)$ eigenvalues. If the $(K-1)$ eigenvalues are constrained by

$$|\lambda_{k-1}(n)| < 1 \quad \text{for} \quad n \in \{1, 2..., K-1\}, \tag{1.8}$$

the linear SIC structure is guaranteed to converge. It is however difficult to analytically show that this is always true. The Gershgorin circle theorem [7] does not provide a tight enough bound to ensure (1.8). During our investigation we have however only found violations to (1.8) for systems with two or more identical spreading codes. In Figure 1.3 we have generated a scatter plot of the eigenvalue location for a multitude of simulation experiements with $7 \leq N \leq 128$ and $4 \leq K \leq 128$. There are no eigenvalues beyond the boundaries of (1.8) and only a few corresponding to singular systems on the periphery. We therefore conclude that the linear SIC structure will converge for all practical cases. Letting

$$|\lambda_{k-1}(n)| = \exp\left(-\frac{1}{\tau_{k-1}(n)}\right) \Rightarrow \tau_{k-1}(n) = \frac{-1}{\ln\left(|\lambda_{k-1}(n)|\right)}, \tag{1.9}$$

we have

$$\begin{aligned}\frac{v_{i,k}(n)}{v_{1,k}(n)} &= |\lambda_{k-1}(n)|^{i-1}\exp(j(i-1)\phi_{k-1}(n)) \\ &= \exp\left(\frac{i-1}{\tau_{k-1}(n)}\right)\exp(j(i-1)\phi_{k-1}(n)),\end{aligned}$$

where $\exp(j(i-1)\phi_{k-1}(n))$ is the phase of the n^{th} complex eigenvalue. It follows that the absolute value of each element of $\mathbf{v}_{i,k}$ decreases exponentially with each

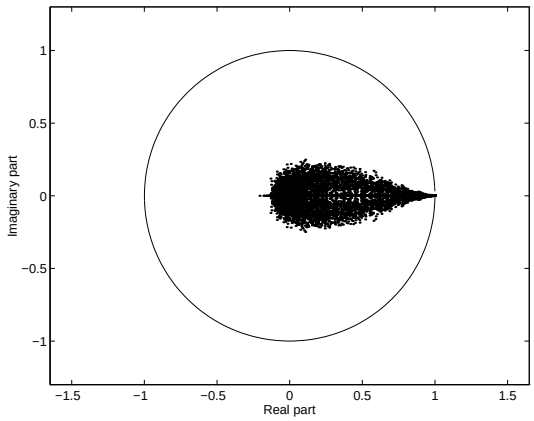

Figure 1.3 A scatter plot of the eigenvalue location for a multitude of simulation experiments with $7 \leq N \leq 128$ and $4 \leq K \leq 128$.

additional stage with a time constant of $\tau_{k-1}(n)$. A similar expression for $\mathbf{r}'_{i,k}$ is obtained from (1.6)

$$r'_{i,k}(j) = \sum_{n=1}^{N} u_n(j)\lambda_{k-1}^{i-1}(n)v_{1,k}(n). \tag{1.10}$$

Eqn. (1.10) shows that each element in the residual received vector $\mathbf{r}'_{i,k}$ converges as a weighted sum of exponentials. The time $\tau_{k-1}(n)$ required for each term to reach $1/e$ of its initial value is given by Eqn. (1.9). However, the overall time constant τ, defined as the time required for the summation term in Eqn. (1.10) to decay to $1/e$ of its initial value, cannot be expressed in a simple closed form. The slowest rate of convergence is attained when $u_n(j)v_{1,k}(n) = 0$ for all n except for the mode corresponding to the largest eigenvalue $\lambda_{\max}$. The fastest rate of convergence is attained when only $u_n(j)v_{1,k}(n) \neq 0$ for the n corresponding to the smallest eigenvalue $\lambda_{\min}$.

Accordingly, the overall time constant τ for any element in the residual received vector is bounded by

$$\frac{-1}{\ln(|\lambda_{\min}|)} \leq \tau \leq \frac{-1}{\ln(|\lambda_{\max}|)}.$$

For a large eigenvalue spread, the convergence rate is limited by the largest eigenvalue or the slowest rate of convergence.

It has been observed that $\mathbf{\Phi}_k$ and $\mathbf{\Phi}_j$ have the same eigenvalues, i.e.,

$$|\mathbf{\Phi}_k - \lambda\mathbf{I}| = |\mathbf{\Phi}_j - \lambda\mathbf{I}|, \quad \text{for all } k, j. \tag{1.11}$$

It is again difficult to show this analytically. However for $K = 2$ and $K = 3$, it is possible to identify terms and confirm (1.11). For $K > 3$ this observation is based on numerical experiments. This then means that all the users experience similar convergence characteristics, i.e., they have the same convergence rate.

For theoretical convergence, $\mathbf{r}'_{i,k} \to \mathbf{0}$ for $i \to \infty$. For practical convergence then $\mathbf{r}'_{i,k} \preceq \mathbf{0}$ for some finite i. For an arbitrary $\epsilon < 1$ we will consider convergence to have occured when

$$\|\mathbf{r}'_{i,k}\|^2 \leq \|\mathbf{r}'_{1,k}\|^2 \epsilon. \tag{1.12}$$

We will denote this behaviour as ϵ-convergence. Based on Eqn. (1.6)

$$\|\mathbf{r}'_{i,k}\|^2 = \left(\sum_{n=1}^{N} \mathbf{u}_n \lambda_{k-1}^{i-1}(n) v_{1,k}(n)\right)^{\mathrm{H}} \left(\sum_{j=1}^{N} \mathbf{u}_j \lambda_{k-1}^{i-1}(j) v_{1,k}(j)\right)$$

A condition for ϵ-convergence can be expressed as a function of the overall time constant τ. However, in order to relate the rate of ϵ-convergence to the number of stages, it is more convenient to consider the individual time constants $\tau_{k-1}(n)$. An upper bound on $\|\mathbf{r}'_{i,k}\|^2$ is obtained by letting $\lambda_{k-1}(n) = \lambda_{\max}$ for all n, i.e.,

$$\begin{aligned} \|\mathbf{r}'_{i,k}\|^2 &\leq |\lambda_{\max}|^{2(i-1)} \left(\sum_{n=1}^{N} \mathbf{u}_n v_{1,k}(n)\right)^{\mathrm{H}} \left(\sum_{j=1}^{N} \mathbf{u}_j v_{1,k}(j)\right) \\ &= |\lambda_{\max}|^{2(i-1)} \|\mathbf{r}'_{1,k}\|^2. \end{aligned} \tag{1.13}$$

Comparing (1.12) and (1.13) we can conclude that ϵ-convergence is obtained if

$$\begin{aligned} |\lambda_{\max}|^{2(i_\epsilon - 1)} &\leq \epsilon \\ \Rightarrow i_\epsilon &\geq \frac{\log \epsilon}{2 \log |\lambda_{\max}|} + 1. \end{aligned} \tag{1.14}$$

In general, a heavy loaded system has several eigenvalues close to one, while lightly loaded systems have smaller eigenvalues. A heavily loaded system therefore converges quite slowly as compared to a lightly loaded system. For example, if $|\lambda_{\max}| = 0.95$ and $\epsilon = 0.01$, $i_\epsilon = 46$ as compared to the case of $|\lambda_{\max}| = 0.5$ with a corresponding $i_\epsilon = 5$.

1.4 NUMERICAL RESULTS

A 5-user numerical example based on a symbol-synchronous system with perfect power control, using a randomly selected set of short codes of length 15 is considered here. The system is described by its cross-correlation matrix:

$$R = \frac{1}{15} \begin{bmatrix} 15 & -5 & -5 & -7 & -9 \\ -5 & 15 & 3 & 1 & 3 \\ -5 & 3 & 15 & 1 & 7 \\ -7 & 1 & 1 & 15 & 1 \\ -9 & 3 & 7 & 1 & 15 \end{bmatrix}.$$

In Figure 1.4 the performance of the 5-user system is shown. The performance is averaged over all users. The averaged performance of the conventional detector and the linear SIC scheme with 1,3 and 5 stages are presented. Also included

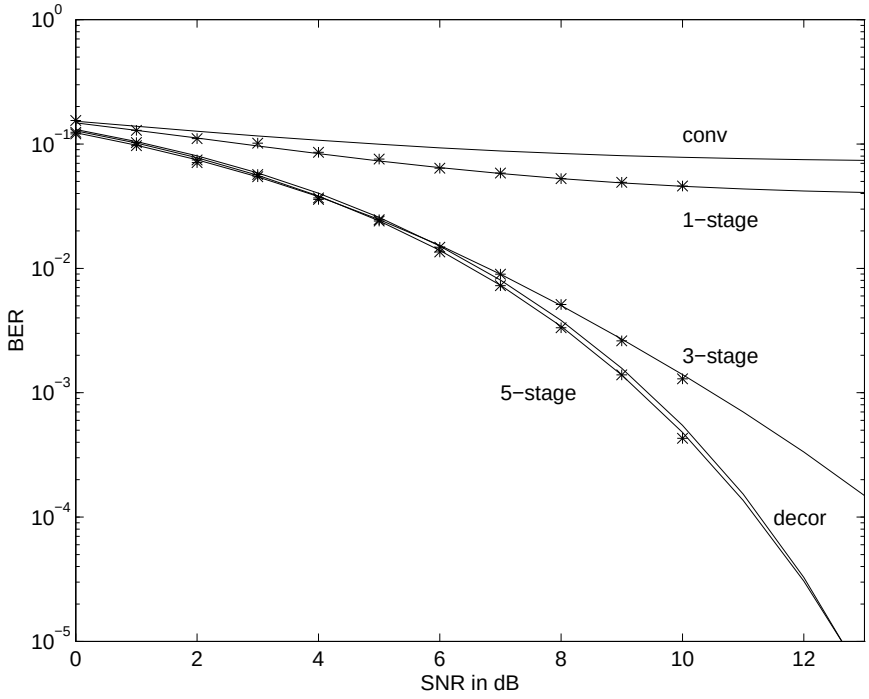

Figure 1.4 BER performance for the conventional detector and the linear SIC schemes with 1, 3 and 5 stages, respectively, using a randomly selected set of short codes.

is the performance of the decorrelating detector. Monte Carlo simulations are indicated with stars. In this case the largest eigenvalue influencing the convergence rate is $\lambda_{\max} = 0.5416$. With $\epsilon = 0.01$, we have ϵ-convergence for a 5-stage detector. This is clearly illustrated in Figure 1.4 as the performance of the 5-stage detector is virtually identical to the performance of the decorrelator.

In Figure 1.5 the averaged BER performance is presented as a function of the number of stages at an SNR of 5 dB and 15 dB for the 5-user case. All users are using the same number of stages for the curves labled "SIC". The corresponding decorrelator performance and the minimum achievable BER for linear SIC, when allowing each user to employ the corresponding optimal number of stages are also included in Figure 1.5. The optimal number of stages for the 5 users are $\{4, 5, 3, 4, 6\}$ at 5 dB and $\{7, 8, 7, 7, 10\}$ at 15 dB. With this example, it is illustrated that the performance is generally better at an intermediate number of stages as compared to the decorrelator. It is also clear that by letting users having a different number of cancellation stages, the performance is further improved. This effect is however diminishing for increasing SNR. All the observed improvements are in terms of only a fraction of a dB.

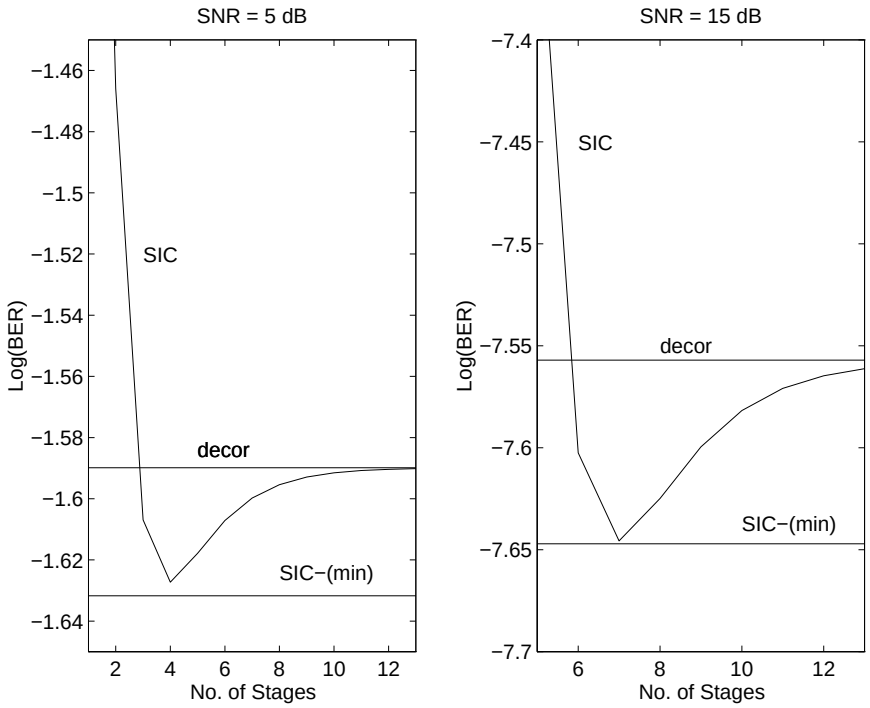

Figure 1.5 BER performance as a function of the number of stages for a 5-user system at an E_b/N_0 of 5 dB and 15 dB, respectively. The decorrelator performance and the minimum achievable BER for linear SIC are also included.

1.5 CONCLUDING REMARKS

In this paper we have used a matrix-algebraic approach to linear SIC to derive some quantitative statements concerning convergence in general and convergence rate in particular. It is demonstrated that the multi-stage linear SIC will converge to the decorrelating detector for all practical cases. The concept of ϵ-convergence is introduced as a tool for determining the number of stages necessary for practical convergence. It is observed that the BER does not generally decrease monotonically as the number of stages increase. In fact, the minimum overall BER is achieved prior to convergence. The optimum number of stages is determined by system load, correlation between users and the noise level and does not have to be the same for different users.

1.6 APPENDIX A

According to the diagonalisability theorem [7] $\mathbf{\Phi}_k$ must have N linearly independent eigenvectors for (1.5) to be valid. If the geometrical multiplicity is equal to the algebraic multiplicity for each eigenvalue, then $\mathbf{\Phi}_k$ is non-defective and has N linearly independent eigenvectors. We therefore consider the eigen-

problem for $\mathbf{\Phi}_k$. From (1.3) and (1.7)

$$\mathbf{\Phi}_k = \prod_{j=k}^{1} \left(\mathbf{I} - \mathbf{s}_j \mathbf{s}_j^\top\right) \prod_{m=K}^{k+1} \left(\mathbf{I} - \mathbf{s}_m \mathbf{s}_m^\top\right) = \mathbf{I} - \mathbf{\Psi}_k,$$

where

$$\mathbf{\Psi}_k = \sum_{i=k}^{1} \mathbf{s}_i \mathbf{s}_i^\top \prod_{m=i-1}^{1} \left(\mathbf{I} - \mathbf{s}_m \mathbf{s}_m^\top\right) \prod_{l=K}^{k+1} \left(\mathbf{I} - \mathbf{s}_l \mathbf{s}_l^\top\right) + \sum_{j=K}^{k+1} \mathbf{s}_j \mathbf{s}_j^\top \prod_{m=j-1}^{k+1} \left(\mathbf{I} - \mathbf{s}_m \mathbf{s}_m^\top\right).$$

Both summations are over decreasing indices so for $k = K$, the last term is zero. Furthermore, the eigenvalues of $\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$ are related through $\lambda_{\mathbf{\Phi}_k} = 1 - \lambda_{\mathbf{\Psi}_k}$. Let

$$\begin{aligned} \mathbf{b}_{k+1} &= \mathbf{s}_{k+1} \\ \mathbf{b}_{k+i} &= \mathbf{s}_{k+i} - \sum_{j=i-1}^{1} \mathbf{s}_{k+i}^\top \mathbf{s}_{k+j} \mathbf{b}_{k+j}, \end{aligned}$$

where the index is evaluated modulo K. Defining $\mathbf{B}_k = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_K)$, it is clear that $\mathbf{B}_k$ is nothing but a linear combination of $\mathbf{A}$, i.e., $\mathbf{B}_k = \mathbf{A}\mathbf{P}_k$ where

$$\begin{aligned} \mathbf{P}_k &= (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_K) \\ \mathbf{p}_{k+1} &= \left(\mathbf{0}_k^\top, 1, 0, \cdots, 0\right)^\top \\ \mathbf{p}_{k+i} &= \left(\mathbf{0}_{k+i-1}^\top, 1, 0, \cdots, 0\right)^\top - \sum_{j=i-1}^{1} \mathbf{s}_{k+i}^\top \mathbf{s}_{k+j} \mathbf{p}_{k+j}. \end{aligned}$$

$\mathbf{P}_K$ is an upper left triangular matrix while $\mathbf{P}_k$, $k \neq K$, is a column-shifted upper left triangular matrix. Based on $\mathbf{B}_k$ and $\mathbf{P}_k$ we can express

$$\mathbf{\Psi}_k = \mathbf{A}\mathbf{B}_k^\top = \mathbf{A}\mathbf{P}_k^\top \mathbf{A}^\top.$$

Lemma 1 $\mathrm{Rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B}_k) = \mathrm{rank}(\mathbf{\Psi}_k)$

Proof: Since $\mathbf{P}_k$ is a full rank matrix, $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B}_k)$. The non-zero eigenvalues of $\mathbf{\Psi}_k = \mathbf{A}\mathbf{B}_k^\top$ are the same as the non-zero eigenvalues of $\mathbf{A}^\top\mathbf{B}_k$ [7]. It then follows that:

$$\mathrm{rank}(\mathbf{\Psi}_k) = \mathrm{rank}(\mathbf{A}^\top\mathbf{B}_k) = \mathrm{rank}(\mathbf{A}^\top\mathbf{A}\mathbf{P}_k) = \mathrm{rank}(\mathbf{A}),$$

which concludes the proof. ∎

$\mathbf{\Psi}_K$ therefore has K non-zero eigenvalues determined as the non-zero eigenvalues of $\mathbf{A}^\top\mathbf{B}_k = \mathbf{A}^\top\mathbf{A}\mathbf{P}_k$. The remaining $(N - K)$ eigenvalues are equal to zero. These eigenvalues correspond to the noise space and the multiplicity is directly determined by $\mathrm{rank}(\mathbf{A})$. It is then clear that the algebraic multiplicity for the zero eigenvalue is equal to the geometric multiplicity. Let us consider the non-zero eigenvalues.

Lemma 2 *$\mathbf{s}_{k+1}$ is orthogonal to all $\mathbf{b}_i$, $i \neq k+1$.*

Proof: Let individual elements of $\mathbf{A}^\top\mathbf{B}_k$ be denoted by $\mathbf{A}^\top\mathbf{B}_k(i,j)$ and consider the elements on the $(k+1)^{\text{th}}$ row of $\mathbf{A}^\top\mathbf{B}_k$:

$$\begin{aligned}\mathbf{A}^\top\mathbf{B}_k(k+1,k+1) &= \mathbf{s}_{k+1}^\top\mathbf{b}_{k+1} = \mathbf{s}_{k+1}^\top\mathbf{s}_{k+1} = 1\\ \mathbf{A}^\top\mathbf{B}_k(k+1,k+2) &= \mathbf{s}_{k+1}^\top\mathbf{b}_{k+2} = \mathbf{s}_{k+1}^\top(\mathbf{s}_{k+2} - \mathbf{s}_{k+1}^\top\mathbf{s}_{k+2}\mathbf{s}_{k+1}) = 0.\end{aligned}$$

Now by induction, assuming that $\mathbf{s}_{k+1}^\top\mathbf{b}_{k+j} = 0$ for all $j = 2, 3...i-1$,

$$\mathbf{A}^\top\mathbf{B}_k(k+1,k+i) = \mathbf{s}_{k+1}^\top\mathbf{b}_{k+i} = \mathbf{s}_{k+1}^\top(\mathbf{s}_{k+i} - \mathbf{s}_{k+1}^\top\mathbf{s}_{k+i}\mathbf{s}_{k+1}) = 0,$$

where $i = 2, 3...K$ and the index is evaluated modulo K. ∎

It then follows that $\mathbf{\Psi}_k$ has at least one eigenvalue equal to one. Additional eigenvalues equal to one occur only when some orthogonal codes are included. It is however not straightforward to determine the multiplicity of this eigenvalue. Based on simulations for a multitude of N and K, it has been observed that when no users are orthogonal, the one eigenvalue is simple. When it is not simple, then in some cases $\mathbf{\Phi}_k$ is defective, especially if successive users are orthogonal. It is also a rather difficult task to determine the multiplicity behaviour for the remaining eigenvalues. For $K = 2$ and $K = 3$ it is possible to determine the eigenvalues explicitly using conventional techniques. It can then be shown that given a non-orthogonal system, the eigenvalues are simple and fulfill (1.8). For $K > 3$ it has been observed based on simulations that for such a system all remaining eigenvalues are simple. Assuming a non-orthogonal system we therefore conclude that $\mathbf{\Phi}_k$ is non-defective.

References

[1] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code–division multiple–access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, April 1990.

[2] P. Dent, B. Gudmundson and M. Ewerbring, "CDMA-IC: A novel code divisions multiple access scheme based on interference cancellation," in *Proc. 3rd IEEE Int. Symp. PIMRC '92*, pp. 98–102, Oct. 1992. Boston, Massachussets.

[3] M. Kawabe, S. Sato, H. Sugimoto, T. Sato and A. Fukasawa, "Interference Cancellation system using estimations of propagation parameters," in *Proc. JTC-CSCC '94*, pp. 173–178, 1994. Japan.

[4] L. K. Rasmussen, A.-L. Johansson and T. J. Lim, "One-shot filtering equivalence for linear successive interference cancellation in CDMA," in *Proc. IEEE VTC '97*, pp. 2163-2167, May 1996. Phoenix, Arizona.

[5] K. Jamal and E. Dahlman, "Multi-stage serial interference cancellation for DS-CDMA," in *Proc. IEEE VTC '96*, April 1996. Atlanta, Georgia.

[6] S. Haykin, *Adaptive Filter Theory.* Prentice–Hall, 3rd ed., 1996.

[7] G. W. Stewart and J. Sun, *Matrix Perturbation Theory.* Academic Press, 1990.