Performance of Reduced–Rank Equalization

Yakun Sun and Michael L. Honig

Dept. of ECE

2145 Sheridan Road

Northwestern University

Evanston, IL 60208

Abstract

We evaluate the performance of reduced-rank equalizers for both single-input, single-output (SISO) and multi-input/multi-output (MIMO) frequency-selective channels. Each equalizer filter is constrained to lie in a Krylov subspace, and can be implemented as a reduced-rank Multi-Stage Wiener Filter (MSWF). Both reduced–rank linear and decision-feedback equalizers are considered. Our results are asymptotic as the filter length goes to infinity. For SISO channels, the output Mean Squared Error is expressed in terms of the moments of the channel spectrum. For MIMO channels, both successive and parallel interference cancellation are considered. The asymptotic performance in that case requires the computation of moments, which depend on shifted versions of the channel impulse response for different users. Those are also expressed in terms of the MIMO channel frequency response. Numerical results are presented, which show that near full-rank performance can be achieved with relatively low-rank equalizers.

I. Introduction

Reduced–rank estimation and filtering has been studied for many applications, such as space-time processing, channel estimation, and equalization [1], [2]. Reduced-rank estimation can accurately approximate the filter parameters with relatively few observations or training data [3, Sec. 8.4]. In this paper, we apply reduced-rank estimation and filtering to equalization, and study the performance for suppression of both intersymbol interference (ISI) and co-channel interference (CCI).

In a reduced–rank filter, the input (observation) vectors are projected onto a lower dimensional subspace. The filter is then optimized within this subspace. If the subspace is selected

This work was supported by ARO under grant DAAD190310119 and NSF under grant CCR-0073686.

appropriately, this can reduce the amount of training, or observation data, needed to estimate the filter. Furthermore, once the subspace is determined, the computational complexity associated with estimating the filter coefficients decreases with the subspace dimension.

The performance of reduced-rank estimation depends on the choice of the projection subspace. A well-known class of reduced-rank methods projects the incoming signal onto a subspace spanned by a subset of eigenvectors of the input covariance matrix (e.g., [4]–[6]). When used for interference suppression, the subspace dimension must generally scale with the dimension of the signal subspace [7], which compromises the benefits of reduced-rank estimation as the load (number of interfering signals normalized by filter length) increases. Additional discussion about subspace decompositions used for reduced-rank estimation and filtering can be found in [5], [8]–[12].

Our focus is on reduced-rank algorithms, which are based on the Multi-Stage Wiener Filter (MSWF) [13]. In previous work, the MSWF has been applied to suppression of multiple access interference in a Code-Division Multiple Access (CDMA) system [7], [14], [15]. (See also [11], [12], which discuss a related ”Auxiliary Vector” reduced-rank method.) It has been shown in [7] that for a target performance, the MSWF can offer a dramatic reduction in subspace dimension relative to other reduced–rank techniques, such as principal components. This can lead to a significant reduction in training overhead relative to a full-rank algorithm [16]. Furthermore, it has been shown in [7] that the subspace associated with the MSWF is a Krylov space, for which the basis vectors are easily computed. In [17] it is shown that the output Signal-to-Interference Plus Noise Ratio (SINR) for the reduced-rank MSWF converges exponentially to the full-rank SINR as the rank increases.

Here we study the performance of the reduced-rank MSWF when used for channel equalization. Earlier work on reduced-rank equalization, based on an eigen-decomposition of the input covariance matrix, is presented in [18], [19]. The reduced-rank MSWF has been considered for equalization in [20]–[22], and a reduced-rank decision-feedback equalizer, based on the MSWF, is presented in [23], [24]. Adaptive reduced-rank estimation is applied to a turbo DFE in [25]. The reduced-rank DFE can provide a significant performance gain, relative to the conventional full-rank DFE, when training is limited.

Our main results characterize the asymptotic performance of reduced–rank linear and decision–feedback equalizers, where asymptotic means that the filter length tends to infinity with a fixed channel. The performance is measured in terms of Mean Square Error (MSE). Because an

infinite impulse response (IIR) equalizer can often be accurately approximated as a finite impulse response (FIR) equalizer, our results can be used to estimate the rank required for FIR equalizers. We consider both single-input/single-output and multiple-user/multiple-antenna channel models. In the latter case, both successive (S-) and parallel (P-) DFEs are considered, in which the users are decoded either successively or in parallel. In the case of the SDFE, we assume that *all* bits for the set of decoded users are available for feedback cancellation. (This corresponds to detection of a received packet.) The asymptotic performance is expressed in terms of the moments of the channel spectrum (for the linear equalizer), and inner products containing the channel spectrum and phase shifted versions of the channel frequency response (for the DFE). We also show that the geometric average of full-rank MMSE across users is the same for both the S- and PDFEs.

Numerical results are presented, which show the asymptotic performance of the reduced–rank equalizers vs. rank. In all cases considered the reduced–rank equalizers can achieve close to full-rank performance when the subspace dimension, or rank, is at least ten.

In Sec. II, the single-user/single antenna system is considered, and we evaluate the minimum MSE (MMSE) for both linear and decision-feedback equalizers. In Sec. III, the results are generalized to the multiple-user/multiple-antenna model with linear, S- and PDFEs. Section IV concludes the paper, and proofs are given in the appendices.

II. Reduced-Rank Equalization

We first present reduced-rank linear and decision-feedback equalizers for the single-user/single-antenna system model, and evaluate the asymptotic performance as the filter length tends to infinity. In the next section we consider the more complicated multi-user/multi-antenna scenario.

A. Single-Input/Single-Output (SISO) Model

In this case, we form the vector of $2N_c + 1$ received signal samples

$$\begin{aligned} \mathbf{y}(i) &= [y(i+N_c), \cdots, y(i), \cdots, y(i-N_c)]^T \\ &= \mathbf{H}\mathbf{s}(i) + \mathbf{n}(i) \end{aligned} \tag{1}$$

where $[\cdot]^T$ denotes transpose, $\mathbf{H}$ is the $(2N_c+1) \times (2N_c+2m+1)$ Toeplitz channel matrix

$$\mathbf{H} = \begin{bmatrix} h(-m) & \cdots & h(m) & 0 & \cdots & 0 \\ 0 & h(-m) & \cdots & h(m) & \cdots & 0 \\ & & \ddots & & \ddots & \\ 0 & \cdots & 0 & h(-m) & \cdots & h(m) \end{bmatrix} \tag{2}$$

corresponding to a channel impulse response with length $2m+1$, $\mathbf{s}(i) = [s(i+m+N_c), \cdots, s(i), \cdots, s(i-m-N_c)]^T$ is a $(2N_c+2m+1)\times 1$ vector of transmitted symbols, and $\mathbf{n}(i) = [n(i+N_c), \cdots, n(i), \cdots, n(i-N_c)]^T$ is complex white Gaussian noise with covariance $\sigma^2\mathbf{I}$.

For simplicity, we assume the transmitted symbols are with unit power, i.e., $E\left[|s(i)|^2\right] = 1$. The Minimum Mean Square Error (MMSE) linear equalizer is given by

$$\mathbf{c} = \arg\min_{\mathbf{c}} \; E\left[\left|s(i) - \mathbf{c}^\dagger \mathbf{y}(i)\right|^2\right] = \mathbf{R}^{-1}\mathbf{h} \tag{3}$$

where $[\cdot]^\dagger$ denotes conjugate transpose, $\mathbf{R}$ is the input covariance matrix

$$\begin{aligned} \mathbf{R} &= E\left[\mathbf{y}(i)\mathbf{y}^\dagger(i)\right] \\ &= \mathbf{H}\mathbf{H}^\dagger + \sigma^2\mathbf{I} \end{aligned} \tag{4}$$

and $\mathbf{h}$ is the channel steering vector,

$$\begin{aligned} \mathbf{h} &= E\left[\mathbf{y}(i)s^*(i)\right] \\ &= \mathbf{H}_{m+N_c+1} = [h(N_c), h(N_c-1), \cdots, h(-N_c)]^T \end{aligned} \tag{5}$$

where $\mathbf{M}_n$ is the n^{th} column of the matrix $\mathbf{M}$, and $h(i) = 0$ for $|i| > m$. In what follows, we allow the length of the impulse response, represented by m, to be infinite. When the channel is not known at the receiver, the covariance matrix and the channel steering vector can be estimated adaptively from a sequence of training symbols and the corresponding received signal vectors [25].

B. Reduced–Rank (RR) Linear Equalization

An RR equalizer can potentially reduce the training overhead and complexity associated with adaptive equalization. This becomes an important issue with frequency-selective Multiple-Input/Multiple-Output (MIMO) channels, since the number of equalizer coefficients increases with the product of number of transmit antennas, number of receive antennas, and the channel delay spread.

The basic idea in RR filtering is to project the received vectors onto a lower–dimensional subspace. Let $\mathbf{S}_D$ be the $(2N_c+1)\times D$ matrix with columns that are the basis vectors for a D–dimensional subspace, and let $\tilde{\mathbf{y}}(i) = \mathbf{S}_D^\dagger \mathbf{y}(i)$ be the projected input. In what follows, all D–dimensional (projected) quantities are denoted with a "tilde". The sequence of projected received vectors is the input to a tapped–delay line filter, represented by the D–vector $\tilde{\mathbf{c}}$. Selecting $\tilde{\mathbf{c}}$ to minimize the RR MSE $E\{|s(i) - \tilde{\mathbf{c}}^\dagger\tilde{\mathbf{y}}(i)|^2\}$ gives

$$\tilde{\mathbf{c}} = (\mathbf{S}_D^\dagger \mathbf{R}\mathbf{S}_D)^{-1}\mathbf{S}_D^\dagger \mathbf{h}. \tag{6}$$

In what follows, we take

$$\mathbf{S}_D = \left[\begin{array}{ccccc} \mathbf{h} & \mathbf{Rh} & \mathbf{R}^2\mathbf{h} & \cdots & \mathbf{R}^{D-1}\mathbf{h} \end{array} \right] \tag{7}$$

which defines a Krylov subspace. It has been shown in [7] that the associated RR filter is equivalent to the MSWF [13]. Additional discussion on related RR methods is given in [26]. Analogous to the results in [7], [14], which apply to interference suppression for CDMA, our analysis and numerical results show that the performance of the RR linear equalizer can approach full–rank performance with low rank, which does not increase with the equalizer length. Reducing the filter rank D can reduce the amount of training needed in an adaptive mode.

C. Performance of the RR Linear Equalizer

The MMSE for the full-rank linear equalizer is given by

$$\mathcal{M}_{LE} = 1 - \mathbf{h}^\dagger \mathbf{R}^{-1} \mathbf{h} \tag{8}$$

where $\mathbf{R}$ and $\mathbf{h}$ are defined by (4) and (5), respectively. The MSE for the RR linear equalizer can be expressed in terms of the channel moments $\gamma_n = \mathbf{h}^\dagger \mathbf{R}^n \mathbf{h}$. Let

$$\bar{\gamma}_D = [\gamma_0, \gamma_1, \cdots, \gamma_{D-1}]^T; \tag{9}$$

$$\Gamma_D = \left[\begin{array}{cccc} \gamma_1 & \gamma_2 & \cdots & \gamma_D \\ \gamma_2 & \gamma_3 & \cdots & \gamma_{D+1} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_D & \gamma_{D+2} & \cdots & \gamma_{2D-1} \end{array} \right] \tag{10}$$

The MMSE for the RR linear equalizer is

$$\begin{aligned} \mathcal{M}_{RRLE} &= 1 - \mathbf{h}^\dagger \mathbf{S}_D (\mathbf{S}_D^\dagger \mathbf{R} \mathbf{S}_D)^{-1} \mathbf{S}_D^\dagger \mathbf{h} \\ &= 1 - \bar{\gamma}_D^\dagger \Gamma_D^{-1} \bar{\gamma}_D \end{aligned} \tag{11}$$

The preceding result assumes that the equalizer has $2N_c+1$ taps. We now present the analogous result as $N_c \to \infty$. In that case, the MMSE for the RR linear equalizer can be computed from (11), where the moments γ_n, $n = 0, \cdots, 2D-1$, are replaced by the asymptotic moments $\gamma_n^\infty = \lim_{N_c \to \infty} \gamma_n$.

Theorem 1: The asymptotic moments are given by

$$\gamma_n^\infty = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(|H(e^{j\omega})|^2 + \sigma^2 \right)^n |H(e^{j\omega})|^2 d\omega \tag{12}$$

Proof: The covariance matrix of the received vector, $\mathbf{R}$ defined in (4), is an $N \times N$ Toeplitz matrix with elements $\{R_i\}_{i=-(N-1),\cdots,N-1}$, where $N = 2N_c + 1$. The sequence $\{R_i\}$ is the autocorrelation of the channel impulse response, $R_i = \sum_{j=-m}^{m} h(j)h^*(j+i) + \sigma^2\delta_i$, where $\delta_i = 1$ for $i = 0$ and is zero otherwise, and $R_i = 0$ when $|i| > 2m$. We have

$$\begin{aligned}
\gamma_n^{\infty} &= \lim_{N\to\infty} \mathbf{h}^{\dagger}\mathbf{R}^n\mathbf{h} \\
&= \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}[\mathbf{H}^{\dagger}\mathbf{R}^n\mathbf{H}] \\
&= \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}[\mathbf{R}^n(\mathbf{R}-\sigma^2\mathbf{I})] \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\lambda_i^{n+1}(\mathbf{R}) - \sigma^2\lambda_i^n(\mathbf{R})\right) \qquad (13)
\end{aligned}$$

where $\lambda_i(\mathbf{R})$, $i = 1, \cdots, N$ are the eigenvalues of $\mathbf{R}$. Since $\mathbf{R}$ is a Toeplitz matrix, it follows from [27, Theorem 4.1] (restated as Theorem 9 in Appendix A) that

$$\begin{aligned}
\gamma_n^{\infty} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S^{n+1}(e^{j\omega})d\omega - \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} S^n(e^{j\omega})d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(|H(e^{j\omega})|^2 + \sigma^2\right)^n |H(e^{j\omega})|^2 d\omega
\end{aligned}$$

where $S(e^{j\omega}) = \sum_{k=-2m}^{2m} R_k e^{-jk\omega}$ is the power spectrum of the received signal. ■

As $D \to \infty$, the MMSE for the RR equalizer converges to the full-rank MMSE given by [28]

$$\mathcal{M}_{LE}^{\infty} = 1 - \lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\{\mathbf{H}^{\dagger}\mathbf{R}^{-1}\mathbf{H}\} = 1 - \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|H(e^{j\omega})|^2}{|H(e^{j\omega})|^2 + \sigma^2}d\omega \qquad (14)$$

A direct proof of (14) is given in Appendix A.

D. RR Decision–Feedback Equalizer

We now apply RR filtering to a DFE. Given the discrete-time model (1), for purposes of analysis we define the output of the DFE as the $2N_b \times 1$ vector

$$\mathbf{z}(i) = \mathbf{F}^{\dagger}\mathbf{y}(i) - \mathbf{B}^{\dagger}\mathbf{s}(i) \qquad (15)$$

where the feedforward filter $\mathbf{F}$ is a $(2N_c + 1) \times (2N_b + 1)$ matrix, and the feedback filter $\mathbf{B}$ is a $(2N_b + 1) \times (2N_b + 1)$ matrix. Letting $N_b = N_c + m$, the vector of feedback symbols $\mathbf{s}(i) = [s(i + N_b), s(i + N_b - 1), \cdots, s(i - N_b)]^T$. The feedback matrix $\mathbf{B}$ is strictly lower triangular, corresponding to causal feedback. In what follows, the feedback decisions are assumed to be correct. We wish to select $\mathbf{F}$ and $\mathbf{B}$ to minimize $E[\|\mathbf{z}(i) - \mathbf{s}(i)\|^2]$, which implies that the $(N_b+1)^{st}$ column of $\mathbf{F}$, and the $(N_b+1)^{st}$ column of $\mathbf{B}$ are selected to minimize $E[|z(i)-s(i)|^2]$.

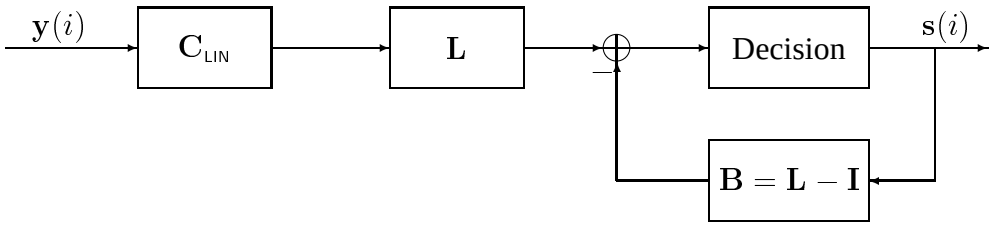

Fig. 1: Block diagram of an optimal (MMSE) DFE.

The columns of $\mathbf{F}$ and $\mathbf{B}$ can be interpreted as a sequence of time-varying filters, which are used to detect the sequence of symbols $s(i - N_b), \cdots, s(i + N_b)$.

The optimal feedforward filter can be expressed as the concatenation of the linear equalizer with an error whitening filter [29], as shown in Figure 1. Specifically, let $\mathbf{C}_{\mathsf{LIN}} = \mathbf{R}^{-1}\mathbf{H}$, and let the corresponding error be

$$\mathbf{e}(i) = \mathbf{s}(i) - \mathbf{C}_{\mathsf{LIN}}^{\dagger}\mathbf{y}(i) \tag{16}$$

which has covariance matrix

$$\begin{aligned} \mathbf{R}_e &= E[\mathbf{e}(i)\mathbf{e}^{\dagger}(i)] = \mathbf{I} - \mathbf{C}_{\mathsf{LIN}}^{\dagger}\mathbf{H} - \mathbf{H}^{\dagger}\mathbf{C}_{\mathsf{LIN}} + \mathbf{C}_{\mathsf{LIN}}^{\dagger}\mathbf{R}\mathbf{C}_{\mathsf{LIN}} & (17)\\ &= \mathbf{I} - \mathbf{H}^{\dagger}\mathbf{R}^{-1}\mathbf{H} & (18) \end{aligned}$$

The feedback matrix can then be computed from the Cholesky decomposition $\mathbf{L}^{\dagger}\mathbf{R}_e\mathbf{L} = \mathbf{D}_e$ where $\mathbf{L}$ is a lower triangular matrix with ones along the diagonal. The optimized feedforward filter is

$$\mathbf{F} = \mathbf{C}_{\mathsf{LIN}}\mathbf{L} \tag{19}$$

and $\mathbf{B} = \mathbf{L} - \mathbf{I}$.

One version of a RR DFE can be obtained by substituting RR approximations for the full-rank filters $\mathbf{C}_{\mathsf{LIN}}$ and $\mathbf{B}$. That is, the $(N_b + 1 + q)^{th}$ column of $\mathbf{C}_{\mathsf{LIN}}$, $-N_b \leq q \leq N_b$, is replaced by the RR approximation

$$\mathbf{c}_q = \mathbf{S}_{D,q}(\mathbf{S}_{D,q}^{\dagger}\mathbf{R}\mathbf{S}_{D,q})^{-1}\mathbf{S}_{D,q}^{\dagger}\mathbf{h}_q \tag{20}$$

where $\mathbf{S}_{D,q} = [\mathbf{h}_q, \mathbf{R}\mathbf{h}_q, \cdots, \mathbf{R}^{D-1}\mathbf{h}_q]$, and $\mathbf{h}_q = \mathbf{H}_{N_b+1+q}$, denotes the *shifted* channel impulse response. A full-rank feedback filter $\mathbf{B}$ can be computed by performing a Cholesky decomposition of the associated error covariance matrix given by (17). Note that the error covariance matrix is no longer given by (18) with a RR approximation for $\mathbf{C}_{\mathsf{LIN}}$. We refer to this RR DFE as the *RR DFE-Linear (RR DFE–L)*, since only the linear equalizer part of the feedforward filter is approximated by a RR filter.

To obtain a RR feedback filter, we recall that the $(N_b+1)^{\text{st}}$ column of $\mathbf{B}$ can be interpreted as an error prediction filter for the sequence of errors, $\left\{e(i) = s(i) - [\mathbf{C}_{\mathsf{LIN}}]^{\dagger}_{N_b+1}\mathbf{y}(i)\right\}$. That is, the column contains the non–zero coefficients $v_1, \cdots, v_{N_b}$, which are selected to minimize $E[|e(i) - \sum_{j=1}^{N_b} v_j e(i-j)|^2] = E[|e(i) - \mathbf{v}^{\dagger}\mathbf{e}_{N_b+2:2N_b+1}(i)|^2]$, where the vector $\mathbf{m}_{m:n}$ contains the m^{th} to n^{th} components of $\mathbf{m}$. The full-rank feedback filter is

$$\mathbf{v} = \check{\mathbf{R}}_e^{-1}\check{\mathbf{h}}_e \tag{21}$$

where

$$\check{\mathbf{R}}_e = E[\mathbf{e}_{N_b+2:2N_b+1}(i)\mathbf{e}^{\dagger}_{N_b+2:2N_b+1}(i)] = [\mathbf{R}_e]_{N_b+2:2N_b+1} \tag{22}$$

$$\check{\mathbf{h}}_e = E[e^*(i)\mathbf{e}_{N_b+2:2N_b+1}(i)] \tag{23}$$

and the matrix $\mathbf{M}_{m:n}$ contains the m^{th} to n^{th} columns and rows of $\mathbf{M}$. The RR feedback filter with rank D_b is given by

$$\tilde{\mathbf{v}} = \tilde{\check{\mathbf{R}}}_e^{-1}\tilde{\check{\mathbf{h}}}_e = (\mathbf{S}_e^{\dagger}\check{\mathbf{R}}_e\mathbf{S}_e)^{-1}\mathbf{S}_e^{\dagger}\check{\mathbf{h}}_e \tag{24}$$

where the matrix of basis vectors

$$\mathbf{S}_e = [\check{\mathbf{h}}_e, \check{\mathbf{R}}_e\check{\mathbf{h}}_e, \cdots, \check{\mathbf{R}}_e^{D_b-1}\check{\mathbf{h}}_e] \tag{25}$$

and the RR approximation of the non-zero components of the feedback filter is $\mathbf{S}_e(\mathbf{S}_e^{\dagger}\check{\mathbf{R}}_e\mathbf{S}_e)^{-1}\mathbf{S}_e^{\dagger}\check{\mathbf{h}}_e$. We will refer to the DFE with both RR linear and error estimation filters as the *RR DFE*.

We remark that when the channel is unknown at the receiver, a Least Squares (LS) version of the RR DFE is obtained by replacing the MMSE linear equalizer in (16) with the LS version, given a training sequence. The estimated feedback filter $\mathbf{v}$ is an LS error prediction filter, again given by (21) or (24), where the estimated linear prediction error covariance matrix and error steering vector are given by the corresponding sample averages [25].

E. RR DFE Performance

From the discussion in Section II-D, to compute the DFE filters and the associated performance, we must compute the error covariance matrix $\mathbf{R}_e$. The correlation among errors at the output of the RR linear equalizer is defined in terms of *shifted* channel moments $\gamma_{D,q} = \mathbf{h}^{\dagger}\mathbf{R}^D\mathbf{h}_q$, where $\mathbf{h}_q$ is the shifted channel impulse response.

Theorem 2: The asymptotic shifted channel moments are given by

$$\begin{aligned}\gamma_{n,q}^{\infty} &= \lim_{N\to\infty} \mathbf{h}^{\dagger}\mathbf{R}^n\mathbf{h}_q \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(|H(e^{j\omega})|^2 + \sigma^2\right)^n |H(e^{j\omega})|^2 e^{jq\omega} d\omega \end{aligned} \tag{26}$$

When the channel impulse response is finite-length, $\gamma_{n,q}^{\infty} = 0$ for $q > 2m(n+1)$.

The proof is given in Appendix B.

We now proceed to compute the asymptotic MSE for the RR DFE by computing the correlation between output errors for a rank-D linear equalizer. That is, the error at time $i-q$ is $e(i-q) = s(i-q) - [\mathbf{C}_{\mathsf{LIN}}]_{N_b+1+q}^{\dagger}\mathbf{y}(i)$, and

$$\begin{aligned} \varepsilon_q &= E\{e(i)e^*(i-q)\} \\ &= \delta_q - 2\bar{\gamma}_D^{\dagger}\Gamma_D^{-1}\bar{\gamma}_{D,q} + \bar{\gamma}_D^{\dagger}\Gamma_D^{-1}\Gamma_{D,q}\Gamma_D^{-1}\bar{\gamma}_D \end{aligned} \tag{27}$$

for $q \leq Q$, where $Q = 4mD$, $\bar{\gamma}_D$ and Γ_D are defined in (9)-(10),

$$\bar{\gamma}_{D,q} = [\gamma_{0,q}, \gamma_{1,q}, \cdots, \gamma_{D-1,q}]^T \tag{28}$$

$$\Gamma_{D,q} = \begin{bmatrix} \gamma_{1,q} & \gamma_{2,q} & \cdots & \gamma_{D,q} \\ \gamma_{2,q} & \gamma_{3,q} & \cdots & \gamma_{D+1,q} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{D,q} & \gamma_{D+1,q} & \cdots & \gamma_{2D-1,q} \end{bmatrix} \tag{29}$$

and $\varepsilon_q = 0$ for $q > Q$.

The $N_b \times N_b$ error covariance matrix is Toeplitz, i.e.,

$$\mathbf{R}_e = \begin{bmatrix} \varepsilon_0 & \varepsilon_1 & \varepsilon_2 & \cdots \\ \varepsilon_1^* & \varepsilon_0 & \varepsilon_1 & \cdots \\ \varepsilon_2^* & \varepsilon_1^* & \varepsilon_0 & \cdots \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix} \tag{30}$$

and the N_b-vector of errors is

$$\mathbf{h}_e = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{N_b}]^{\dagger} \tag{31}$$

The asymptotic MSE of the RR DFE-L can be computed from a Cholesky decomposition of the covariance matrix of the sequence of linear equalizer errors. Alternatively, using the fact that the output MSE is the mean squared prediction error for this sequence, we can write

$$\mathcal{M}_{RRDFE-L}^{\infty} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S_{\varepsilon}(e^{j\omega})d\omega\right\} \tag{32}$$

where $S_{\varepsilon}(e^{j\omega}) = \sum_{q=-\infty}^{\infty} \varepsilon_q e^{-jq\omega}$ is the power spectrum of the error sequence, and $\varepsilon_{-q} = \varepsilon_q^*$.

We now evaluate the output MSE for the RR DFE in which both the linear and error estimation filters are replaced by the corresponding RR approximations. Here we must assume a *finite-length* channel impulse response. Let

$$\psi_n = \check{\mathbf{h}}_e^{\dagger}\check{\mathbf{R}}_e^{n}\check{\mathbf{h}}_e \tag{33}$$

which we will refer to as the n^{th} moment of the error covariance matrix, where $\check{\mathbf{R}}_e$ and $\check{\mathbf{h}}_e$ are defined by (22) and (23). The MSE for the RR DFE is given by the analogous form to (11),

$$\begin{aligned}\mathcal{M}_{RRDFE} &= \min E\{|e(i) - \tilde{\mathbf{v}}^\dagger \tilde{\mathbf{e}}_{N_b+2:2N_b+1}(i)|^2\} \\ &= E\{|e(i)|^2\} - \check{\mathbf{h}}_e^\dagger \mathbf{S}_e (\mathbf{S}_e^\dagger \check{\mathbf{R}}_e \mathbf{S}_e)^{-1} \mathbf{S}_e^\dagger \check{\mathbf{h}}_e \\ &= \mathcal{M}_{RRLE} - \bar{\psi}_{D_b}^\dagger \Psi_{D_b}^{-1} \bar{\psi}_{D_b} \end{aligned} \tag{34}$$

where

$$\bar{\psi}_{D_b} = [\psi_0, \psi_1, \cdots, \psi_{D_b-1}]^T \tag{35}$$

and

$$\Psi_{D_b} = \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_{D_b} \\ \psi_2 & \psi_3 & \cdots & \psi_{D_b+1} \\ \vdots & \vdots & \vdots & \vdots \\ \psi_{D_b} & \psi_{D_b+1} & \cdots & \psi_{2D_b-1} \end{bmatrix} \tag{36}$$

Given a finite-length channel impulse response, the autocorrelation of the error at the output of the linear filter is zero after a finite time shift. Note that $\check{\mathbf{R}}_e$ and $\check{\mathbf{h}}_e$ have exactly the same structure as $\mathbf{R}_e$ and $\mathbf{h}_e$, the only difference being the effective length of the feedback filter. Specifically, $\mathbf{R}_e$ and $\mathbf{h}_e$ with a feedback filter of length N_b are the same as $\check{\mathbf{R}}_e$ and $\check{\mathbf{h}}_e$ with a feedback filter of length $2N_b + 1$. Therefore if the size of the feedback filter is large enough, then we can simply use $\mathbf{R}_e$ and $\mathbf{h}_e$ to compute the moments of the error covariance matrix. That is, we can write $\psi_n = \mathbf{h}_e^\dagger \mathbf{R}_e^n \mathbf{h}_e$, since $\mathbf{h}_e$ has a finite number of nonzero components at the top, which does not increase with the filter length. This enables us to compute the performance of the RR DFE as $N \to \infty$ and $N_b \to \infty$ by evaluating the performance for infinite N but finite N_b.

Theorem 3: With a finite-length channel impulse response the asymptotic error moment can be evaluated as

$$\psi_n^\infty = \lim_{(N,N_b)\to\infty} \psi_n = \mathbf{h}_{e,M}^\dagger \mathbf{R}_{e,M}^n \mathbf{h}_{e,M} \tag{37}$$

where $M \geq 1 + \lceil \frac{n+1}{2} \rceil Q$, $\mathbf{h}_{e,M} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_Q, 0, \cdots, 0]_{M\times 1}^\dagger$, and $\mathbf{R}_{e,M} = [\mathbf{R}_e]_{1:M}$.

Proof: It is easily observed that for a Hermitian Toeplitz matrix $\mathbf{R}$ with $2Q+1$ non–zero diagonal elements, the $Q \times Q$ upper–left block of $\mathbf{R}^n$ remains constant when the size of $\mathbf{R}$ is larger than $\lceil \frac{n+1}{2} \rceil Q$. Since only the first Q entries of $\mathbf{h}_e$ may be nonzero, the asymptotic error moment can be computed from a finite–size $M \times M$ matrix $\mathbf{R}_{e,M}$ provided that $M \geq 1 + \lceil \frac{n+1}{2} \rceil Q$.

■

As the ranks of both feedforward and feedback filters tend to infinity, the MSE converges to the classical expression for the full-rank MMSE [28],

$$\mathcal{M}_{DFE}^{\infty} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left[\frac{\sigma^2}{|H(e^{j\omega})|^2+\sigma^2}\right] d\omega\right\} \tag{38}$$

A direct proof that the MMSE for the finite–length DFE converges to (38) is given in Appendix A.

F. Computation of Channel Moments

The expressions (12) and (26) can be computed via the residue theorem, i.e.,

$$\begin{aligned}\gamma_{n,q}^{\infty} &= \frac{1}{2\pi j}\oint_{|z|=1} \left(|H(z)|^2+\sigma^2\right)^n |H(z)|^2 z^q \frac{1}{z} dz \\ &= \sum \text{Residues}\left\{(|H(z)|^2+\sigma^2)^n |H(z)|^2 z^{q-1}\right\}\end{aligned} \tag{39}$$

For example, if $H(z) = \frac{1-az^{-1}}{\sqrt{1+|a|^2}}$, then the asymptotic channel moments, corresponding to $q=0$, can be explicitly expressed as

$$\gamma_n^{\infty} = T_{n+1} - \sigma^2 T_n \tag{40}$$

where

$$\begin{aligned}T_n &= \sum \text{Residue}\left\{(|H(z)|^2+\sigma^2)^n z^{-1}\right\} \\ &= \sum_{i=0}^{\lfloor n/2\rfloor} \frac{n!}{(n-2i)!(i!)^2}(1+\sigma^2)^{n-2i}\left(\frac{|a|}{1+|a|^2}\right)^{2i}\end{aligned} \tag{41}$$

and the residues in the first equality are evaluated by expanding the argument in powers of z and keeping the term which multiplies z^{-1}.

From (41) we see that the performance of the linear filter only depends on σ^2 and $|a|$, the distance of the zero to the origin. The shifted channel moments can be obtained as

$$\gamma_{n,q}^{\infty} = T_{n+1,q} - \sigma^2 T_{n,q} \tag{42}$$

where

$$T_{n,q} = \left(-\frac{1+|a|^2}{a^*}\right)^q \sum_{i=q}^{\lfloor (n+q)/2\rfloor} \frac{n!}{(n+q-2i)!i!(i-q)!}(1+\sigma^2)^{n+q-2i}\left(\frac{|a|}{1+|a|^2}\right)^{2i} \tag{43}$$

for $q \le n$ and $T_{n,q} = 0$ for $q > n$. Note that $T_{n,q}$ depends on the phase of the channel zero, as well as the magnitude.

We also observe from (12) and (26) that $\{\gamma_{n,q}\}$ is the inverse Discrete Fourier Transform of $(|H(e^{j\omega})|^2+\sigma^2)^n|H(e^{j\omega})|^2$ and $\gamma_n = \gamma_{n,0}$. Hence given the channel frequency response, the inverse FFT can be used to compute the moments efficiently.

G. Numerical Results

Fig. 2 shows the performance of RR equalizers for the channel $H(z) = (1 - az^{-1})/\sqrt{1 + |a|^2}$. In this case, the channel moments are given by (41) and (43). The figure shows plots of MSE vs. $|a|$ for different ranks with SNR$= 20$ dB. The solid and dashed curves are for the RR linear equalizer and RR DFE-L, respectively. As $|a| \rightarrow 1$, the impulse response of the full-rank equalizer effectively increases in length, and the performance degrades. These plots show that a RR linear equalizer and a RR DFE–L with rank 8 can achieve near full-rank linear equalizer and DFE performance respectively, independent of $|a|$.

Fig. 3 shows MSE vs. rank for the RR linear equalizer and RR DFE-L, and also for the RR DFE. For the RR DFE, each curve corresponds to a particular rank for the feedback filter, and shows MSE vs. rank for the linear equalizer part of the feedforward filter. These results were generated using a specific realization of an 11–tap Rayleigh fading channel with equal powers across taps. The SNR is 20 dB. The results show that in each case, a relatively low rank is needed to achieve near full-rank performance.

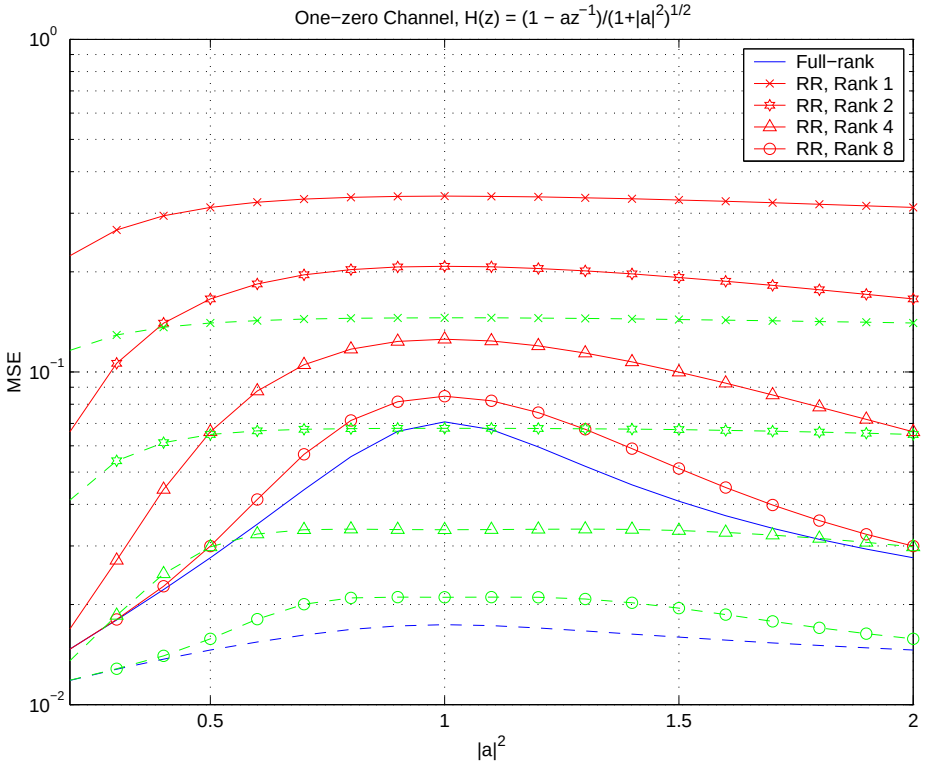

Fig. 2: Performance of the RR linear equalizer and RR DFE-L with a one-zero channel. The solid curves correspond to the linear equalizer, and the dashed curves correspond to the DFE.

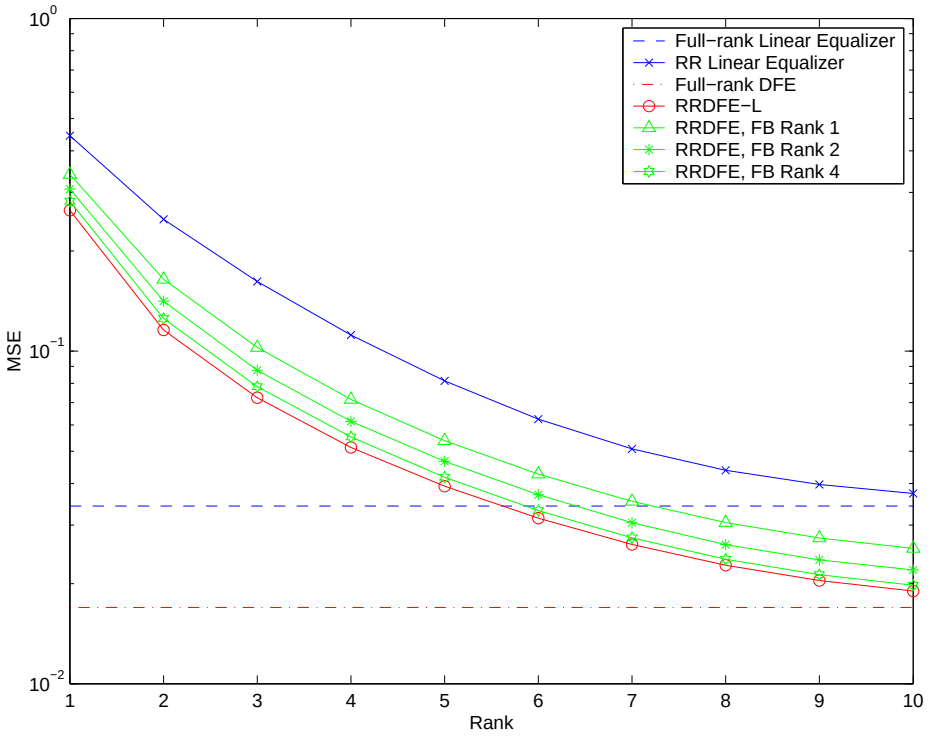

Fig. 3: MMSE vs. rank for RR linear and decision–feedback equalizers.

III. DFE WITH MULTIPLE USERS AND RECEIVE ANTENNAS

We now extend the preceding results to $K \geq 1$ users and $L \geq 1$ receive antennas. Let $\mathbf{y}_l(i)$ be the vector of received samples on antenna l at time i. Then we have

$$\begin{aligned} \mathbf{y}(i) &= [\mathbf{y}_1^T(i), \cdots, \mathbf{y}_L^T(i)]^T \qquad (44) \\ &= [\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_K] \begin{bmatrix} \mathbf{s}_1(i) \\ \vdots \\ \mathbf{s}_K(i) \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1(i) \\ \vdots \\ \mathbf{n}_L(i) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,K} \\ & \vdots & \\ \mathbf{H}_{L,1} & \cdots & \mathbf{H}_{L,K} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1(i) \\ \vdots \\ \mathbf{s}_K(i) \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1(i) \\ \vdots \\ \mathbf{n}_L(i) \end{bmatrix} \qquad (45) \end{aligned}$$

where $\mathbf{H}_{l,k}$ is the channel from user k to antenna l, as defined in (2), and $\mathbf{s}_k(i)$ contains the symbols for user k. That is, $\mathbf{y}(i)$ is obtained by stacking the received vectors across antennas. The channel frequency response, $\mathcal{H}(e^{j\omega})$, is given by a $L \times K$ matrix. That is,

$$\mathcal{H}(e^{j\omega}) = \left[\mathcal{H}_1(e^{j\omega}), \cdots, \mathcal{H}_K(e^{j\omega})\right] \qquad (46)$$

where $\mathcal{H}_k(e^{j\omega}) = [H_{1,k}(e^{j\omega}) \cdots H_{L,k}(e^{j\omega})]^T$, and $H_{l,k}(e^{j\omega})$ is the frequency response of the channel from user k to antenna l.

A. Linear Equalizer

For the full-rank linear equalizer, the asymptotic MMSE for user 1 is given by [30]

$$\mathcal{M}_{LE,1}^{\infty} = 1 - \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{H}_1^{\dagger}(e^{j\omega}) \left(\mathcal{H}(e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega}) + \sigma^2 \mathbf{I}\right)^{-1} \mathcal{H}_1(e^{j\omega}) d\omega \tag{47}$$

In Appendix C, we give a different derivation from the one given in [30]. (This technique is used later to compute the MMSE for a MIMO DFE.) The MMSE for the reduced-rank linear equalizer is given by (11), where the MIMO channel moment for user 1 is given by

$$\gamma_n^{(1),\infty} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{H}_1^{\dagger}(e^{j\omega}) \left[\mathcal{H}(e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega}) + \sigma^2 \mathbf{I}\right]^n \mathcal{H}_1(e^{j\omega}) d\omega. \tag{48}$$

B. Successive-DFE (SDFE)

In this section and the next, the performance of full– and RR DFEs are presented for MIMO channels. The DFE cancels both ISI and Co–Channel–Interference (CCI) from other users. We first consider a Successive (S)-DFE, and subsequently consider a Parallel (P)-DFE in the next section.

The output of the DFE for user k is

$$\mathbf{z}_k(i) = \mathbf{F}_k^{\dagger}\mathbf{y}(i) - \mathbf{B}_k^{\dagger}\mathbf{s}_k(i) \tag{49}$$

where $\mathbf{s}_k(i)$ is the vector of interfering symbols being fed back for cancellation. For the SDFE, $\mathbf{s}_k(i)$ consists of the *previously decoded* symbols for user k, i.e., $\{s_k(i-1), s_k(i-2), \cdots\}$, along with the *entire* symbol sequences for users $1, 2, \cdots, k-1$, i.e., $\{\cdots, s_p(i+1), s_p(i), s_p(i-1), \cdots; p = 1, 2, \cdots, k-1\}$. There is no feedback from users $k+1, \cdots, K$. This type of DFE is different from the conventional MIMO DFE discussed in [29]–[32] in that both pre- and post-cursor interference from other users is canceled. (The SDFE considered here is also described in [33].) Here we give closed-form expressions for the MMSE corresponding to both full–rank and RR SDFEs. We first consider an SDFE with a full–rank feedback filter, which we refer to as an SDFE-L, and subsequently consider a reduced–rank feedback filter.

According to the discussion in Section II-D, the covariance matrix of the linear prediction errors corresponding to users $1, \cdots, k$ can be written as

$$\mathbf{R}_e^{(k)} = \mathbf{I} - \mathbf{H}_{1:k}^{\dagger}\mathbf{R}^{-1}\mathbf{H}_{1:k} \tag{50}$$

where $\mathbf{H}_{1:k} = [\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_k]$. Since the users are decoded successively, the MMSE can be computed via the Cholesky decomposition

$$\mathbf{L}_k^\dagger \mathbf{R}_e^{(k)} \mathbf{L}_k = \mathbf{D}_e^{(k)} = \mathsf{diag}\left[\mathcal{M}_k, \cdots, \mathcal{M}_k, \mathcal{M}_{k-1}, \cdots, \mathcal{M}_{k-1}, \cdots, \mathcal{M}_1, \cdots, \mathcal{M}_1\right] \tag{51}$$

where the diagonal matrix $\mathbf{D}_e^{(k)}$ contains the MMSEs for users $1, 2, \cdots, k$. This decomposition can be used to compute user k's MMSE with an infinite-length feedback filter.

The asymptotic performance of the SDFE is given by the following theorem where $|\mathbf{A}|$ denotes the determinant of the matrix $\mathbf{A}$.

Theorem 4: The asymptotic MMSE for user k with an SDFE is

$$\mathcal{M}_{SDFE,k}^{\infty} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left[\frac{\left|\mathbf{I} - \mathcal{H}_{1:k}^\dagger(e^{j\omega})\left(\mathcal{H}(e^{j\omega})\mathcal{H}^\dagger(e^{j\omega}) + \sigma^2\mathbf{I}\right)^{-1}\mathcal{H}_{1:k}(e^{j\omega})\right|}{\left|\mathbf{I} - \mathcal{H}_{1:k-1}^\dagger(e^{j\omega})\left(\mathcal{H}(e^{j\omega})\mathcal{H}^\dagger(e^{j\omega}) + \sigma^2\mathbf{I}\right)^{-1}\mathcal{H}_{1:k-1}(e^{j\omega})\right|}\right] d\omega\right\} \tag{52}$$

where $\mathcal{H}_{1:k}(e^{j\omega}) = [\mathcal{H}_1(e^{j\omega}), \mathcal{H}_2(e^{j\omega}), \cdots, \mathcal{H}_k(e^{j\omega})]$.

The proof is given in Appendix D.

An RR SDFE can be derived as before by defining the appropriate Krylov space in terms of the received covariance matrix and the channel steering vectors for different users. Note that for user k, $\mathbf{C}_{\mathsf{LIN}} = [\mathbf{C}_{\mathsf{LIN}}^{(k)}, \mathbf{C}_{\mathsf{LIN}}^{(k-1)}, \cdots, \mathbf{C}_{\mathsf{LIN}}^{(1)}]$, where $\mathbf{C}_{\mathsf{LIN}}^{(k)}$ is the $N \times N_b$ matrix with each column in the form of (20), i.e., a shifted version of the RR approximation of the linear equalizer for user k. Suppose the ranks for the users associated with the linear RR equalizer are $D_1, D_2, \cdots, D_K$. For two users k_1 and k_2 the error correlation computed in (27) then becomes

$$\begin{aligned}
\varepsilon_q^{(k_1,k_2)} &= E\left\{e_{k_1}(i)e_{k_2}^*(i-q)\right\} \\
&= E\left\{\left[s_{k_1}(i) - [\mathbf{C}_{\mathsf{LIN}}^{(k_1)}]_{N_b+1}^\dagger \mathbf{y}(i)\right]\left[s_{k_2}(i-q) - [\mathbf{C}_{\mathsf{LIN}}^{(k_2)}]_{N_b+1+q}^\dagger \mathbf{y}(i)\right]^*\right\} \\
&= \delta_q\delta_{k_1-k_2} - [\mathbf{C}_{\mathsf{LIN}}^{(k_1)}]_{N_b+1}^\dagger \mathbf{h}_q^{(k_2)} - (\mathbf{h}^{(k_1)})^\dagger [\mathbf{C}_{\mathsf{LIN}}^{(k_2)}]_{N_b+1+q} + [\mathbf{C}_{\mathsf{LIN}}^{(k_1)}]_{N_b+1}^\dagger \mathbf{R}[\mathbf{C}_{\mathsf{LIN}}^{(k_2)}]_{N_b+1+q} \\
&= \delta_q\delta_{k_1-k_2} - \bar{\gamma}_{D_{k_1}}^{(k_1)\dagger}[\Gamma_{D_{k_1}}^{(k_1)}]^{-1}\bar{\gamma}_{D_{k_1},q}^{(k_1,k_2)} - \bar{\gamma}_{D_{k_2},q}^{(k_1,k_2)T}[\Gamma_{D_{k_2}}^{(k_2)}]^{-1}\bar{\gamma}_{D_{k_2}}^{(k_2)} \\
&\quad + \bar{\gamma}_{D_{k_1}}^{(k_1)\dagger}[\Gamma_{D_{k_1}}^{(k_1)}]^{-1}\Gamma_{D_{k_1},D_{k_2},q}^{(k_1,k_2)}[\Gamma_{D_{k_2}}^{(k_2)}]^{-1}\bar{\gamma}_{D_{k_2}}^{(k_2)}
\end{aligned} \tag{53}$$

where the *cross–shifted channel moment* $\gamma_{n,q}^{(k_1,k_2)} = \mathbf{h}^{(k_1)\dagger}\mathbf{R}^n\mathbf{h}_q^{(k_2)}$, $\mathbf{h}_q^{(k)}$ is the shifted channel impulse response for user k,

$$\bar{\gamma}_{n,q}^{(k_1,k_2)} = [\gamma_{0,q}^{(k_1,k_2)}, \gamma_{1,q}^{(k_1,k_2)}, \cdots, \gamma_{n-1,q}^{(k_1,k_2)}]^T, \tag{54}$$

and

$$\Gamma_{n_1,n_2,q}^{(k_1,k_2)} = \begin{bmatrix} \gamma_{1,q}^{(k_1,k_2)} & \gamma_{2,q}^{(k_1,k_2)} & \cdots & \gamma_{n_2,q}^{(k_1,k_2)} \\ \gamma_{2,q}^{(k_1,k_2)} & \gamma_{3,q}^{(k_1,k_2)} & \cdots & \gamma_{n_2+1,q}^{(k_1,k_2)} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{n_1,q}^{(k_1,k_2)} & \gamma_{2,q}^{(k_1,k_2)} & \cdots & \gamma_{n_2+n_1-1,q}^{(k_1,k_2)} \end{bmatrix}. \tag{55}$$

When $n_1 = n_2$, $k_1 = k_2$ or $q = 0$, we drop the repeated exponent, user index, or the shifted index, respectively. In analogy with the single-user DFE analyzed previously, $\varepsilon_q^{(k_1,k_2)} = 0$ when $q \geq 2m(D_{k_1} + D_{k_2})$.

Theorem 5: As $N \to \infty$, the cross–shifted channel moment $\gamma_{n,q}^{(k_1,k_2)}$ converges to

$$\begin{aligned} \gamma_{n,q}^{(k_1,k_2),\infty} &= \lim_{N\to\infty} \mathbf{h}^{(k_1)\dagger} \mathbf{R}^n \mathbf{h}_q^{(k_2)} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{H}_{k_1}^{\dagger}(e^{j\omega}) \left[\mathcal{H}(e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega}) + \sigma^2 \mathbf{I}\right]^n \mathcal{H}_{k_2}(e^{j\omega}) e^{jqw} d\omega \end{aligned} \tag{56}$$

The proof is analogous to the proof of Theorem 1, where Corollary 2 in Appendix A is applied, and is omitted.

Combining (56) and (53), the linear prediction error covariance matrix can be written as

$$\mathbf{R}_e = \begin{bmatrix} \mathbf{R}_e^{(K)} & \mathbf{R}_e^{(K,K-1)} & \cdots & \mathbf{R}_e^{(K,1)} \\ \mathbf{R}_e^{(K-1,K)} & \mathbf{R}_e^{(K-1)} & \cdots & \mathbf{R}_e^{(K-1,1)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}_e^{(1,K)} & \mathbf{R}_e^{(1,K-1)} & \cdots & \mathbf{R}_e^{(1)} \end{bmatrix} \tag{57}$$

where $\mathbf{R}_e^{(k_1,k_2)}$ is an $N_b \times N_b$ Hermitian Toeplitz matrix for $1 \leq k_1, k_2 \leq K$, which has as its first row $\{\varepsilon_q^{(k_1,k_2)}\}_{q=0,1,\cdots,N_b}$, where $\varepsilon_{-q}^{(k_1,k_2)} = [\varepsilon_q^{(k_1,k_2)}]^*$.

As $N_b \to \infty$, in analogy with (32) and (51), Cholesky decomposition of the prediction error covariance matrix gives a block diagonal matrix, where the kth block contains the asymptotic MSE of the RR SDFE-L for the kth user. We can therefore write

$$\mathcal{M}_{RRSDFE-L,k}^{\infty} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \frac{|\mathcal{S}_k(e^{j\omega})|}{|\mathcal{S}_{k-1}(e^{j\omega})|} d\omega\right\} \tag{58}$$

where

$$\mathcal{S}_k(e^{j\omega}) = \begin{bmatrix} S_{k,k}(e^{j\omega}) & S_{k,k-1}(e^{j\omega}) & \cdots & S_{k,1}(e^{j\omega}) \\ S_{k-1,k}(e^{j\omega}) & S_{k-1,k-1}(e^{j\omega}) & \cdots & S_{k-1,1}(e^{j\omega}) \\ \vdots & \vdots & \vdots & \vdots \\ S_{1,k}(e^{j\omega}) & S_{1,k-1}(e^{j\omega}) & \cdots & S_{1,1}(e^{j\omega}) \end{bmatrix} \tag{59}$$

and

$$S_{k_1,k_2}(e^{j\omega}) = \sum_{q=-\infty}^{\infty} \varepsilon_q^{(k_1,k_2)} e^{jq\omega} \tag{60}$$

The asymptotic MSE for the RR SDFE with a reduced–rank feedback filter can be evaluated by extending the previous results for the single-user DFE. Namely, (34)–(36) still hold, where the moment of the error covariance matrix is replaced by a different moment $\psi_s^{(k)} = \mathbf{h}_e^{(k)\dagger}(\bar{\mathbf{R}}_e^{(k)})^s \mathbf{h}_e^{(k)}$. The matrix $\bar{\mathbf{R}}_e^{(k)}$ is obtained by deleting the first row and the first column of the $k \times k$ blocks in the lower-right corner of $\mathbf{R}_e^{(K)}$ in (57), and $\mathbf{h}_e^{(k)}$ is the first column of $\bar{\mathbf{R}}_e^{(k)}$, i.e.,

$$\mathbf{h}_e^{(k)} = [\varepsilon_1^{(k)}, \varepsilon_2^{(k)}, \cdots, \varepsilon_0^{(k,k-1)}, \varepsilon_1^{(k,k-1)}, \cdots, \varepsilon_0^{(k,1)}, \varepsilon_1^{(k,1)}, \cdots]^\dagger \tag{61}$$

and

$$\bar{\mathbf{R}}_e^{(k)} = \begin{bmatrix} \mathbf{R}_e^{(k)-} & \mathbf{R}_e^{(k,k-1)-} & \cdots & \mathbf{R}_e^{(k,1)-} \\ \mathbf{R}_e^{(k-1,k)-} & \mathbf{R}_e^{(k-1)} & \cdots & \mathbf{R}_e^{(k-1,1)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}_e^{(1,k)-} & \mathbf{R}_e^{(1,k-1)} & \cdots & \mathbf{R}_e^{(1)} \end{bmatrix} \tag{62}$$

where

$$\mathbf{R}_e^{(k,p)-} = \begin{bmatrix} \varepsilon_1^{(k,p)*} & \varepsilon_0^{(k,p)} & \varepsilon_1^{(k,p)} & \cdots \\ \varepsilon_2^{(k,p)*} & \varepsilon_1^{(k,p)*} & \varepsilon_0^{(k,p)} & \cdots \\ \varepsilon_3^{(k,p)*} & \varepsilon_2^{(k,p)*} & \varepsilon_1^{(k,p)*} & \cdots \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix} \tag{63}$$

for $p = 1, \cdots, k-1$ is the corresponding Toeplitz matrix with the first row deleted, and $\mathbf{R}_e^{(p,k)-} = [\mathbf{R}_e^{(k,p)-}]^\dagger$. The matrix $\mathbf{R}_e^{(k)-}$ is obtained by deleting the first column and the first row of $\mathbf{R}_e^{(k)}$, and asymptotically $\lim_{N\to\infty} \mathbf{R}_e^{(k)-} = \lim_{N\to\infty} \mathbf{R}_e^{(k)}$.

We also need the following corollary to Theorem 3, which says that the asymptotic (N and $N_b \to \infty$) MSE is achieved with finite N_b. Namely, we define $\mathbf{h}_{e,M}^{(k)}$ and $\bar{\mathbf{R}}_{e,M}^{(k)}$ as $kM \times 1$ and $kM \times kM$ versions of $\mathbf{h}_e^{(k)}$ and $\bar{\mathbf{R}}_e^{(k)}$, defined by (61)–(62), respectively. That is,

$$\begin{aligned} \mathbf{h}_{e,M}^{(k)} &= [\varepsilon_1^{(k)}, \cdots, \varepsilon_M^{(k)}, \varepsilon_0^{(k,k-1)}, \cdots, \varepsilon_{M-1}^{(k,k-1)}, \varepsilon_0^{(k,1)}, \cdots, \varepsilon_{M-1}^{(k,1)}]^\dagger \\ \bar{\mathbf{R}}_{e,M}^{(k)} &= \begin{bmatrix} \mathbf{R}_{e,M^+}^{(k)} & \mathbf{R}_{e,M^+}^{(k,k-1)-} & \cdots & \mathbf{R}_{e,M^+}^{(k,1)-} \\ \mathbf{R}_{e,M^+}^{(k-1,k)-} & \mathbf{R}_{e,M^+}^{(k-1)} & \cdots & \mathbf{R}_{e,M^+}^{(k-1,1)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}_{e,M^+}^{(1,k)-} & \mathbf{R}_{e,M^+}^{(1,k-1)} & \cdots & \mathbf{R}_{e,M^+}^{(1)} \end{bmatrix} \end{aligned}$$

where A_{M^+} denotes the upper left $M \times M$ block in matrix A.

Corollary 1: When the channel impulse response has finite length, the asymptotic error moment can be evaluated from a finite-size error covariance matrix, namely,

$$\psi_m^{(k)\infty} = \mathbf{h}_{e,M}^{(k)\dagger} \left[\bar{\mathbf{R}}_{e,M}^{(k)}\right]^m \mathbf{h}_{e,M}^{(k)} \tag{64}$$

where $M \geq 1 + \lceil \frac{s+1}{2} \rceil Q$, and $Q = \max\{2m(D_{k_1} + D_{k_2}); 1 \leq k_1, k_2 \leq k\}$.
This corollary is a straightforward extension of Theorem 3.

Substituting (64) into (34)–(36) gives the asymptotic MSE for the RR SDFE for user k.

C. Parallel DFE (PDFE)

As previously discussed, the PDFE refers to the *conventional* DFE, in which the feedback depends on decoded symbols from previous sample intervals [29]–[32]. In this section we evaluate the performance of a PDFE with a RR approximation for the feedforward linear equalizer only (RR PDFE-L), and the performance with RR approximations for both the feedforward and feedback filters (RR PDFE). The feedback consists of previously decoded symbols from all users, i.e., $\{s_1(1), \cdots, s_1(i); \cdots; s_{k-1}(1), \cdots, s_{k-1}(i); s_k(1), \cdots, s_k(i-1); \cdots; s_K(1), \cdots, s_K(i-1)\}$.

For the full-rank PDFE, the error covariance sequence at the output of the linear equalizer is the same as for the SDFE, since it does not depend on the feedback. However, before applying a Cholesky decomposition to the error covariance matrix, the outputs of the linear equalizer must be re-ordered to reflect the decoding order. Namely, the linear prediction covariance matrix in this case is

$$\begin{aligned} \mathbf{R}_{e,PDFE} &= \mathbf{K}_{NK}(\mathbf{I} - \mathbf{H}^\dagger \mathbf{R}^{-1} \mathbf{H})\mathbf{K}_{KN} \\ &= \begin{bmatrix} \mathbf{R}_e(0) & \mathbf{R}_e(1) & \mathbf{R}_e(2) & \cdots \\ \mathbf{R}_e^\dagger(1) & \mathbf{R}_e(0) & \mathbf{R}_e(1) & \cdots \\ \mathbf{R}_e^\dagger(2) & \mathbf{R}_e^\dagger(1) & \mathbf{R}_e(0) & \cdots \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix} \end{aligned} \tag{65}$$

where

$$\mathbf{R}_e(q) = \begin{bmatrix} \varepsilon_q^{(K)} & \varepsilon_q^{(K,K-1)} & \cdots & \varepsilon_q^{(K,1)} \\ \varepsilon_q^{(K-1,K)} & \varepsilon_q^{(K-1)} & \cdots & \varepsilon_q^{(K-1,1)} \\ \vdots & \ddots & \ddots & \vdots \\ \varepsilon_q^{(1,K)} & \varepsilon_q^{(1,K-1)} & \cdots & \varepsilon_q^{(1)} \end{bmatrix}, \tag{66}$$

$\mathbf{K}_{NK}$ is a permutation matrix (see Appendix C), and $\varepsilon_q^{(j,k)}$ is defined by (53).

Unlike the SDFE error covariance matrix, which has a finite number of infinite–sized blocks, $\mathbf{R}_{e,PDFE}$ has an infinite number of finite-size blocks. Namely, the (k,m) block entry, $\mathbf{R}_e^{(k,m)}$, in

the SDFE error covariance matrix is the error covariance matrix corresponding to users k and m, whereas each entry in $\mathbf{R}_{e,PDFE}$ is the error covariance matrix across users for a specific time shift.

We have

$$|\mathbf{R}_{e,PDFE}| = |\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{R}^{-1}\mathbf{H}|$$

Hence the geometric average of asymptotic MSE across users is

$$\begin{aligned}
\bar{\mathcal{M}}^{\infty}_{PDFE} &= (\mathcal{M}^{\infty}_{1}\cdots\mathcal{M}^{\infty}_{K})^{\frac{1}{K}}_{PDFE} \\
&= \lim_{N_b\to\infty} |\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{R}^{-1}\mathbf{H}|^{\frac{1}{KN}} \\
&= \exp\left(\frac{1}{2K\pi}\int_{-\pi}^{\pi} \ln\left[|\mathbf{I} - \mathcal{H}^{\dagger}(e^{j\omega})(\mathcal{H}(e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega}) + \sigma^2\mathbf{I})^{-1}\mathcal{H}(e^{j\omega})|\right] d\omega\right) \\
&= (\mathcal{M}^{\infty}_{1}\cdots\mathcal{M}^{\infty}_{K})^{\frac{1}{K}}_{SDFE}
\end{aligned}$$

The full–rank PDFE therefore has the same geometric average asymptotic performance across users as the full–rank SDFE.

If each user has a rank-D linear equalizer then the linear prediction error autocorrelation can computed from (53)-(56) with $D_{k_1} = D_{k_2} = D$. The geometric average of the asymptotic MSE for the RR PDFE-L is given by

$$\bar{\mathcal{M}}^{\infty}_{RRPDFE-L} = \exp\left\{\frac{1}{2K\pi}\int_{-\pi}^{\pi} \ln\left|\mathcal{S}_K(e^{j\omega})\right| d\omega\right\} \tag{67}$$

where $\mathcal{S}_K(e^{j\omega})$ is given by (59).

If user k has a rank-D_b feedback filter, then the asymptotic MSE for the RR PDFE can be obtained in an analogous form to (34) , i.e.,

$$\mathcal{M}_{RRPDFE,k} = \bar{\mathcal{M}}_{RRLE-L} - \bar{\psi}^{(k)\dagger}_{D_b}[\Psi^{(k)}_{D_b}]^{-1}\bar{\psi}^{(k)}_{D_b} \tag{68}$$

Following the method for computing the moments of the error covariance matrix associated with the RR SDFE in the last section, we delete the first k columns and rows of $\mathbf{R}_{e,PDFE}$ to get

$$\mathbf{h}^{(k)}_{e} = [\varepsilon^{(k,k-1)}_{0}, \varepsilon^{(k,k-2)}_{0}, \cdots, \varepsilon^{(k,K)}_{1}, \varepsilon^{(k,K-1)}_{1}, \cdots, \varepsilon^{(k,K)}_{2}, \varepsilon^{(k,K-1)}_{2}, \cdots]^{\dagger} \tag{69}$$

and

$$\bar{\mathbf{R}}^{(k)}_{e} = \begin{bmatrix} \mathbf{R}^{(k)-}_{e}(0) & \mathbf{R}^{(k)-}_{e}(1) & \mathbf{R}^{(k)-}_{e}(2) & \cdots \\ [\mathbf{R}^{(k)-}_{e}(1)]^{\dagger} & \mathbf{R}_{e}(0) & \mathbf{R}_{e}(1) & \cdots \\ [\mathbf{R}^{(k)-}_{e}(2)]^{\dagger} & \mathbf{R}^{\dagger}_{e}(1) & \mathbf{R}_{e}(0) & \cdots \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix} \tag{70}$$

where $\mathbf{R}_e^{(k)-}(i), i \geq 1$, is the $(k-1) \times K$ matrix

$$\mathbf{R}_e^{(k)-}(i) = \begin{bmatrix} \varepsilon_i^{(k-1,K)} & \cdots & \varepsilon_i^{(k-1)} & \cdots & \varepsilon_1^{(k-1,1)} \\ \varepsilon_i^{(k-2,K)} & \cdots & \varepsilon_i^{(k-2,k-1)} & \cdots & \varepsilon_1^{(k-2,1)} \\ \varepsilon_i^{(k-3,K)} & \cdots & \varepsilon_i^{(k-3,k-1)} & \cdots & \varepsilon_1^{(k-3,1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varepsilon_i^{(1,K)} & \cdots & \varepsilon_i^{(1,k-1)} & \cdots & \varepsilon_1^{(1)} \end{bmatrix} \tag{71}$$

and $\mathbf{R}_e^{(k)-}(0) = \left[\varepsilon_0^{(i,j)}\right]_{i,j=k-1,\cdots,1}$, is a $(k-1) \times (k-1)$ matrix, and is obtained from $\mathbf{R}_e^{(k)}(0)$ by deleting the first row and column. Finally, $\psi_n^{(k)} = \mathbf{h}_e^{(k)\dagger}[\bar{\mathbf{R}}_e^{(k)}]^n \mathbf{h}_e^{(k)}$.

When the length of the channel impulse response is finite, $\varepsilon_q^{(k_1,k_2)} = 0$ for $q > Q = 4mD$. Consequently, $\mathbf{h}_e^{(k)} = \left[\left[\mathbf{h}_{e,Q}^{(k)}\right]^T, 0, 0 \cdots\right]^T$, and $\mathbf{R}_e^{(k)}(q) = \mathbf{0}$ for $q > Q$, so that the asymptotic moment of the error covariance matrix can be computed as $\psi_s^{(k)\infty} = \mathbf{h}_{e,M}^{(k)\dagger}[\bar{\mathbf{R}}_{e,M}^{(k)}]^s \mathbf{h}_{e,M}$, for any finite $M \geq \lceil \frac{s+1}{2} \rceil Q + 1$.

D. Numerical Results

Fig. 4 shows performance results (output MSE vs. rank) for the RRSDFE–L and RRSDFE with a MIMO channel. There are two equal–power users with SNR$= 10$ dB and two receive antennas. All constituent single-input/single-output channels have seven equal-power independent Rayleigh fading taps. The figure shows that for the SDFE the second user experiences better performance than the first user due to the enhanced interference suppression and cancellation. These results also show that full–rank performance can be nearly achieved with a relatively low–rank equalizer. Results for the RRSDFE are shown with a rank-one feedback filter for the first user, and rank-one and rank-two feedback filters for the second user. The RRSDFE performance is close to the RRSDFE–L with a full-rank feedback filter.

IV. Conclusions

The output MSE for RR linear and decision-feedback equalizers has been evaluated asymptotically as the filter length tends to infinity. The RR equalizers are constrained to lie in a Krylov subspace, which can reduce complexity and reduce the amount of training needed in an adaptive mode. The RR DFE is based on the representation of the feedforward filter as a linear equalizer followed by an error whitening filter. Namely, we considered RR approximations for only the linear equalizer, and for both feedforward filters.

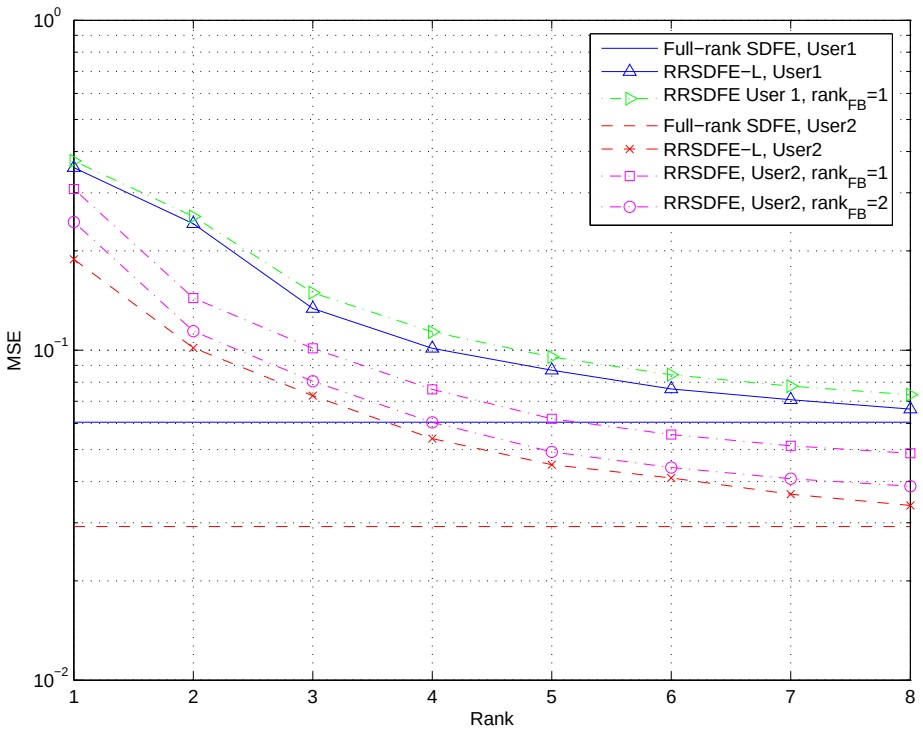

Fig. 4: MMSE vs. rank for RR linear and decision–feedback equalizers with a MIMO channel. There are two equal–power users and two receive antennas.

In all cases the output MSE is easily computed in terms of the channel frequency response, and indicates the degradation in performance, relative to full–rank equalization, as a function of the equalizer rank. For the single-user SISO channel, the output MSE is expressed in the term of moments of the channel spectrum, which can be easily computed via an inverse FFT of the channel power spectrum. Numerical examples show that full-rank equalizer performance can be nearly achieved with low–rank equalizers.

The MIMO channel model considered allows for multiple users and multiple receive antennas. The cancellation of ISI and CCI can be done in different ways, depending on which symbols are assumed to be available for feedback. We have evaluated output MSE with both successive and parallel interference cancellation. With full-rank equalization, we have shown that the geometric average of MMSE across users is the same for both the SDFE and PDFE, although each generally gives different MSEs for different users. The output MSE for the reduced-rank versions of the MIMO equalizers requires the computation of inner products of shifted channel impulse responses with moments of the received covariance matrix. The asymptotic MSE can be computed easily in terms of the MIMO channel frequency response. The results can be used to quantify the

tradeoff between the benefits of rank reduction, in terms of complexity and convergence with finite training, with the increase in achievable MMSE.

APPENDIX

A. Asymptotic Eigenvalue Distribution of Products of Toeplitz Matrices

Here we introduce and extend some well-known properties of large Toeplitz matrices, which are needed to derive the MSE expressions for the DFEs considered. Following [27], we first define the notion of asymptotic equivalence of two sequences of matrices. Given an $N \times N$ matrix $\mathbf{A}_N$, the strong norm (or spectral norm) is defined as

$$\|\mathbf{A}_N\|_s^2 = \max_{\mathbf{x}} \frac{\mathbf{x}^\dagger \mathbf{A}_N^\dagger \mathbf{A}_N \mathbf{x}}{\mathbf{x}^\dagger \mathbf{x}}, \tag{72}$$

and the weak norm (or scaled Frobenius norm) is defined as

$$|\mathbf{A}_N|_w^2 = \frac{1}{N}\mathsf{trace}[\mathbf{A}_N^\dagger \mathbf{A}_N] = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|a_{i,j}|^2 \tag{73}$$

where $\mathbf{x}$ is an $N \times 1$ vector. Two sequences of $N \times N$ matrices $\{\mathbf{A}_N\}$ and $\{\mathbf{B}_N\}$ are said to be asymptotically equivalent if

$$\|\mathbf{A}_N\|_s, \|\mathbf{B}_N\|_s \leq M < \infty \qquad \forall N \geq 0 \tag{74}$$

$$\lim_{N\to\infty} |\mathbf{A}_N - \mathbf{B}_N|_w = 0 \tag{75}$$

Theorem 6: ([27] Theorem 2.3) Let $\{\mathbf{A}_N\}$ and $\{\mathbf{B}_N\}$ be asymptotically equivalent sequences of Hermitian matrices with eigenvalues $\{\alpha_{N,k}\}$ and $\{\beta_{N,k}\}$ respectively. Since $\mathbf{A}_N$ and $\mathbf{B}_N$ are bounded, there exist finite a and A such that $a \leq \alpha_{N,k}$ and $\beta_{N,k} \leq A$ for $k = 0, 1, \cdots, N-1$. For any function $F(x)$ continuous on $[a, A]$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} F(\alpha_{N,k}) = \lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} F(\beta_{N,k}) \tag{76}$$

The following theorem extends the notion of asymptotic equivalence to products of matrices. In what follows we drop the sequence notation, and state that $\mathbf{A}_N$ is asymptotically equivalent to $\mathbf{B}_N$.

Theorem 7: ([34] Theorem 2.1) If $\mathbf{A}_N$ is asymptotically equivalent to $\hat{\mathbf{A}}_N$, and $\mathbf{B}_N$ is asymptotically equivalent to $\hat{\mathbf{B}}_N$, then $\mathbf{A}_N\mathbf{B}_N$ is asymptotically equivalent to $\hat{\mathbf{A}}_N\hat{\mathbf{B}}_N$. If $\min\{\lambda(\mathbf{A}_N), \lambda(\hat{\mathbf{A}}_N)\} \geq a > 0$, where a is a constant for $\forall N \geq 0$, then $\mathbf{A}_N^{-1}\mathbf{B}_N$ is asymptotically equivalent to $\hat{\mathbf{A}}_N^{-1}\hat{\mathbf{B}}_N$.

The following theorem extends Theorem 6 to sequences of block matrices.

Theorem 8: If $\mathbf{A}_N^{(i,j)}$ is asymptotically equivalent to $\hat{\mathbf{A}}_N^{(i,j)}, i,j = 1,\cdots,K$, and the block matrix $\mathbf{A}_N = [\mathbf{A}_N^{(i,j)}]$, then $\mathbf{A}_N$ is asymptotically equivalent to $\hat{\mathbf{A}}_N = [\hat{\mathbf{A}}_N^{(i,j)}]$.

Proof: It is straightforward to show that (75) holds, since

$$|\mathbf{A}_N - \hat{\mathbf{A}}_N|_w^2 = \sum_{i,j=1}^{K} |\mathbf{A}_N^{(i,j)} - \hat{\mathbf{A}}_N^{(i,j)}|_w^2 \to 0$$

for finite K. Let $\mathbf{x} = \left[\mathbf{x}_1^T \ \cdots \ \mathbf{x}_K^T\right]^T$, where $\mathbf{x}_i$ is an $N \times 1$ vector. We can again apply the Cauchy–Schwartz inequality to show that (74) is satisfied:

$$\begin{aligned}
\frac{\mathbf{x}^\dagger \mathbf{A}_N^\dagger \mathbf{A}_N \mathbf{x}}{\|\mathbf{x}\|^2} &= \frac{\sum_{i,j} \mathbf{x}_i^\dagger \left(\sum_{k=1}^K \mathbf{A}_N^{(k,i)\dagger} \mathbf{A}_N^{(k,j)}\right) \mathbf{x}_j}{\|\mathbf{x}\|^2} \\
&\le \frac{\sum_{i,j,k} \sqrt{(\mathbf{x}_i^\dagger \mathbf{A}_N^{(k,i)\dagger} \mathbf{A}_N^{(k,i)} \mathbf{x}_i)(\mathbf{x}_j^\dagger \mathbf{A}_N^{(k,j)\dagger} \mathbf{A}_N^{(k,j)} \mathbf{x}_j)}}{\|\mathbf{x}\|^2} \\
&\le \frac{\sum_{i,j,k} \sqrt{M\|\mathbf{x}_i\|^2 M\|\mathbf{x}_j\|^2}}{\|\mathbf{x}\|^2} = KM \frac{\sum_{i,j} \|\mathbf{x}_i\| \|\mathbf{x}_j\|}{\|\mathbf{x}\|^2} \\
&\le \frac{1}{2} KM \frac{\sum_{i,j} (\|\mathbf{x}_i\|^2 + \|\mathbf{x}_j\|^2)}{\|\mathbf{x}\|^2} = K^2 M \frac{\sum_i \|\mathbf{x}_i\|^2}{\|\mathbf{x}\|^2} \\
&= K^2 M < \infty
\end{aligned}$$

where $M = \max_{i,j} \|\mathbf{A}_N^{(i,j)}\|_s^2$. ■

It is shown in [27] that the Toeplitz matrix $\mathbf{T}_N$ with first row $t_0, \cdots, t_{N-1}$ is asymptotically equivalent to the circulant matrix $\mathbf{C}_N$, with first row $c_0, \cdots, c_{N-1}$, where

$$c_i = \begin{cases} t_i & 0 \le i \le \lfloor \frac{N}{2} \rfloor - 1 \\ t_{i-N} & \lceil \frac{N}{2} \rceil \le i \le N-1 \\ 0 & i = \frac{N-1}{2} \qquad \text{when N is odd} \end{cases} \tag{77}$$

Furthermore, since $\mathbf{C}_N$ has the eigen-decomposition [35]

$$\mathbf{C}_N = \frac{1}{N} \mathbf{W} \Lambda \mathbf{W}^\dagger \tag{78}$$

where $\mathbf{W}$ is the DFT matrix and $\Lambda = \mathsf{diag}[\lambda_0, \lambda_1, \cdots, \lambda_{N-1}]$, where

$$\lambda_k = \sum_{i=0}^{N-1} c_i e^{-j\frac{2ik\pi}{N}}, \qquad k = 0, \cdots, N-1 \tag{79}$$

we have the following theorem.

Theorem 9: ([27] Theorem 4.1) If $\mathbf{T}_N$ is a Hermitian Toeplitz matrix with eigenvalues $\lambda_{N,k}$, then for any function $F(x)$ continuous on the appropriate interval, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} F(\lambda_{N,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(\hat{t}(e^{j\omega})) d\omega \tag{80}$$

where $\hat{t}(e^{j\omega})$ is the Fourier Transform of the sequence $\{t_{-(N-1)}, \cdots, t_0, \cdots, t_{N-1}\}$.

The following corollary to Theorems 7 and 8 states that products of block matrices with Toeplitz blocks are asymptotically equivalent to the products of the corresponding block matrices with circulant blocks.

Corollary 2: Let $\mathbf{T}_{1,N}, \mathbf{T}_{2,N}$ be two block matrices with Toeplitz blocks $\mathbf{T}_{1,N}^{(i,j)}$ and $\mathbf{T}_{2,N}^{(i,j)}$, $1 \leq i,j \leq K$, respectively. If each block is asymptotically equivalent to a circulant matrix $\mathbf{C}_{l,N}^{(i,j)}, l = 1,2$, defined as in (77), then, $\mathbf{T}_{1,N}\mathbf{T}_{2,N}$ is asymptotically equivalent to $\mathbf{C}_{1,N}\mathbf{C}_{2,N}$. If $\mathbf{T}_{1,N}$ is positive definite, with eigenvalues bounded away from zero, then $\mathbf{T}_{1,N}^{-1}\mathbf{T}_{2,N}$ is asymptotically equivalent to $\mathbf{C}_{1,N}^{-1}\mathbf{C}_{2,N}$, where $\mathbf{C}_{1,N} = [\mathbf{C}_{1,N}^{(i,j)}]$ and $\mathbf{C}_{2,N} = [\mathbf{C}_{2,N}^{(i,j)}]$.

Finally, we observe that Theorem 9 can be used to show directly that the MMSEs for finite–length linear and decision-feedback equalizers converge to the classical formulas (14) and (38), respectively. Namely, for the linear equalizer we have

$$\begin{aligned}
\mathcal{M}_{LE}^{\infty} &= \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}(\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{R}^{-1}\mathbf{H}) \\
&= \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}(\mathbf{R}\mathbf{R}^{-1} - \mathbf{H}\mathbf{H}^{\dagger}\mathbf{R}^{-1}) \\
&= \lim_{N\to\infty} \frac{\sigma^2}{N}\mathrm{trace}\left\{\mathbf{R}^{-1}\right\} \\
&= \lim_{N\to\infty} \frac{\sigma^2}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sigma^2}{S(e^{j\omega})}d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sigma^2}{|H(e^{j\omega})|^2+\sigma^2}d\omega
\end{aligned}$$

For the DFE, we have $|\mathbf{R}_e| = |\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{R}^{-1}\mathbf{H}| = |\mathbf{D}_e| = \mathcal{M}_{DFE}^{N}$, so that

$$\begin{aligned}
\mathcal{M}_{DFE}^{\infty} &= \lim_{N\to\infty}\left(|\mathbf{I} - \mathbf{H}^{\dagger}\mathbf{R}^{-1}\mathbf{H}|\right)^{1/N} \\
&= \lim_{N\to\infty}\sigma^2|\mathbf{R}^{-1}|^{\frac{1}{N}} \\
&= \lim_{N\to\infty}\sigma^2\prod_{i=1}^{N}(\lambda_i^{-1}(\mathbf{R}))^{\frac{1}{N}} \\
&= \exp\{\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\ln\frac{\sigma^2}{\lambda_i}\} \\
&= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left[\frac{\sigma^2}{|H(e^{j\omega})|^2+\sigma^2}\right]d\omega\right\}
\end{aligned}$$

where the limit and $\exp$ can be interchanged provided that the exponent is bounded (since all derivatives of $\exp(\cdot)$ are continuous).

B. Proof of Theorem 2

We have

$$\begin{aligned}\gamma_{n,q}^{\infty} &= \lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\{\mathbf{H}^{\dagger}\mathbf{R}^{n}\mathbf{H}_{q}\}\\ &= \lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\{\mathbf{R}^{n}\mathbf{R}_{q}\} \qquad (81)\end{aligned}$$

where $\mathbf{H}_q$ is a Toeplitz matrix, analogous to (2), with center column $\mathbf{h}_q$, and $\mathbf{R}_q = \mathbf{H}_q\mathbf{H}^{\dagger}$ is a Toeplitz matrix, with elements $\{R_{-N-q},\cdots,R_0-\sigma^2,\cdots,R_{N-q}\}$, which are the shifted autocorrelations without noise. Let $\mathbf{C}$ and $\mathbf{C}_q$ be circulant matrices, which approximate $\mathbf{R}$ and $\mathbf{R}_q$ respectively, according to (77). From the discussion in Appendix A, $\mathbf{R}^n\mathbf{R}_q$ is asymptotically equivalent to $\mathbf{C}^n\mathbf{C}_q$, so that

$$\begin{aligned}\gamma_{n,q}^{\infty} &= \lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\{\mathbf{C}^{n}\mathbf{C}_{q}\}\\ &= \lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\{\mathbf{V}\mathbf{\Lambda}^{n}\mathbf{V}^{\dagger}\mathbf{V}\mathbf{\Lambda}_{q}\mathbf{V}^{\dagger}\}\\ &= \lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\{\mathbf{\Lambda}^{n}\mathbf{\Lambda}_{q}\}\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{p=0}^{N-1}(|H(e^{j\frac{2\pi p}{N}})|^2+\sigma^2)^n|H(e^{j\frac{2\pi p}{N}})|^2e^{jpq\frac{2\pi}{N}}\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(|H(e^{j\omega})|^2+\sigma^2)^n|H(e^{j\omega})|^2e^{jq\omega}d\omega\end{aligned}$$

where the last two equalities follow from the eigen-decomposition of $\mathbf{C}$ and $\mathbf{C}_q$ given in (78)-(79), i.e.,

$$\mathbf{V} = \frac{1}{\sqrt{N}}\left\{\exp\left(-j2\pi pk/N\right); p,k = 0,1,\cdots,N-1\right\} \qquad (82)$$

and

$$\mathbf{\Lambda} = \mathsf{diag}\left[|H(e^{j0})|^2+\sigma^2,\cdots,|H(e^{j2\pi(N-1)/N})|^2+\sigma^2\right] \qquad (83)$$

$$\mathbf{\Lambda}_q = \mathsf{diag}\left[|H(e^{j0})|^2e^{jq0},\cdots,|H(e^{j2\pi(N-1)/N})|^2e^{jq2\pi(N-1)/N}\right] \qquad (84)$$

To show that $\gamma_{n,q}^{\infty}=0$ for $q>2m(n+1)$ we note that $\mathbf{R}^n$ is a $4mn$-diagonal matrix, and $\mathbf{R}_q$ is non-zero only in the upper-right triangular block bounded by the $(q-2m)^{th}$ diagonal. If $q-2m>2mn$, then all diagonal elements of $\mathbf{R}^n\mathbf{R}_q$ are zero, so that $\gamma_{n,q}=0$ according to (81).

C. Alternative Derivation of MMSE for MIMO Linear Equalizer

From (45), the covariance matrix of the received vector can be partitioned in block form as

$$\mathbf{R} = \mathbf{H}\mathbf{H}^{\dagger} + \sigma^2\mathbf{I} = \begin{bmatrix} \mathbf{R}_{1,1} & \mathbf{R}_{1,2} & \cdots & \mathbf{R}_{1,L} \\ \mathbf{R}_{2,1} & \mathbf{R}_{2,2} & \cdots & \mathbf{R}_{2,L} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}_{L,1} & \mathbf{R}_{L,2} & \cdots & \mathbf{R}_{L,L} \end{bmatrix} \tag{85}$$

where $\mathbf{R}_{i,j} = \sum_{k=1}^{K} \mathbf{H}_{i,k}\mathbf{H}_{k,j}^{\dagger} + \sigma^2\delta_{i-j}\mathbf{I}$ is a Toeplitz matrix. With an infinite-length, full-rank linear equalizer the MMSE for user 1 is given by

$$\begin{aligned} \mathcal{M}_{LE,1}^{\infty} &= \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\mathbf{I} - \mathbf{H}_1^{\dagger}\mathbf{R}^{-1}\mathbf{H}_1\right\} \\ &= 1 - \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\mathbf{R}^{-1}\mathbf{H}_1\mathbf{H}_1^{\dagger}\right\} \\ &= 1 - \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\mathbf{R}^{-1}\mathbf{R}_1\right\} \end{aligned} \tag{86}$$

where $\mathbf{R}_1 = \mathbf{H}_1\mathbf{H}_1^{\dagger}$ can also be partitioned into Toeplitz blocks. Let $\mathbf{C}$ and $\mathbf{C}_1$ be the corresponding matrices with circulant blocks, which approximate $\mathbf{R}$ and $\mathbf{R}_1$, respectively, as in (77). According to Corollary 2 in Appendix A, $\mathbf{R}^{-1}\mathbf{R}_1$ is asymptotically equivalent to $\mathbf{C}^{-1}\mathbf{C}_1$, hence

$$\lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\mathbf{R}^{-1}\mathbf{R}_1\right\} = \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\mathbf{C}^{-1}\mathbf{C}_1\right\} \tag{87}$$

and from (86) we have

$$\begin{aligned} \mathcal{M}_{LE,1}^{\infty} &= 1 - \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\mathbf{C}^{-1}\mathbf{C}_1\right\} \\ &= 1 - \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\bar{\mathbf{V}}\bar{\mathbf{\Lambda}}^{-1}\bar{\mathbf{V}}^{\dagger}\bar{\mathbf{V}}\bar{\mathbf{\Lambda}}_1\bar{\mathbf{V}}^{\dagger}\right\} \\ &= 1 - \lim_{N\to\infty} \frac{1}{N}\mathsf{trace}\left\{\bar{\mathbf{\Lambda}}^{-1}\bar{\mathbf{\Lambda}}_1\right\} \end{aligned} \tag{88}$$

where

$$\bar{\mathbf{V}} = \mathsf{diag}\left[\mathbf{V}, \mathbf{V}, \cdots, \mathbf{V}\right],$$

and $\mathbf{V}$ is defined in (82). $\bar{\mathbf{\Lambda}}$ and $\bar{\mathbf{\Lambda}}_1$ can be partitioned into $L \times L$ diagonal blocks given by

$$\begin{aligned} \mathbf{\Lambda}_{i,j} &= \mathsf{diag}\left[\sum_{k=1}^{K} H_{i,k}(e^{j2\pi n/N})H_{k,j}^{*}(e^{j2\pi n/N}) + \sigma^2\delta_{i-j}; n = 0, 1, \cdots, N-1\right] \\ \mathbf{\Lambda}_{1;i,j} &= \mathsf{diag}\left[H_{i,1}(e^{j2\pi k/N})H_{j,1}^{*}(e^{j2\pi k/N}); n = 0, 1, \cdots, N-1\right] \end{aligned}$$

where $1 \le i, j \le L$.

Using the vec–permutation matrix $\mathbf{K}_{r,s}$ [36], (88) can be written as

$$\begin{aligned}\mathcal{M}_{LE,1}^{\infty} &= 1-\lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\left\{\mathbf{K}_{L,N}\mathcal{R}^{-1}\mathbf{K}_{L,N}^{T}\mathbf{K}_{L,N}\mathcal{R}_{1}\mathbf{K}_{L,N}^{T}\right\}\\ &= 1-\lim_{N\to\infty}\frac{1}{N}\mathsf{trace}\left\{\mathcal{R}^{-1}\mathcal{R}_{1}\right\}\end{aligned}$$

where $\mathbf{K}_{L,N}$ is an $LN\times LN$ permutation matrix defined by

$$[\mathbf{K}_{L,N}]_{i,j}=\begin{cases}1, & \text{if } \begin{array}{l} i=aN+b+1, 0\le a\le L-1\\ j=bL+a+1, 0\le b\le N-1\end{array}\\ 0, & \text{otherwise}\end{cases}$$

The block–diagonal matrices $\mathcal{R}$ and $\mathcal{R}_1$ are given by

$$\begin{aligned}\mathcal{R} &= \mathsf{diag}\left[\mathcal{H}(e^{j2\pi n/N})\mathcal{H}^{\dagger}(e^{j2\pi n/N})+\sigma^{2}\mathbf{I}; n=0,1,\cdots,N-1\right]\\ \mathcal{R}_{1} &= \mathsf{diag}\left[\mathcal{H}_{1}(e^{j2\pi n/N})\mathcal{H}_{1}(e^{j2\pi n/N})^{\dagger}; n=0,1,\cdots,N-1\right]\end{aligned}$$

so that

$$\begin{aligned}\mathcal{M}_{LE,1}^{\infty} &= 1-\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mathcal{H}_{1}^{\dagger}(e^{j2\pi n/N})\left[\mathcal{H}(e^{j2\pi n/N})\mathcal{H}^{\dagger}(e^{j2\pi n/N})+\sigma^{2}\mathbf{I}\right]^{-1}\mathcal{H}_{1}(e^{j2\pi n/N})\\ &= 1-\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{H}_{1}^{\dagger}(e^{j\omega})\left[\mathcal{H}(e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega})+\sigma^{2}\mathbf{I}\right]^{-1}\mathcal{H}_{1}(e^{j\omega})d\omega.\end{aligned}$$

D. Proof of Theorem 4

From (51), we have

$$\begin{aligned}\mathcal{M}_{k}^{\infty} &= \frac{\mathcal{M}_{1}^{\infty}\mathcal{M}_{2}^{\infty}\cdots\mathcal{M}_{k}^{\infty}}{\mathcal{M}_{1}^{\infty}\mathcal{M}_{2}^{\infty}\cdots\mathcal{M}_{k-1}^{\infty}}\\ &= \lim_{N_b\to\infty}\frac{\left|\mathbf{R}_{e}^{(k)}\right|^{\frac{1}{N_b'}}}{\left|\mathbf{R}_{e}^{(k-1)}\right|^{\frac{1}{N_b'}}}\end{aligned}\tag{89}$$

where $N_b'=2N_b+1$. According to (50) and Appendix A, we can compute

$$\begin{aligned}\lim_{N_b\to\infty}\left|\mathbf{R}_{e}^{(k)}\right|^{\frac{1}{N_b'}} &= \lim_{N_b\to\infty}\left|\mathbf{I}-\mathbf{H}_{1:k}^{\dagger}\mathbf{R}^{-1}\mathbf{H}_{1:k}\right|^{\frac{1}{N_b'}}\\ &= \lim_{N_b\to\infty}\left|\mathbf{I}-\mathbf{R}^{-1}\mathbf{H}_{1:k}\mathbf{H}_{1:k}^{\dagger}\right|^{\frac{1}{N_b'}}\\ &= \lim_{N_b\to\infty}\left|\mathbf{I}-\mathbf{R}^{-1}\mathbf{R}_{1:k}\right|^{\frac{1}{N_b'}}\\ &= \lim_{N_b\to\infty}\left|\mathbf{I}-\mathbf{C}^{-1}\mathbf{C}_{1:k}\right|^{\frac{1}{N_b'}}\end{aligned}\tag{90}$$

where $\mathbf{C}$ and $\mathbf{C}_{1:k}$ are the circulant approximations of $\mathbf{R}$ and $\mathbf{R}_{1:k}$, respectively. Applying the eigen–decomposition for $\mathbf{C}$ and $\mathbf{C}_{1:k}$, as in the proof of (47), (90) becomes

$$\lim_{N_b\to\infty} |\mathbf{R}_e^{(k)}|^{\frac{1}{N_b'}} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left|\mathbf{I} - \mathcal{H}_{1:k}^{\dagger}(e^{j\omega})(\mathcal{H}(e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega}) + \sigma^2\mathbf{I})^{-1}\mathcal{H}_{1:k}(e^{j\omega})\right| d\omega\right\} \quad (91)$$

Substituting (91) into (89) gives (52). We remark that (89) and (91) provide a recursive method for obtaining the sequence of user MMSEs in the order decoded.

REFERENCES

[1] I. I. Scharf. The SVD and reduced rank signal processing. *Signal Processing*, 25(2):113–133, 1991.

[2] E. Lindskog and C. Tidestav. Reduced rank space-time equalization. In *Personal Indoor Mobile Radio Communications Conference*, volume 3, pages 1081–1085, 1998.

[3] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, New York, 1993.

[4] L. L. Scharf and D. W. Tufts. Rank reduction for modeling stationary signals. *IEEE Transactions on Signal Processing*, ASSP-35(3):350–355, March 1987.

[5] X. Wang and H. V. Poor. Blind multiuser detection: A subspace approach. *IEEE Transactions on Information Theory*, 44(2), March 1998.

[6] J.S. Goldstein and I.S. Reed. Reduced rank adaptive filtering. *IEEE Trans. Signal Processing*, 45(2):492–496, February 1997.

[7] M. L. Honig and W. Xiao. Performance of reduced-rank linear interference suppression. *IEEE Transactions on Information Theory*, 47(5):1928–1946, July 2001.

[8] M. L. Honig. Adaptive linear interference suppression for packet DS-CDMA. *European Transactions on Telecommunications*, 9(2):173–182, March-April 1998.

[9] M. L. Honig. A comparison of subspace adaptive filtering techniques for DS-CDMA interference suppression. In *IEEE International Military Communications Conference (MILCOM)*, Monterey, CA, Nov 1997.

[10] A. M. Haimovich and Y. Bar-Ness. An eigenanalysis interference canceler. *IEEE Transactions on Signal Processing*, 39(1):76–84, 1991.

[11] D. A. Pados and S. N. Batalama. Low-complexity blind detection of DS/CDMA signals: Auxiliary-vector receivers. *IEEE Transactions on Communications*, 45(12):1586–1594, December 1997.

[12] D. A. Pados and S. N. Batalama. Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays. *IEEE Transactions on Communications*, 47(9):1406–1415, September 1999.

[13] J.S. Goldstein, I.S. Reed, and L.L. Scharf. A multistage representation of the Wiener filter based on orthogonal projections. *IEEE Transactions on Information Theory*, 44(7), November 1998.

[14] M. L. Honig and J. S. Goldstein. Adaptive reduced-rank interference suppression based on the multistage wiener filter. *IEEE Transactions on Communications*, 50(6):986–994, June 2002.

[15] L. Li, A. Tulino, and S. Verdú. Asymptotic eigenvalue moments for linear multiuser detection. *Communications in Information and Systems*, 1(3):273–304, Fall 2001.

[16] W. Xiao and M. L. Honig. Convergence analysis of adaptive full-rank and multistage reduced-rank interference suppression. In *Proceeding of CISS'2000*, Princeton, NJ, March 2000.

[17] P. Loubaton and W. Hachem. Asymptotic analysis of reduced-rank wiener filters. In *IEEE Information Theory Workshop*, March 2003.

[18] L. Tong and S. Perreau. Multichannel blind channel estimation: from subspa ce to maximum likelihood methods. *Proceedings of the IEEE*, 86(10):1951–1968, October 1998.

[19] G. Long, F. Ling, and J. G. Proakis. Fractionally-spaced equalizers based on singular value decomposition. In *Proc. International Conf. on Acoustics, Speech, and Signal Processing*, volume 3, pages 1514–1517, April 1988.

[20] S. Chowdhury, M. D. Zoltowski, and J. S. Goldstein. Reduced-rank adaptive MMSE equalization for the forward link in high-speed CDMA. In *Proceedings of IEEE Midwest Symposium on Circuits and Systems*, volume 1, pages 6–9, August 2000.

[21] G. Dietl, M. D. Zoltowski, and M. Joham. Reduced-rank equalization for edge via conjugate gradient implementation of multi-stage nested wiener filter. In *IEEE Vehicular Technolgy Conference*, volume 3, pages 1912–1916, Oct. 2001.

[22] B. Mouhouche, K. A. Meraijm, N. Ibpahim, and P. Loubaton. Reduced-rank adaptive chip-level MMSE equalization for the forward link of long-code DS-CDMA systems. In *Seventh International Symposium on Signal Processing and its Applications*, volume 1, pages 497 – 500, July 2003.

[23] Y. Sun and M. L. Honig. Performance of reduced-rank equalization. In *IEEE Internatinal Symposium on Information Theory*, 2002.

[24] G. Dietl, C. Mensing, W. Utschick, J. A. Nossek, and M. D. Zoltowski. Multi-stage mmse decision feedback equalization for edge. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 509–512, April 2003.

[25] Y. Sun, V. Tripathi, and M. L. Honig. Adaptive, turbo, reduced-rank equalization for mimo channels. submitted to IEEE Transactions on Wireless Communications.

[26] W. Chen, U. Mitra, and P. Schniter. On the equivalence of three reduced rank linear estimators with applications to ds-cdma. *IEEE Transactions on Information Theory*, 48(9):2609–2614, September 2002.

[27] R. M. Gray. On the asymptotic eigenvalue distribution of toeplitz matrices. *IEEE Transactions on Information Theory*, 18(6):725–730, November 1972.

[28] J. G. Proakis. *Digital Communications*. McGraw-Hill, New York, 1995.

[29] N. Al-Dhahir and J. M. Cioffi. MMSE decision-feedback equalizers: Finite-length results. *IEEE Transactions on Information Theory*, 41(4):961–975, July 1995.

[30] S. L. Ariyavisitakul, J. H. Winters, and I. Lee. Optimum space-time processors with dispersive interference: Unified analysis and required filter span. *IEEE Transactions on Communications*, 47(7):1073–1083, July 1999.

[31] A. Duel-Hallen. Equalizers for muiltple input/multiple output channels and PAM systems with cyclostationary input sequences. *IEEE Journal on Selected Areas in Communications*, 10(3):630–639, 1992.

[32] N. Al-Dhahir and A. H. Sayed. The finite-length Multi-Input Multi-Output MMSE-DFE. *IEEE Transactions on Signal Processing*, 48(10):2921–2936, October 2000.

[33] R. Schlagenhaufer, B. R. Petersen, and A. B. Sesay. Delayed–decision–feedback equalization for multiuser systems. In *Wireless 99*, volume 1, pages 71–84, Calgary, Alberta, Canada, July 1999.

[34] R. M. Gray. Toeplitz and circulant matrices: A review. `http://www-ee.stanford.edu/~gray/toeplitz.pdf`.

[35] P. Lancaster. *Theory of Matrices*. Academic Press, New York, 1969.

[36] H. V. Henderson and S. R. Searle. The vec–permutation matrix, the vec operator and Kronecker products: A review. *Linear and Multilinear Algebra*, 9:271–288, 1981.