

Large System Transient Analysis of Adaptive Least Squares Filtering

Weimin Xiao, *Member, IEEE*, and Michael L. Honig, *Fellow, IEEE*

Abstract—The performance of adaptive least squares (LS) filtering is analyzed for the suppression of multiple-access interference. Both full-rank LS filters and reduced-rank LS filters, which reside in a lower dimensional Krylov space, are considered with training, and without training but with known signature for the desired user. We compute the large system limit of output signal-to-interference-plus-noise ratio (SINR) as a function of normalized observations, load, and noise level. Specifically, the number of users K , the degrees of freedom N , and the number of training symbols or observations i all tend to infinity with fixed ratios K/N and i/N . Our results account for an arbitrary power distribution over the users, data windowing (e.g., recursive LS (RLS) with exponential windowing), and initial diagonal loading of the covariance matrix to prevent ill-conditioning. Numerical results show that the large system analysis accurately predicts the simulated convergence performance of the algorithms considered with moderate degrees of freedom (typically $N = 32$). Given a fixed, short training length, the relative performance of full- and reduced-rank filters depends on the selected rank and diagonal loading. With an optimized diagonal loading factor, the performance of full- and reduced-rank filters are similar. However, full-rank performance is generally much more sensitive to the choice of diagonal loading factor than reduced-rank performance.

Index Terms—Adaptive filter, large system analysis, least squares (LS), reduced-rank filters.

I. INTRODUCTION

ADAPTIVE least squares (LS) filtering is a standard technique, which has been widely studied and applied to many communications applications, such as equalization and interference suppression, echo cancellation, and array processing [1]–[3]. Typically, the performance of the adaptive LS filter is described in terms of an output fidelity measure, such as mean-squared error (MSE) or signal-to-interference-plus-noise Ratio (SINR), which depends on the amount of training, or observed data. Related reduced-rank filtering techniques (adaptive and nonadaptive) have been presented in [4]–[12]. A reduced-rank

LS filter resides in a lower dimensional subspace, and can potentially perform better than a conventional (full-rank) LS filter with short training lengths [5], [11]–[14]. In addition, for some applications, the reduced-rank filter may be less complex than the analogous full-rank filter.

In this paper, we analyze the performance of adaptive full- and reduced-rank LS filtering when used to suppress multiple-access interference. Relevant applications are code-division multiple access (CDMA) with short signatures, and communications through multiple-input/multiple-output (MIMO) channels. Our analytical approach differs from prior work [1], [15]–[17] on the convergence of adaptive algorithms in that we focus on *large system* performance with randomly assigned signatures. Namely, we evaluate the output SINR of the adaptive filter as the number of users (or transmit antennas) K , the degrees of freedom N (e.g., processing gain or number of receive antennas), and the number of training symbols or observations i all tend to infinity with fixed ratios $\alpha = K/N$ and $\eta = i/N$. In this case, α is the normalized system load and η is the number of training symbols or observations normalized by the degrees of freedom. This approach is motivated by recent large system performance analyses of the linear minimum mean-square error (MMSE) receiver for CDMA with random spreading [18], [19]. Related work, which also has taken this approach, is presented in [20].

Our analytical results are used to compare the performance of full- and reduced-rank LS filters. The reduced-rank filters reside in a lower dimensional Krylov subspace, and have been proposed and studied in [8], [10]–[14], [19], [21]–[24]. (See also [25]–[29], which consider closely related filters.) This choice is motivated by the multistage Wiener filter (MSWF) implementation presented in [8], and the subsequent analysis in [19]. The MSWF is relatively simple compared with other reduced-rank filters, which require an eigendecomposition of the sample covariance matrix [4], [6], [9], [20]. Furthermore, it is shown in [19], [22] that the MSWF can achieve essentially full-rank performance with much lower rank than the other reduced-rank filters. For the MSWF, this rank does not scale with system size K and N , unlike the reduced-rank filters based on eigendecomposition. Adaptive filters based on the MSWF are presented in [12] along with simulation results, which show that the adaptive MSWF converges significantly faster than a full-rank LS filter.

The large system SINR is evaluated for full- and reduced-rank LS filters with a training sequence, and without a training sequence, but with known signature for the desired symbol sequence (i.e., corresponding to a CDMA user). Our results account for an arbitrary power distribution over the users, arbitrary data windowing, and initial diagonal loading of the

Manuscript received July 12, 2002; revised March 14, 2005. This work was supported in part by the Army Research Office under Grant DAAD19-99-1-0288 and by the National Science Foundation under Grant CCR-0073686. The material in this paper was presented in part at the Conference on Information Sciences and Systems, Princeton, NJ, March 2000 and at the 39th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2001.

W. Xiao was with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 USA. He is now with Motorola, Arlington Heights, IL 60004 USA (e-mail: wxiao1@motorola.com).

M. L. Honig is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 USA (e-mail: mh@ece.northwestern.edu).

Communicated by D. N. C. Tse, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2005.850081

sample covariance matrix. Data windowing allows past data to be discounted, the standard example being recursive LS (RLS) filtering with exponential windowing. Diagonal loading of the sample covariance matrix refers to initializing the sample covariance matrix as a small positive constant times the identity matrix, and is needed to prevent ill-conditioning with RLS filtering.

As $\eta \rightarrow \infty$, the output SINR of the (conventional) full-rank LS adaptive filter converges to the large system SINR for the MMSE filter. In contrast, for RLS adaptation with exponential weighting, as the normalized training length $\eta \rightarrow \infty$, the output SINR does not generally converge to the SINR for the MMSE filter. This is because exponential weighting effectively creates a finite-length data window, which prevents the adaptive filter from converging to the corresponding MMSE filter.

Large system convergence analysis requires evaluation of the moments of the corresponding sample covariance matrix. Without data windowing, the moments can be evaluated from expressions given in [30], [31]. We show that evaluation of the moments with data windowing is equivalent to solving a combinatorial (coloring) problem, which reduces to a simpler coloring problem without data windowing. These problems have closed-form solutions. An independent derivation of the large system moments considered here, which uses the theory of noncrossing partitions, is given in [23]. (See also [24], [32].) In that work, the moments are used to evaluate the large system performance of complex linear MMSE full- and reduced-rank filters with multiple antennas.

Numerical results are presented, which show that the large system analysis accurately predicts the performance of simulated finite systems (i.e., $N \geq 32$ for nearly all cases considered). A reduced-rank LS filter with appropriate rank typically requires significantly less training than the analogous full-rank LS filter, which is consistent with the results in [12]. We show, however, that given a specific training length (η), diagonal loading can significantly improve the performance of the full-rank LS filter. In contrast, the reduced-rank filter is less sensitive to both diagonal loading and exponential windowing. (Related results, based on simulation, are presented in [11].) Also, when N is sufficiently large, a low-rank filter requires less computation than the full-rank filter. We also present numerical results, which illustrate the effect of exponential windowing on the convergence of full-rank RLS filters. Additional large system numerical results, which also take into account the combination of partial despreading [33] with the reduced-rank filters considered here, are presented in [34] and [35].

We start with an analysis of full-rank LS filters. The model and filters are specified in the next section, and in Section III, we present our results on large system transient behavior (i.e., output SINR as a function of normalized training, or observations). The analytical approach, which is used to evaluate the large system moments of the sample covariance matrix, is described in Section IV. Expressions for the (negative) moments needed to evaluate the full-rank large system SINR are given in Section V. We then present large system transient results for reduced-rank LS filters in Section VI. A method for computing the positive moments needed to evaluate the reduced-rank SINR, and full-rank SINR with data windowing, is presented in Sec-

tion VII. Numerical results are given in Section VIII. Proofs and derivations are presented in the Appendices.

II. LINEAR LEAST SQUARES FILTERS

To simplify the analysis, we assume the idealized complex baseband model

$$\mathbf{r}(m) = \mathbf{S}\mathbf{A}\mathbf{b}(m) + \mathbf{n}(m) \quad (1)$$

where $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_K]$ is the $N \times K$ matrix of random signatures with independent and identically distributed (i.i.d.) elements, \mathbf{s}_k is the signature for user k , \mathbf{A} is the diagonal matrix of amplitudes, $\mathbf{b}(m)$ is the vector containing K symbols at time m (e.g., across users or antennas), and $\mathbf{n}(m)$ is the noise vector, which has covariance matrix $\sigma^2\mathbf{I}$. We assume that the symbol variance $E[|b_k(m)|^2] = 1$ for all k and m , where $b_k(m)$ is the m th symbol corresponding to the k th symbol stream (i.e., user or transmit antenna).

Throughout this paper, we will assume that the elements of signature matrix \mathbf{S} are Gaussian, although numerical results indicate that our analysis is applicable to more general distributions (e.g., binary). We denote $\mathbf{P} = \mathbf{A}\mathbf{A}^\dagger$ as the diagonal matrix of received powers, which are chosen from a distribution with finite moments. We also assume that \mathbf{S} , \mathbf{A} , $\mathbf{b}(m)$, and \mathbf{n} are independent.

In this section and the next, we consider only full-rank filters. Results for reduced-rank filters are subsequently presented in Section VI. In what follows, we assume that user 1 is the desired user. The RLS filter at time i , denoted as $\mathbf{c}(i)$, minimizes the cost function

$$\mathcal{E}(i) = \sum_{m=1}^i w_m |b_1(m) - \mathbf{c}^\dagger(i)\mathbf{r}(m)|^2 \quad (2)$$

where “ \dagger ” denotes complex conjugate transpose, $\{b_1(1), \dots, b_1(i)\}$ is part of a training sequence, and $\{w_m\}$ is a data windowing sequence. For example, exponential data windowing implies that $w_m = w^{i-m}$. The LS solution with diagonal loading is

$$\mathbf{c}(i) = \hat{\mathbf{R}}^{-1}(i)\hat{\mathbf{s}}_1(i) \quad (3)$$

where

$$\hat{\mathbf{R}}(i) = \frac{1}{i} \sum_{m=1}^i w_m \mathbf{r}(m)\mathbf{r}^\dagger(m) + \bar{\mu}(i)\mathbf{I} \quad (4)$$

$$\hat{\mathbf{s}}_1(i) = \frac{1}{i} \sum_{m=1}^i w_m \mathbf{r}(m)b_1^*(m) \quad (5)$$

and $\bar{\mu}(i) \geq 0$ is the diagonal loading term, which can be used to avoid ill-conditioning of the matrix $\hat{\mathbf{R}}(i)$ for small i . (Note that $\mathcal{E}(i)$ in (2) is minimized when $\bar{\mu}(i) = 0$.) Similarly, given the desired user’s signature \mathbf{s}_1 , we define a “blind” LS filter as

$$\mathbf{c}(i) = \hat{\mathbf{R}}^{-1}(i)\mathbf{s}_1. \quad (6)$$

In the presence of a time-varying channel, the recursions

$$\hat{\mathbf{R}}(i) = w\hat{\mathbf{R}}(i-1) + \mathbf{r}(i)\mathbf{r}^\dagger(i) \quad (7)$$

$$\hat{\mathbf{s}}_1(i) = w\hat{\mathbf{s}}_1(i-1) + b_1^*(i)\mathbf{r}(i) \quad (8)$$

with $w < 1$ can be used to track the time-varying steering vector and covariance matrix. (This corresponds to (4) and (5) with exponential weighting and $\bar{\mu}(i) = 0$.) In what follows, an *LS filter* refers to (3)–(5) with $w_m = 1$ for all m , whereas an *RLS filter* implies nonuniform data weighting (e.g., exponential with $w < 1$). As $i \rightarrow \infty$, the LS filter $\mathbf{c}(i)$ converges almost surely to the MMSE filter $\mathbf{R}^{-1}\mathbf{s}_1$, where \mathbf{R} is the covariance matrix

$$\mathbf{R} = E[\mathbf{r}(m)\mathbf{r}^\dagger(m)] = \mathbf{S}\mathbf{P}\mathbf{S}^\dagger + \sigma^2\mathbf{I}. \quad (9)$$

We emphasize that this is not true for RLS filters.

Let

$$\mathcal{R}(i) = [\mathbf{r}(1) \quad \cdots \quad \mathbf{r}(i)] \quad (10)$$

$$\mathbf{N}(i) = [\mathbf{n}(1) \quad \cdots \quad \mathbf{n}(i)] \quad (11)$$

$$\mathbf{B}(i) = [\mathbf{b}(1) \quad \cdots \quad \mathbf{b}(i)] \quad (12)$$

be the $N \times i$ matrices of received vectors, noise vectors, and $K \times i$ matrix of transmitted symbols, respectively. We can write

$$\hat{\mathbf{R}}(i) = \frac{1}{i}\mathcal{R}(i)\mathbf{W}(i)\mathcal{R}^\dagger(i) + \bar{\mu}(i)\mathbf{I} \quad (13)$$

$$= \frac{1}{i}[\mathbf{S}\mathbf{A}\mathbf{B}(i) + \mathbf{N}(i)]\mathbf{W}(i)[\mathbf{S}\mathbf{A}\mathbf{B}(i) + \mathbf{N}(i)]^\dagger + \bar{\mu}(i)\mathbf{I} \quad (14)$$

$$\hat{\mathbf{s}}_1(i) = \frac{1}{i}\mathcal{R}(i)\mathbf{W}(i)[\mathbf{B}^\dagger]_1 \quad (15)$$

where $[\mathbf{M}]_m$ denotes the m th column of the matrix \mathbf{M} , and $\mathbf{W}(i)$ is a diagonal data windowing matrix. For the LS filter without diagonal loading, $\mathbf{W}(i) = \mathbf{I}$ and $\bar{\mu}(i) = 0$. With exponential weighting $\mathbf{W}(i) = \text{diag}[w^{i-1}, w^{i-2}, \dots, w, 1]$.

The output SINR for RLS and LS filters with i training symbols, conditioned on the signature matrix \mathbf{S} , can be evaluated as

$$\beta_F(i) = \frac{P_1|\hat{\mathbf{s}}_1^\dagger(i)\hat{\mathbf{R}}^{-1}(i)\mathbf{s}_1|^2}{\sigma^2\hat{\mathbf{s}}_1^\dagger(i)\hat{\mathbf{R}}^{-2}(i)\hat{\mathbf{s}}_1(i) + \hat{\mathbf{s}}_1^\dagger(i)\hat{\mathbf{R}}^{-1}(i)\mathbf{S}_I\mathbf{A}_I^2\mathbf{S}_I^\dagger\hat{\mathbf{R}}^{-1}(i)\hat{\mathbf{s}}_1(i)} \quad (16)$$

where \mathbf{S}_I is the $N \times (K-1)$ matrix of signatures for the interferers (i.e., columns two through K of \mathbf{S}), \mathbf{A}_I is the diagonal matrix of interference amplitudes, and P_1 is the received power for user 1. The same expression applies to the blind RLS and LS filters, where $\hat{\mathbf{s}}_1(i)$ is replaced by \mathbf{s}_1 .

We are interested in evaluating the average SINR as a function of i where the average is with respect to the symbols, noise, and the random spreading sequences. This appears to be intractable; however, we are able to evaluate the large system limit of $\beta_F(i)$ as $(K, N, i) \rightarrow \infty$, $K/N = \alpha$, $i/N = \eta$. We find that this limit accurately predicts the performance of finite systems.

III. LARGE SYSTEM SINR

In this section, we give expressions for the large system output SINR for the LS and RLS filters presented in the preceding section. These expressions depend on the moments of the sample covariance matrix. In analogy with prior large system results for MMSE filters [18], the SINR converges to a

deterministic limit for fixed α and η . Computation of the large system moments is discussed in subsequent sections.

We first define a large system data window length \bar{L} . To motivate this definition, observe that any finite window length L becomes negligible in the large system limit as $i \rightarrow \infty$. For example, in the case of exponential weighting, if $w < 1$ then for any fixed η , $w^i \rightarrow 0$ as $(K, N, i) \rightarrow \infty$ and the effective large system window length is zero. Therefore, we must let the window length $L \rightarrow \infty$ in proportion to N . The effective window length is therefore defined as

$$\bar{L} = \frac{L}{N}.$$

Defining the average window length with exponential windowing as $L = 1/(1-w)$, the large system window length is then

$$\bar{L} = \frac{1}{(1-w)N}$$

and for fixed \bar{L} , we have for a finite system

$$w = 1 - \frac{1}{N\bar{L}}.$$

As $N \rightarrow \infty$, $w \rightarrow 1$ with rate depending on the effective window length.

In what follows, we will need the large system moments of \mathbf{W} defined as

$$f_k(\bar{L}) = \lim_{(N,i,L) \rightarrow \infty} \frac{1}{i} \text{trace}[\mathbf{W}^k(i)]$$

and assumed to be finite. For exponential windowing with fixed \bar{L} , we have

$$\begin{aligned} f_k(\bar{L}) &= \lim_{i \rightarrow \infty} \frac{1}{i} \sum_{m=1}^i (w^{i-m})^k = \frac{1}{\eta} \int_0^\eta e^{-\frac{tk}{\bar{L}}} dt \\ &= \frac{\bar{L}}{k\eta} \left(1 - e^{-\frac{k\eta}{\bar{L}}}\right) \end{aligned} \quad (17)$$

We will also need the function

$$\mathcal{W}_{k,m}(x) = \lim_{(N,i,L) \rightarrow \infty} \frac{1}{N} \sum_{t=1}^i \frac{w_t^k}{(1+w_t x)^m}. \quad (18)$$

To assess the effect of the diagonal loading term $\bar{\mu}(i)$ in (14), we normalize $\bar{\mu}(i)$ with respect to the training interval i . That is, if $\hat{\mathbf{R}}(0) = \bar{\mu}(0)\mathbf{I}$, then in the large system limit, $\bar{\mu}(i)$ in (14) should decrease as $(1/i)$. We therefore let $\bar{\mu}(i) = \mu(N/i) = (\mu/\eta)$, where $\mu \geq 0$ is a constant.

Let $\hat{\mathbf{R}}_I$ be the sample covariance matrix for the received interference plus noise

$$\begin{aligned} \hat{\mathbf{R}}_I(i) &= \frac{1}{i}[\mathbf{S}_I\mathbf{A}_I\mathbf{B}_I(i) + \mathbf{N}(i)]\mathbf{W}(i)[\mathbf{S}_I\mathbf{A}_I\mathbf{B}_I(i) \\ &\quad + \mathbf{N}(i)]^\dagger + \bar{\mu}(i)\mathbf{I} \end{aligned} \quad (19)$$

where $\mathbf{B}_I(i)$ is the $(K-1) \times i$ matrix of transmitted symbols from interferers, and let

$$\hat{\gamma}_n(\bar{L}, \mu) = \frac{1}{N} \text{trace} \left(\hat{\mathbf{R}}_I^n \right). \quad (20)$$

We will often omit the dependence on \bar{L} and $\bar{\mu}$.

Theorem 1: If \mathbf{S}_I has i.i.d. Gaussian elements, then as $(K, N, i) \rightarrow \infty$, with $(K/N) = \alpha$ and $(i/N) = \eta$, $\hat{\gamma}_n(\bar{L}, \mu)$ converges in probability to a deterministic limit $\hat{\gamma}_n^\infty(\bar{L}, \mu)$ for $n \geq 0$.

An outline of the proof is given in the next section, and the formal proof is given in Appendix I. Convergence of the eigenvalue distribution of $\hat{\mathbf{R}}_I$ to a deterministic limit then follows from the Moment Convergence Theorem (see [36]). This, in turn, implies the convergence of $\hat{\gamma}_{-1}$ and $\hat{\gamma}_{-2}$ to the corresponding limits, which appear in the expressions for large system SINR.

The next theorem presents the large system SINR for the most general model, which includes diagonal loading, a general power distribution, and a general data window. Although this expression is complicated, the corresponding expression without diagonal loading and uniform data windowing is relatively simple, and is presented as a corollary. Another corollary presents the expression for large system SINR with data windowing and infinite training. That can be used to compute the degradation in steady-state SINR due to different windowing schemes.

Theorem 2: As $(K, N, i) \rightarrow \infty$, with $(K/N) = \alpha$ and $(i/N) = \eta$, the output SINR for the RLS filter with i observations, given in (16), converges in probability to

$$\begin{aligned} \beta_F^\infty &= a_1^2 P_1 (\hat{\gamma}_{-1}^\infty)^2 \times \left\{ \sigma^2 \left(a_1^2 \hat{\gamma}_{-2}^\infty + a_2^2 \hat{\psi}_{-2}^\infty \right) \right. \\ &+ \alpha \left(a_1^2 - \frac{a_2^2 \mu \mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\eta^2 \mathcal{W}_{1,2}(\phi_{-1}^\infty)} \right) \\ &\times E_P \left[\frac{P \hat{\gamma}_{-2}^\infty + P^2 (\hat{\gamma}_{-1}^\infty)^2 \hat{\psi}_{-2}^\infty}{\left\{ 1 + P \hat{\gamma}_{-1}^\infty \left[f_1(\bar{L}) - \hat{\psi}_{-1}^\infty \right] \right\}^2} \right] \\ &\left. + \alpha \frac{a_2^2 \mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\eta \mathcal{W}_{1,2}(\phi_{-1}^\infty)} E_P \left[\frac{P \hat{\gamma}_{-1}^\infty}{1 + P \hat{\gamma}_{-1}^\infty \left[f_1(\bar{L}) - \hat{\psi}_{-1}^\infty \right]} \right] \right\}^{-1} \end{aligned} \quad (21)$$

where $F(P)$ is the limit distribution of received powers across users, $E_P[\cdot]$ denotes expectation over the power distribution $F(P)$

$$a_1 = \begin{cases} 1, & \text{blind RLS} \\ A_1 \left(f_1(\bar{L}) - \hat{\psi}_{-1}^\infty \right), & \text{with training} \end{cases} \quad (22)$$

$$a_2 = \begin{cases} -A_1 \hat{\gamma}_{-1}^\infty, & \text{blind RLS} \\ 1, & \text{with training} \end{cases} \quad (23)$$

and

$$\hat{\psi}_{-2}^\infty = \frac{\mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\mathcal{W}_{1,2}(\phi_{-1}^\infty) \eta} \left(\hat{\gamma}_{-1}^\infty - \frac{\mu \hat{\gamma}_{-2}^\infty}{\eta} \right) \quad (24)$$

$$\hat{\psi}_{-1}^\infty = \frac{\mathcal{W}_{2,1}(\phi_{-1}^\infty)}{\mathcal{W}_{1,1}(\phi_{-1}^\infty) \eta} \left(1 - \frac{\mu \hat{\gamma}_{-1}^\infty}{\eta} \right) \quad (25)$$

¹The authors thank M. Peacock for correcting a mistake, which appears in the corresponding theorem in [37], and in our original submission.

and

$$\phi_{-1}^\infty = \frac{\sigma^2}{\eta} \hat{\gamma}_{-1}^\infty + \frac{\alpha}{\eta} E_P \left[\frac{P \hat{\gamma}_{-1}^\infty}{1 + P \hat{\gamma}_{-1}^\infty \left[f_1(\bar{L}) - \hat{\psi}_{-1}^\infty \right]} \right] \quad (26)$$

The proof is given in Appendix II.

For the case where each user is received with the same power, $P = P_1 = 1$, (21)–(23) become

$$\begin{aligned} \beta_F^\infty &= a_1^2 (\hat{\gamma}_{-1}^\infty)^2 \times \left\{ \sigma^2 \left(a_1^2 \hat{\gamma}_{-2}^\infty + a_2^2 \hat{\psi}_{-2}^\infty \right) \right. \\ &+ \alpha \left(\left[a_1^2 - \frac{\mu \mathcal{W}_{2,2}(\phi_{-1}^\infty) a_2^2}{\eta^2 \mathcal{W}_{1,2}(\phi_{-1}^\infty)} \right] \right. \\ &\times \frac{\hat{\gamma}_{-2}^\infty + (\hat{\gamma}_{-1}^\infty)^2 \hat{\psi}_{-2}^\infty}{\left\{ 1 + \hat{\gamma}_{-1}^\infty \left[f_1(\bar{L}) - \hat{\psi}_{-1}^\infty \right] \right\}^2} \\ &\left. \left. + \frac{a_2^2 \mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\eta \mathcal{W}_{1,2}(\phi_{-1}^\infty)} \frac{\hat{\gamma}_{-1}^\infty}{1 + \hat{\gamma}_{-1}^\infty \left(f_1(\bar{L}) - \hat{\psi}_{-1}^\infty \right)} \right) \right\}^{-1} \end{aligned} \quad (27)$$

where a_1 and a_2 are defined by (22) and (23) with $A_1 = 1$.

The corresponding result for an adaptive LS filter with uniform data windowing and diagonal loading is obtained by setting $\mu = 0$ and $f_k(\bar{L}) = 1$ in Theorem 2. This is stated as the following corollary where

$$\gamma_n = \frac{1}{N} \text{trace}(\mathbf{R}_I^n) \quad (28)$$

is the n th moment of the interference-plus-noise covariance matrix $\mathbf{R}_I = \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger + \sigma^2 \mathbf{I}$, and

$$\gamma_n^\infty = \lim_{(K,N) \rightarrow \infty} \gamma_n = \lim_{(K,N) \rightarrow \infty} \frac{1}{N} \text{trace}(\mathbf{R}^n) \quad (29)$$

i.e., \mathbf{R}_I and \mathbf{R} have the same large system moments. Note that $P_1 \gamma_{-1}^\infty$ is the large system output SINR for the MMSE receiver [18].

Corollary 1: For the LS filter without diagonal loading, as $(K, N, i) \rightarrow \infty$, the output SINR converges in probability to

$$\beta_F^\infty = \frac{P_1 \gamma_{-1}^\infty}{1 + \frac{1}{\eta-1} (1 + P_1 \gamma_{-1}^\infty)} \quad \text{blind LS} \quad (30)$$

$$\beta_F^\infty = \frac{P_1 \gamma_{-1}^\infty}{1 + \frac{1}{\eta-1} \left(1 + \frac{1}{P_1 \gamma_{-1}^\infty} \right)} \quad \text{with training.} \quad (31)$$

The proof is given in Section V. Corollary 1 is motivated by a similar result for the blind LS filter presented in [20].² Corollary 1 implies the following properties for adaptive LS filters.

- 1) For both LS filters, as $\eta \rightarrow 1$ from above, $\beta_F^\infty \rightarrow 0$. That is, the large system output SINR is zero for $i = N$.
- 2) If the output SINR for the MMSE filter $P_1 \gamma_{-1}^\infty$ is large, then the SINR with training is approximately

$$\beta_F^\infty \approx P_1 \gamma_{-1}^\infty \times \left(1 + \frac{1}{\eta-1} \right)^{-1}$$

²The analogous result in [20] is an approximation, which is valid for large η . Specifically, the factor $(1)/(\eta-1)$ in (30) appears as $(1/\eta)$ in [20]. Training-based algorithms are not considered in [20].

and the degradation due to estimation error is $1 + \frac{1}{\eta-1}$. When $\eta = 2(i = 2N)$, the output SINR for the LS filter with training is approximately $\frac{P_1\gamma_{-1}^\infty}{2}$, i.e., 3 dB below the SINR for the MMSE filter, which is consistent with the classical result presented in [15].

- 3) If the SINR for the MMSE filter satisfies $P_1\gamma_{-1}^\infty > 1$, then the LS filter with training performs better than the blind LS filter, and vice versa if $P_1\gamma_{-1}^\infty < 1$.
- 4) As observed in [20], the blind LS filter is estimation error limited. Namely, as $P_1 \rightarrow \infty$, the output SINR approaches $\eta - 1$. For example, when $\eta = 1$, the output SINR is zero, independent of the power of the desired user. In contrast, the LS filter with training is *not* estimation error limited for $\eta > 1$. Namely, as $P_1 \rightarrow \infty$, the output SINR goes to infinity in proportion with P_1 .

Exponential weighting of the data effectively creates a finite-length window, so that as $\eta \rightarrow \infty$, the RLS filter does not converge to the MMSE filter. The large system SINR with infinite training is given as the following corollary to Theorem 2.

Corollary 2: As $\eta \rightarrow \infty$, the large system output SINR for the RLS filters with exponential windowing converges in probability to

$$\begin{aligned} \beta_F^\infty = & a_1^2 P_1 (\tilde{\gamma}_{-1}^\infty)^2 \times \left\{ \sigma^2 \left(a_1^2 \tilde{\gamma}_{-2}^\infty + a_2^2 \tilde{\psi}_{-2}^\infty \right) \right. \\ & + \alpha \left(a_1^2 - \frac{a_2^2 \mu \mathcal{W}_{2,2}(\tilde{\phi}_{-1}^\infty)}{\mathcal{W}_{1,2}(\tilde{\phi}_{-1}^\infty)} \right) \\ & \times E_P \left[\frac{P\tilde{\gamma}_{-2}^\infty + P^2\tilde{\psi}_{-2}^\infty}{\left[1 + P\tilde{\gamma}_{-1}^\infty (\bar{L} - \tilde{\psi}_{-1}^\infty) \right]^2} \right] \\ & \left. + \alpha \frac{a_2^2 \mathcal{W}_{2,2}(\tilde{\phi}_{-1}^\infty)}{\mathcal{W}_{1,2}(\tilde{\phi}_{-1}^\infty)} E_P \left[\frac{P\tilde{\gamma}_{-1}^\infty}{1 + P\tilde{\gamma}_{-1}^\infty (\bar{L} - \tilde{\psi}_{-1}^\infty)} \right] \right\}^{-1} \end{aligned} \quad (32)$$

where $\tilde{\gamma}_{-k}^\infty = \lim_{\eta \rightarrow \infty} \eta^{-k} \hat{\gamma}_{-k}^\infty$, $k = 1, 2$

$$a_1 = \begin{cases} 1, & \text{blind RLS} \\ A_1 (\bar{L} - \tilde{\psi}_{-1}^\infty), & \text{with training} \end{cases}$$

$$a_2 = \begin{cases} -A_1 \tilde{\gamma}_{-1}^\infty, & \text{blind RLS} \\ 1, & \text{with training} \end{cases}$$

and

$$\tilde{\psi}_{-2}^\infty = \frac{\mathcal{W}_{2,2}(\tilde{\phi}_{-1}^\infty)}{\mathcal{W}_{1,2}(\tilde{\phi}_{-1}^\infty)} (\tilde{\gamma}_{-1}^\infty - \mu \tilde{\gamma}_{-2}^\infty)$$

$$\tilde{\psi}_{-1}^\infty = \frac{\mathcal{W}_{2,1}(\tilde{\phi}_{-1}^\infty)}{\mathcal{W}_{1,1}(\tilde{\phi}_{-1}^\infty)} (1 - \mu \tilde{\gamma}_{-1}^\infty)$$

and

$$\tilde{\phi}_{-1}^\infty = \sigma^2 \tilde{\gamma}_{-1}^\infty + \alpha E_P \left[\frac{P\tilde{\gamma}_{-1}^\infty}{1 + P\tilde{\gamma}_{-1}^\infty [\bar{L} - \tilde{\psi}_{-1}^\infty]} \right]. \quad (33)$$

The proof is given in Appendix II-B. Corollary 2 shows that with infinite training, the asymptotic output SINR depends on both the normalized window length and the diagonal loading constant μ .

Examining the preceding expressions, it becomes apparent that computing the large system SINR requires the evaluation of the asymptotic moments

$$\hat{\gamma}_n^\infty = \frac{1}{N} \text{trace}[\hat{\mathbf{R}}_I^n(i)], \quad \text{for } n = -1 \text{ and } n = -2.$$

A method for computing these moments is described in the next section.

IV. ANALYTICAL APPROACH

In this section, we illustrate our approach to computing the large system moments, which appear in the SINR expressions presented in the preceding section. Equivalently, we wish to determine the large system eigenvalue distribution of the sample covariance matrix $\hat{\mathbf{R}}_I(i)$. We first show how that distribution can be derived in the absence of noise, and without data windowing and diagonal loading. We then show how those features can be included in the basic approach.

Ignoring the presence of noise, setting $w_m = 1$ for all m (no data windowing) and $\mu = 0$ (no diagonal loading), from (19) we have that

$$\begin{aligned} \hat{\mathbf{R}}_I(i) &= \frac{1}{i} \mathbf{S}_I \mathbf{A}_I \mathbf{B}_I(i) \mathbf{B}_I^\dagger(i) \mathbf{A}_I^\dagger \mathbf{S}_I^\dagger \\ &= \mathbf{S}_I \mathbf{V}_B(i) \Lambda_B(i) \mathbf{V}_B^\dagger(i) \mathbf{S}_I^\dagger \\ &= \mathbf{V}_S(i) \Lambda_B(i) \mathbf{V}_S^\dagger(i) \end{aligned} \quad (34)$$

where $\mathbf{V}_B(i)$ and $\Lambda_B(i)$ are the matrices of eigenvectors and eigenvalues associated with the covariance matrix $\frac{1}{i} \mathbf{A}_I \mathbf{B}_I(i) \mathbf{B}_I^\dagger(i) \mathbf{A}_I^\dagger$, and $\mathbf{V}_S(i) = \mathbf{S}_I \mathbf{V}_B(i)$. It is easy to show that $\frac{1}{i} \mathbf{A}_I \mathbf{B}_I(i) \mathbf{B}_I^\dagger(i) \mathbf{A}_I^\dagger$ and $\frac{1}{i} \mathbf{B}_I^\dagger(i) |\mathbf{A}_I|^2 \mathbf{B}_I(i)$ have the same nonzero eigenvalues. The large system eigenvalue distribution of the latter matrix as $(K, i) \rightarrow \infty$ is given in [38]. Consequently, the distribution of diagonal elements of $\Lambda_B(i)$ converges to a deterministic distribution, which depends on the distribution of powers.

It can be shown that for orthonormal $\mathbf{V}_B(i)$ and signature matrix \mathbf{S} with i.i.d. Gaussian elements, the elements of $\mathbf{V}_S(i)$ are i.i.d. Gaussian. (Our numerical results suggest that the elements of $\mathbf{V}_S(i)$ are Gaussian under more general assumptions.) Consequently, we can compute the asymptotic eigenvalue distribution of $\hat{\mathbf{R}}_I(i)$ as $(K, N, i) \rightarrow \infty$ by once again applying the result in [38], where the received ‘‘power distribution’’ is replaced by the asymptotic eigenvalue distribution of $\frac{1}{i} \mathbf{A}_I \mathbf{B}_I(i) \mathbf{B}_I^\dagger(i) \mathbf{A}_I^\dagger$. The fact that $\hat{\gamma}_k^\infty$ converges to a deterministic limit as $(K, N, i) \rightarrow \infty$ (Theorem 1) follows from the fact that the moments of $\mathbf{V} \mathbf{P} \mathbf{V}^\dagger$ converge to a deterministic limit, where \mathbf{V} is a matrix with i.i.d. random variables, and \mathbf{P} is diagonal with diagonal elements having the limit distribution $F(P)$, which has finite moments [18], [19].

The preceding approach can be extended to account for additive noise. Let $\mathbf{N}(i) = \tilde{\sigma} \mathbf{N}_S \mathbf{N}_B(i)$, where \mathbf{N}_S is an $N \times L$ random matrix with i.i.d. elements having the same distribution as the elements of \mathbf{S} , $\mathbf{N}_B(i)$ is an $L \times i$ random matrix with i.i.d. elements having the same distribution as the elements of \mathbf{B} , and

$\tilde{\sigma} = \sqrt{(N\sigma^2)/L}$. As $L \rightarrow \infty$, the elements of $\tilde{\sigma}\mathbf{N}_S\mathbf{N}_B(i)$ converge to i.i.d. Gaussian random variables with zero mean and variance σ^2 . Hence we rewrite $\hat{\mathbf{R}}_I(i)$ as

$$\begin{aligned} \hat{\mathbf{R}}_I(i) &= \frac{1}{i} [\mathbf{S}_I \quad \mathbf{N}_S] \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B}_I(i) \\ \mathbf{N}_B(i) \end{bmatrix} \\ &\quad \times \left([\mathbf{S}_I \quad \mathbf{N}_S] \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B}_I(i) \\ \mathbf{N}_B(i) \end{bmatrix} \right)^\dagger \\ &= [\mathbf{S}_I \quad \mathbf{N}_S] \mathbf{\Omega} [\mathbf{S}_I \quad \mathbf{N}_S]^\dagger \end{aligned} \quad (35)$$

where

$$\begin{aligned} \mathbf{\Omega} &= \frac{1}{i} \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B}_I(i) \\ \mathbf{N}_B(i) \end{bmatrix} \\ &\quad \times [\mathbf{B}_I^\dagger(i) \quad \mathbf{N}_B^\dagger(i)] \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix} \end{aligned} \quad (36)$$

and $\mathbf{\Omega}$, \mathbf{S}_I , and \mathbf{N}_S are independent. The same approach as that described from (34) can now be applied to computing the moments of $\hat{\mathbf{R}}_I(i)$ where $[\mathbf{S}_I \quad \mathbf{N}_S]$ replaces \mathbf{S}_I , and $\mathbf{\Omega}$ replaces $\frac{1}{i}\mathbf{A}_I\mathbf{B}_I\mathbf{B}_I^\dagger\mathbf{A}_I^\dagger$.

With data windowing and $\mu > 0$, we have

$$\begin{aligned} \hat{\mathbf{R}}_I(i) &= \frac{1}{i} [\mathbf{S}_I \quad \mathbf{N}_S] \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B}_I(i) \\ \mathbf{N}_B(i) \end{bmatrix} \mathbf{W}(i) \\ &\quad \times \left([\mathbf{S}_I \quad \mathbf{N}_S] \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B}_I(i) \\ \mathbf{N}_B(i) \end{bmatrix} \right)^\dagger + \mu(i)\mathbf{I} \\ &= [\mathbf{S}_I \quad \mathbf{N}_S] \mathbf{\Omega} [\mathbf{S}_I \quad \mathbf{N}_S]^\dagger + \mu(i)\mathbf{I} \end{aligned} \quad (37)$$

where

$$\mathbf{\Omega} = \frac{1}{i} \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B}_I(i) \\ \mathbf{N}_B(i) \end{bmatrix} \mathbf{W}(i) [\mathbf{B}_I^\dagger(i) \quad \mathbf{N}_B^\dagger(i)] \begin{bmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}\mathbf{I} \end{bmatrix}$$

Once again, the preceding approach applies with the appropriate substitutions in (34). Specifically, with $\mu = 0$, $\hat{\mathbf{R}}_I(i)$ has the form $\mathbf{V}\mathbf{P}\mathbf{V}^\dagger$, where the elements of \mathbf{V} are i.i.d. and \mathbf{P} is a diagonal matrix. The large system distribution of the diagonal elements of \mathbf{P} is now the large system eigenvalue distribution of $\mathbf{\Omega}$.³ Since for matrices \mathbf{M}_1 and \mathbf{M}_2 , the moments of $\mathbf{M}_1\mathbf{M}_2$, and $\mathbf{M}_2\mathbf{M}_1$ are the same, computing the large system moments of $\mathbf{\Omega}$ is equivalent to computing the moments of $\mathbf{W}\mathbf{V}\mathbf{P}\mathbf{V}^\dagger$ where \mathbf{V} has i.i.d. elements, and \mathbf{W} and \mathbf{P} are diagonal matrices.

To include diagonal loading in the analysis, let $\hat{\mathbf{R}}_I(i; \mu)$ denote the sample covariance matrix as a function of diagonal loading factor μ . The moments of $\hat{\mathbf{R}}_I$ can be written as

$$\begin{aligned} \hat{\gamma}_n(\bar{L}, \mu) &= \frac{1}{N} \text{trace} \left[\left(\hat{\mathbf{R}}_I(i; 0) + \frac{\mu}{\eta} \mathbf{I} \right)^n \right] \\ &= \sum_{k=0}^n \frac{n!}{(n-k)!k!} \left(\frac{\mu}{\eta} \right)^{n-k} \left(\frac{1}{N} \text{trace} \left[\hat{\mathbf{R}}_I^n(i; 0) \right] \right) \\ &= \sum_{k=0}^n \frac{n!}{(n-k)!k!} \left(\frac{\mu}{\eta} \right)^{n-k} \hat{\gamma}_k(\bar{L}, 0). \end{aligned} \quad (38)$$

³In what follows, we will compute the large system positive moments of $\hat{\mathbf{R}}_I(i)$, i.e., $\hat{\gamma}_n^\infty, n > 0$. We note that the n th positive moment depends on \mathbf{P} only through the first n positive moments of \mathbf{P} , or equivalently, $\mathbf{\Omega}$.

That is, the moments with diagonal loading can be expressed as a function of the moments without diagonal loading.

We conclude that with arbitrary data windowing and diagonal loading, evaluating $\hat{\gamma}_n(\bar{L}, \mu)$ reduces to evaluating the n th asymptotic moment of $\mathbf{W}\mathbf{V}\mathbf{P}\mathbf{V}^\dagger$. With uniform data windowing, the moments $\hat{\gamma}_n^\infty, n = -1, -2$, can be computed directly using this approach. Those expressions are presented in the next section. For the most general scenario with nonuniform data windowing and a nonuniform power distribution, we show in the next section how to compute $\hat{\gamma}_{-1}^\infty$ and $\hat{\gamma}_{-2}^\infty$ in terms of positive moments $\hat{\gamma}_n^\infty, n > 0$. In Section VII, we present a combinatorial method for computing the positive moments, which are also needed to evaluate the SINR for reduced-rank LS filters, to be discussed in Section VI. We remark that the Stieltjes transform of the large system eigenvalue distribution of $\mathbf{W}\mathbf{V}\mathbf{P}\mathbf{V}^\dagger$ is derived in [39]. This can also be used to compute the large system moments, although the method presented in Section VII requires much less computation.

V. COMPUTATION OF NEGATIVE MOMENTS

We now present expressions for the negative moments of the sample covariance matrix, which are needed to compute the large system SINR, given by (21) and (27). Here we assume that these limits exist, and ignore associated convergence issues.⁴

Let $\hat{\lambda}_1, \dots, \hat{\lambda}_N$ denote the N eigenvalues of the sample covariance matrix $\hat{\mathbf{R}}$, and the set $\mathcal{A}(x) = \{\lambda_j : \lambda_j < x\}$. The corresponding empirical distribution function is defined as $|\mathcal{A}(x)|/N$. For the time being, we assume uniform data windowing and no diagonal loading. In that case, the discussion in the preceding section combined with the convergence results in [38] implies that as $(K, N, i) \rightarrow \infty$, the empirical eigenvalue distribution converges to a deterministic large system eigenvalue distribution $G_{\hat{\mathbf{R}}}(\cdot)$.

Let the Stieltjes transform of $G_{\hat{\mathbf{R}}}$ be

$$m_{\hat{\mathbf{R}}}(z) = \int \frac{1}{\lambda - z} dG_{\hat{\mathbf{R}}}(\lambda) \quad (39)$$

for $z \in \mathcal{C}^+$, i.e., z must be complex with positive imaginary part. The Stieltjes transform of the large system eigenvalue distribution for the covariance matrix \mathbf{R} in (9) satisfies [38]

$$m_{\mathbf{R}}(z) = \frac{1}{-z + \alpha \int \frac{\tau dF_P(\tau)}{1 + m_{\mathbf{R}}(z)}} \quad (40)$$

where $F_P(\cdot)$ is the large system limit of the power distribution. In Appendix III, we show that the corresponding Stieltjes

⁴Also in what follows, we encounter expressions of the form $X_k^{(N)} = \mathbf{v}^\dagger \mathbf{R}_{I,k}^{-n} \mathbf{v}$ for $n = 1, 2$, where \mathbf{v} is a random vector (e.g., a signature) and $\mathbf{R}_{I,k}$ is the interference-plus-noise covariance matrix for user k . (In some cases, $\mathbf{R}_{I,k}$ is replaced by a sample covariance matrix.) We will implicitly assume that the sequence $\{X_k^{(N)}\}$ converges uniformly over k (almost surely) to the corresponding limit so that

$$\lim_{N \rightarrow \infty} (1/N) \sum_{k=1}^N X_k^{(N)} = \lim_{N \rightarrow \infty} 1/N \sum_{k=1}^N \left(\lim_{N \rightarrow \infty} X_k^{(N)} \right).$$

This enables the computation of similar types of limiting sums.

transform for the *sample* covariance matrix without exponential weighting and diagonal loading is

$$m_{\hat{\mathbf{R}}}(z) \left(\frac{\eta - 1 - m_{\hat{\mathbf{R}}}(z)z}{\eta} \right) = \frac{1 + m_{\hat{\mathbf{R}}}(z)z}{\sigma^2 + \alpha \int \frac{PdF_P(P)}{1 + Pm_{\hat{\mathbf{R}}}(z) \left(\frac{\eta - 1 - m_{\hat{\mathbf{R}}}(z)z}{\eta} \right)}}. \quad (41)$$

The following relations are derived in Appendix III:

$$\hat{\gamma}_{-1}^{\infty} = \frac{\gamma_{-1}^{\infty}}{1 - \frac{1}{\eta}} \quad (42)$$

$$\hat{\gamma}_{-2}^{\infty} = \frac{(\eta - 1)\gamma_{-2}^{\infty} + (\gamma_{-1}^{\infty})^2}{(\eta - 1)(1 - \frac{1}{\eta})^2} \quad (43)$$

$$\gamma_{-1}^{\infty} = \frac{1}{\sigma^2 + \alpha \int_0^{\infty} \frac{PdF(P)}{1 + P\gamma_{-1}^{\infty}}} \quad (44)$$

$$\gamma_{-2}^{\infty} = \frac{\gamma_{-1}^{\infty}}{\sigma^2 + \alpha \int_0^{\infty} \frac{PdF(P)}{[1 + P\gamma_{-1}^{\infty}]^2}}. \quad (45)$$

These relations follow from (40) and (41), although the derivations of (43)–(45) in Appendix III are more direct. In particular, we give a simple, direct derivation of the Tse–Hanly formula for the output MMSE (44), [18]. Combining (21) with (42)–(45) gives Corollary 1.

For the case $\mathbf{W}(i) = \mathbf{I}$ and $\mu > 0$, we observe that

$$\hat{\gamma}_{-1} = \lim_{z \rightarrow \mu/\eta} m_{\hat{\mathbf{R}}}(z)$$

where $m_{\hat{\mathbf{R}}}(z)$ satisfies (41). With equal-power users, this gives the condition

$$a_4(\hat{\gamma}_{-1})^4 + a_3(\hat{\gamma}_{-1})^3 + a_2(\hat{\gamma}_{-1})^2 + a_1\hat{\gamma}_{-1} - \frac{1}{\eta} = 0 \quad (46)$$

where

$$\begin{aligned} a_1 &= \left(\frac{\eta - 1}{\eta^2} \right) (\alpha + \sigma^2 - 1) + \frac{\mu}{\eta} \\ a_2 &= \frac{\eta - 1}{\eta^2} \left(\mu + \frac{\eta - 1}{\eta} \sigma^2 \right) + \frac{\mu}{\eta^2} (\alpha + \sigma^2) \\ a_3 &= \left(\frac{\mu}{\eta} \right)^2 + \frac{2(\eta - 1)\mu\sigma^2}{\eta^3} \\ a_4 &= \frac{\mu^2\sigma^2}{\eta^3} \end{aligned}$$

and the solution must satisfy $0 \leq \hat{\gamma}_{-1} \leq (\eta/\mu)$. Similar manipulations as those used to derive (43) give (47) at the bottom of the page, where $0 \leq \hat{\gamma}_{-2}^{\infty} \leq 1/(\mu^2)$.

To compute the SINR versus observations with nonuniform data windowing and diagonal loading, we express the output

SINR of the full-rank filter in terms of the SINR of a *reduced-rank* filter with sufficiently large rank. Namely, we observe that $\hat{\gamma}_{-1}$ is the output SINR of a linear MMSE receiver given the input covariance matrix $\hat{\mathbf{R}}_I(i)$ [18]. It is shown in [19] that

$$\hat{\gamma}_{-1}^{\infty} = \lim_{D \rightarrow \infty} \hat{\gamma}_{0:D-1}^{\dagger} \hat{\Gamma}_{1:D}^{-1} \hat{\gamma}_{0:D-1} \quad (48)$$

where

$$\hat{\gamma}_{l:m} = [\hat{\gamma}_l^{\infty} \quad \cdots \quad \hat{\gamma}_m^{\infty}]^T \quad (49)$$

$$\hat{\Gamma}_{l:l+m} = [\hat{\gamma}_{l:m} \quad \cdots \quad \hat{\gamma}_{l+m:l+2m}]. \quad (50)$$

The expression on the right-hand side of (48) is the output SINR of the corresponding reduced-rank MMSE filter as the rank $D \rightarrow \infty$. (See the discussion in the next section.) Computation of the positive moments $\hat{\gamma}_n^{\infty}$, $n > 0$, is discussed in Section VII. It is observed in [19] that the SINR converges rapidly to the full-rank SINR as D increases, so that taking $D = 10$ is typically adequate.

To compute $\hat{\gamma}_{-2}^{\infty}$, we observe that

$$\hat{\gamma}_{-2}^{\infty} = \int \frac{1}{\lambda^2} dG_{\hat{\mathbf{R}}}(\lambda) = -\frac{\partial}{\partial x} \int \frac{1}{\lambda + x} dG_{\hat{\mathbf{R}}}(\lambda) \Big|_{x=0}. \quad (51)$$

Let

$$\hat{\gamma}_n^{\infty}[x] = \int (\lambda + x)^n dG_{\hat{\mathbf{R}}} = \int \lambda^n dG_{\hat{\mathbf{R}}'} \quad (52)$$

where $\hat{\mathbf{R}}' = \hat{\mathbf{R}} + x\mathbf{I}$, and $\hat{\gamma}_n^{\infty}[0] = \hat{\gamma}_n^{\infty}$. It is easily shown that (48) applies when the moments are computed from $\hat{\mathbf{R}}'$, as well as $\hat{\mathbf{R}}$, so that we can view the moments in (48) as functions of x , as defined in (52). Taking the derivative, as in (51), gives

$$\hat{\gamma}_{-2}^{\infty} = \lim_{D \rightarrow \infty} \left(\hat{\gamma}_{0:D-1}^{\dagger} \hat{\Gamma}_{1:D}^{-1} \hat{\Gamma}_{1:D}^{-1} \hat{\gamma}_{0:D-1} - \hat{\gamma}_{0:D-1}^{\dagger} \hat{\Gamma}_{1:D}^{-1} \hat{\gamma}_{0:D-1} - \hat{\gamma}_{0:D-1}^{\dagger} \hat{\Gamma}_{1:D}^{-1} \hat{\gamma}_{0:D-1} \right) \quad (53)$$

where

$$\hat{\gamma}_{l:m} = [l\hat{\gamma}_{l-1}^{\infty} \quad \cdots \quad m\hat{\gamma}_{m-1}^{\infty}]^T \quad (54)$$

$$\hat{\Gamma}_{l:l+m} = [\hat{\gamma}_{l:m} \quad \cdots \quad \hat{\gamma}_{l+m:l+2m}]. \quad (55)$$

VI. LARGE SYSTEM SINR FOR REDUCED-RANK LS FILTERS

In this section, we present large system SINR expressions for reduced-rank LS filters. These will be used to compare the transient performance of reduced- and full-rank LS filters as a function of training with different diagonal loading factors and exponential weighting factors.

A reduced-rank filter projects the received vectors onto a lower dimensional subspace. The filtering and estimation then occur within this subspace. Let \mathbf{M}_D be the $N \times D$ matrix of

$$\hat{\gamma}_{-2}^{\infty} = \frac{\hat{\gamma}_{-1}^{\infty}}{\mu + \frac{\eta}{\hat{\gamma}_{-1}^{\infty}} \left(1 - \frac{1}{\eta} + \mu \frac{\hat{\gamma}_{-1}^{\infty}}{\eta} \right) \left(1 - \frac{1}{1 + \frac{\sigma^2 \hat{\gamma}_{-1}^{\infty}}{\eta} + \frac{\alpha}{\eta(1 + \hat{\gamma}_{-1}^{\infty}(1 - \frac{1}{\eta}))}} \right)} \quad (47)$$

D basis vectors, which span a D -dimensional subspace, where $D < N$. The projected received vector is given by

$$\tilde{\mathbf{r}}(i) = \mathbf{M}_D^\dagger \mathbf{r}(i) \quad (56)$$

and the filter output at time i is

$$z(i) = \tilde{\mathbf{c}}^\dagger(i) \tilde{\mathbf{r}}(i) \quad (57)$$

where $\tilde{\mathbf{c}}$ is $D \times 1$. Selecting $\tilde{\mathbf{c}}(i)$ to minimize the LS cost function $\sum_{m=1}^i w^{i-m} |b_1(m) - \tilde{\mathbf{c}}^\dagger(i) \tilde{\mathbf{r}}(m)|^2$ gives

$$\tilde{\mathbf{c}}(i) = \left[\mathbf{M}_D^\dagger \hat{\mathbf{R}}(i) \mathbf{M}_D \right]^{-1} \mathbf{M}_D^\dagger \hat{\mathbf{s}}_1(i). \quad (58)$$

The reduced-rank LS filter considered here is presented in [12], and for training length i uses the matrix of basis vectors

$$\mathbf{M}_D(i) = [\hat{\mathbf{s}}_1(i) \hat{\mathbf{R}}(i) \hat{\mathbf{s}}_1 \hat{\mathbf{R}}^2(i) \hat{\mathbf{s}}_1(i) \dots \hat{\mathbf{R}}^{D-1}(i) \hat{\mathbf{s}}_1(i)] \quad (59)$$

which define a Krylov subspace. This reduced-rank LS filter is equivalent to the adaptive MSWF presented in [12]. It is also equivalent to the adaptive auxiliary-vector reduced-rank filter presented in [10], [11], provided that the estimated covariance matrices are the same in each case, and the combining coefficients are selected to minimize the LS criterion [12], [21].

The output SINR for the rank- D adaptive filter with i training symbols can be written as

$$\beta_D(i) = \frac{P_1 \mathcal{S}}{\mathcal{N} + \mathcal{J}} \quad (60)$$

where

$$\mathcal{S} = \left| \hat{\mathbf{g}}_{0:D-1}^\dagger(i) \hat{\mathbf{G}}_{1:D}^{-1}(i) \mathbf{g}_{0:D-1}(i) \right|^2 \quad (61)$$

$$\mathcal{N} = \sigma^2 \hat{\mathbf{g}}_{0:D-1}^\dagger(i) \hat{\mathbf{G}}_{1:D}^{-1}(i) \hat{\mathbf{G}}_{0:D-1}(i) \hat{\mathbf{G}}_{1:D}^{-1}(i) \hat{\mathbf{g}}_{0:D-1}(i) \quad (62)$$

$$\mathcal{J} = \hat{\mathbf{g}}_{0:D-1}^\dagger(i) \hat{\mathbf{G}}_{1:D}^{-1}(i) \hat{\mathbf{G}}_{0:D-1}(i) \hat{\mathbf{G}}_{1:D}^{-1}(i) \hat{\mathbf{g}}_{0:D-1}(i) \quad (63)$$

and

$$g_k(i) = \hat{\mathbf{s}}_1^\dagger(i) \hat{\mathbf{R}}^k(i) \mathbf{s}_1 \quad (64)$$

$$\mathbf{g}_{l:m}(i) = [g_l(i) \quad \dots \quad g_m(i)]^T \quad (65)$$

$$\hat{g}_k(i) = \hat{\mathbf{s}}_1^\dagger(i) \hat{\mathbf{R}}^k(i) \hat{\mathbf{s}}_1(i) \quad (66)$$

$$\hat{\mathbf{g}}_{l:m}(i) = [\hat{g}_l(i) \quad \dots \quad \hat{g}_m(i)]^T \quad (67)$$

$$\hat{\mathbf{G}}_{l:l+m}(i) = [\hat{g}_{l+l+m}(i) \quad \dots \quad \hat{g}_{l+m:l+2m}(i)] \quad (68)$$

$$\check{g}_{lm}(i) = \hat{\mathbf{s}}_1^\dagger(i) \hat{\mathbf{R}}^l(i) \mathbf{S}_l \mathbf{P} \mathbf{S}_l^\dagger \hat{\mathbf{R}}^m(i) \hat{\mathbf{s}}_1(i) \quad (69)$$

$$\check{\mathbf{G}}_{0:D-1}(i) = [\check{g}_{lm}(i)] \quad (D \times D). \quad (70)$$

For the blind adaptive filters, the preceding expressions hold where $\hat{\mathbf{s}}_1(i)$ is replaced by \mathbf{s}_1 .

Theorem 3: As $(K, N, i) \rightarrow \infty$, with $(K/N) = \alpha$ and $(i/N) = \eta$, the output SINR for the reduced-rank LS filter with i observations, given by (60), converges in probability to β_D^∞ , also given by (60), where all terms in (64)–(70) are replaced by their associated large system limits.

Denote the large system limits of $g_k(i)$, $\hat{g}_k(i)$ and $\check{g}_{lm}(i)$ as g_k^∞ , \hat{g}_k^∞ , and \check{g}_{lm}^∞ , respectively. It is straightforward to show by induction that each of these terms is a deterministic function of $\hat{\gamma}_n^\infty$, $1 \leq n \leq 2D + 2$, and other related terms. Recall that computing $\hat{\gamma}_n^\infty$ reduces to computing the asymptotic moments of a matrix with the form \mathbf{WVPV}^\dagger where \mathbf{V} is i.i.d. and \mathbf{W} and \mathbf{P} are diagonal. Computing the previous reduced-rank terms g_k^∞ , \hat{g}_k^∞ , and \check{g}_{lm}^∞ , with nonuniform data windowing

requires the computation of additional moments, which have the forms $\text{trace}[\mathbf{W}^k \mathbf{V} \mathbf{P} \mathbf{V}^\dagger (\mathbf{W} \mathbf{V} \mathbf{P} \mathbf{V}^\dagger)^n]$, for $k = 2, 3$, and $\text{trace}[(\mathbf{W}^2 \mathbf{V} \mathbf{P} \mathbf{V}^\dagger) (\mathbf{W} \mathbf{V} \mathbf{P} \mathbf{V}^\dagger)^m (\mathbf{W}^2 \mathbf{V} \mathbf{P} \mathbf{V}^\dagger) (\mathbf{W} \mathbf{V} \mathbf{P} \mathbf{V}^\dagger)^n]$. It can be shown that these terms also converge as $N \rightarrow \infty$, and that the limits can be efficiently computed. We omit the proof of Theorem 3 along with the associated computations. Further details are presented in [34].

VII. COMPUTATION OF POSITIVE MOMENTS

As discussed in Sections IV and V, the moments $\hat{\gamma}_{-2}^\infty$ and $\hat{\gamma}_{-1}^\infty$, which appear in the expressions for large system SINR for the LS filter, can be computed in terms of the *positive* moments $\hat{\gamma}_n^\infty$, $n > 0$. (This corresponds to a reduced-rank approximation when the number of moments used is finite.) Computing those moments reduces to computing the large system limit of

$$\gamma_n = \frac{1}{M} \text{trace}(\mathbf{W} \mathbf{V} \mathbf{P} \mathbf{V}^\dagger)^n \quad (71)$$

for $n > 0$. As discussed in Sections V and VI, the positive moments are needed to compute the output SINR of the associated reduced-rank LS filter.

In what follows, \mathbf{V} is an $M \times K$ matrix with i.i.d. elements, \mathbf{W} is a diagonal $M \times M$ matrix, and \mathbf{P} is a diagonal $K \times K$ matrix. The large system limit will be indicated by $M \rightarrow \infty$, where $\alpha = K/M$ is constant. We will denote the corresponding n th large system moment as γ_n^∞ . This is a generalization of the large system moment defined by (28), for which $\mathbf{W} = \mathbf{I}$.

It is shown in [31], [38] that $\gamma_n \rightarrow \gamma_n^\infty$ in probability when $\mathbf{W} = \mathbf{I}$. In Appendix I-A, we prove convergence in probability for an arbitrary windowing matrix \mathbf{W} , assuming that the distribution of diagonal elements has finite moments. Here we assume convergence, and present a method for computing the limit. A similar analysis, which relies on the theory of non-crossing partitions, is presented in [32] and [23].

If $\mathbf{W} = \mathbf{I}$, then the large system moments of $\mathbf{V} \mathbf{P} \mathbf{V}^\dagger$ are given by [31]

$$\gamma_n^\infty = \sum_{k=1}^n \alpha^{n-k+1} \sum \frac{n!}{d_1! \dots d_k! k!} E^{d_1}[P] \dots E^{d_k}[P^k] \quad (72)$$

where the inner summation is over all nonnegative solutions to the equations

$$\begin{aligned} d_1 + d_2 + \dots + d_k &= n - k + 1 \\ d_1 + 2d_2 + \dots + kd_k &= n \end{aligned}$$

With equal power users, i.e., $\mathbf{P} = \mathbf{P} \mathbf{I}_K$, where the subscript of \mathbf{I} denotes the dimension, the preceding expression becomes

$$\gamma_n^\infty = P^n \sum_{k=1}^n \frac{1}{k} \binom{n}{k-1} \binom{n-1}{k-1} \alpha^k, \quad n = 1, 2, \dots \quad (73)$$

which was first presented in [30].

In what follows, we show that computation of the large system moments of the random matrix \mathbf{WVPV}^\dagger with arbitrary distributions for \mathbf{P} and \mathbf{W} is equivalent to a combinatorial coloring problem. We start with the case $\mathbf{W} = \mathbf{I}$, which leads to the following simpler coloring problem.

Coloring Problem: Consider n balls arranged in a circle and numbered from 1 to n as shown in Fig. 1. To each ball we assign

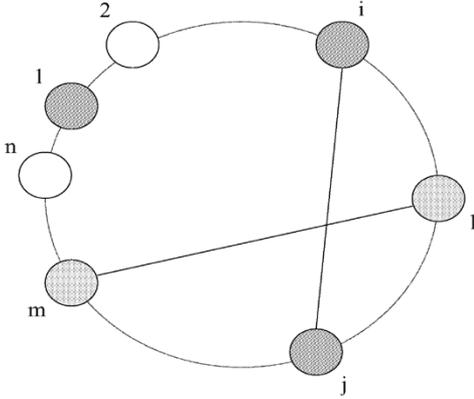


Fig. 1. Illustration of the coloring problem. This is an invalid coloration since lines connecting two balls with the same color cross.

one of k colors. Let \mathcal{C}_m denote the set of balls assigned color $m, 1 \leq m \leq k$, and let $c_m = |\mathcal{C}_m|$, the number of balls in \mathcal{C}_m . The sets $\{\mathcal{C}_m\}$ are sorted so that $c_1 \leq c_2 \leq \dots \leq c_k$. Given (c_1, \dots, c_k) , find the number of colorations that satisfy the following conditions.

- Given two balls $j, l \in \mathcal{C}_s$ and two balls $m, n \in \mathcal{C}_t, s \neq t$, the line that connects j and l does not cross the line that connects m and n . (For example, the coloration shown in Fig. 1 is not valid.)
- Two colorations are counted the same if one is obtained from the other by exchanging colors. (For example, white becomes red, and the original red becomes white.)

We denote the solution to this problem as $\mathcal{X}_k^{(n)}(c_1, \dots, c_k)$.

Theorem 4 (Large System Moments Without Data Windowing): Assume that the diagonal entries of \mathbf{P} have limit distribution $F(\mathbf{P})$ with finite moments $E[P^k]$ for $0 \leq k \leq n$, where $n \geq 0$. The n th large system eigenvalue moment can be evaluated as

$$\gamma_n^\infty = \sum_{k=1}^n \alpha^k \Lambda_k^{(n)} \quad (74)$$

where

$$\Lambda_k^{(n)} = \sum_{(c_1, \dots, c_k) \in \mathcal{S}_k^{(n)}} \mathcal{X}_k^{(n)}(c_1, \dots, c_k) \left(\prod_{t=1}^k E[P^{c_t}] \right) \quad (75)$$

and

$$\mathcal{S}_k^{(n)} = \left\{ (c_1, c_2, \dots, c_k) \mid 1 \leq c_1 \leq c_2 \leq \dots \leq c_k, \sum_{t=1}^k c_t = n \right\} \quad (76)$$

The proof is given in Appendix I-A.

Comparing Theorem 4 with (72), it can be shown that

$$\mathcal{X}_k^{(n)}(c_1, \dots, c_k) = \frac{n!}{d_1! \dots d_{n-k+1}! (n-k+1)!} \quad (77)$$

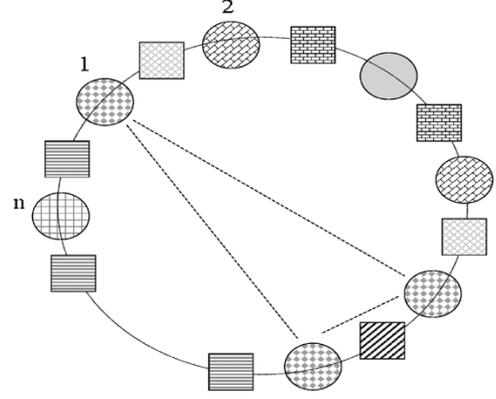


Fig. 2. Illustration of the double-coloring problem.

where $d_j, 1 \leq j \leq n - k + 1$, is the multiplicity of $c_s = j, 1 \leq s \leq k$ in the vector (c_1, \dots, c_k) . Combining (73) with (75), assuming equal powers, gives

$$\begin{aligned} & \frac{\binom{n}{k-1} \binom{n-1}{k-1}}{\binom{k-1}{k-1}} \\ &= \sum_{(c_1, \dots, c_k) \in \mathcal{S}_k^{(n)}} \mathcal{X}_k^{(n)}(c_1, \dots, c_k) \\ &= \sum_{(c_1, \dots, c_{n-k+1}) \in \mathcal{S}_{n-k+1}^{(n)}} \mathcal{X}_{n-k+1}^{(n)}(c_1, \dots, c_{n-k+1}). \quad (78) \end{aligned}$$

Theorem 5 (Convergence of Moments With Windowed Data): If the diagonal entries of \mathbf{P} have limit distribution $F(\mathbf{P})$ with finite moments $E[P^k]$ for $0 \leq k \leq n$, where $n \geq 0$, and \mathbf{W} has finite moments $f_k(\bar{L})$ for $0 \leq k \leq n$, then the n th moment of \mathbf{WVPV}^\dagger converges in probability as $M \rightarrow \infty$.

The proof is given in Appendix I-A.

With data windowing, evaluation of the large system moments relies on the solution to the following coloring problem, which is a more complicated version of the preceding coloring problem.

Double-Coloring Problem: Consider n balls arranged in a circle with one square between each two neighboring balls, and numbered from 1 to n as shown in Fig. 2. To each ball (square) we assign one of $k(n - k + 1)$ colors. Let $\mathcal{C}_m^{(1)}(\mathcal{C}_m^{(2)})$ denote the set of balls (squares) assigned color $m, 1 \leq m \leq k(1 \leq m \leq n - k + 1)$, and let $c_m^{(1)} = |\mathcal{C}_m^{(1)}| (c_m^{(2)} = |\mathcal{C}_m^{(2)}|)$, the number of balls (squares) in $\mathcal{C}_m^{(1)}(\mathcal{C}_m^{(2)})$. The sets $\{\mathcal{C}_m^{(1)}\}(\{\mathcal{C}_m^{(2)}\})$ are sorted so that

$$c_1^{(1)} \leq c_2^{(1)} \leq \dots \leq c_k^{(1)} (c_1^{(2)} \leq c_2^{(2)} \leq \dots \leq c_{n-k+1}^{(2)}).$$

Given $(c_1^{(1)}, \dots, c_k^{(1)}, c_1^{(2)}, \dots, c_{n-k+1}^{(2)})$, find the number of colorations generated by the following process.

- 1) Assign colors to the balls according to a valid coloration defined in the preceding coloring problem.
- 2) Initially assign the same color to all squares.
- 3) Balls in set $\mathcal{C}_m^{(1)}$ divide the circle into $c_m^{(1)}$ segments. Assign one of $c_m^{(1)}$ colors to the squares within each segment, replacing the color previously assigned.

- 4) Repeat step 3 for $m = 2, \dots, k$. (Steps 1 and 2 correspond to $m = 1$.)

Two colorations are again counted the same if one is obtained from the other by exchanging colors among balls or squares.

Let

$$\bar{\mathcal{X}}(c_1^{(1)}, \dots, c_k^{(1)}, c_1^{(2)}, \dots, c_{n-k+1}^{(2)})$$

denote the solution to this problem. The following theorem expresses the large system moments in terms of $\bar{\mathcal{X}}$.

Theorem 6 (Large System Moments With Windowed Data):

The n th large system moment of \mathbf{WVPV}^\dagger for $n \geq 0$ can be evaluated as $\gamma_n^\infty = \sum_{k=1}^n \alpha^k \mathcal{H}_k^{(n)}$ where

$$\begin{aligned} \mathcal{H}_k^{(n)} &= \sum_{(c_1^{(1)}, \dots, c_k^{(1)}) \in \mathcal{S}_k^{(n)}} \sum_{(c_1^{(2)}, \dots, c_{n-k+1}^{(2)}) \in \mathcal{S}_{n-k+1}^{(n)}} \\ &\times \bar{\mathcal{X}}(c_1^{(1)}, \dots, c_k^{(1)}, c_1^{(2)}, \dots, c_{n-k+1}^{(2)}) \\ &\times \left(\prod_{t=1}^k E[P^{c_t^{(1)}}] \right) \left(\prod_{s=1}^{n-k+1} f_{c_s^{(2)}}(\bar{L}) \right) \end{aligned} \quad (79)$$

and

$$\mathcal{S}_k^{(n)} = \left\{ (c_1, c_2, \dots, c_k) \mid 1 \leq c_1 \leq c_2 \leq \dots \leq c_k, \sum_{t=1}^k c_t = n \right\}$$

The proof is given in Appendices I.A and B.

It is observed in [23] that computation of

$$\bar{\mathcal{X}}(c_1^{(1)}, \dots, c_k^{(1)}, c_1^{(2)}, \dots, c_{n-k+1}^{(2)})$$

is equivalent to counting factorizations of cycles in the symmetric group (see [40] and [41]). An explicit expression for this term is given in [42, Theorem 2.2]. Namely

$$\bar{\mathcal{X}}(c_1^{(1)}, \dots, c_k^{(1)}, c_1^{(2)}, \dots, c_{n-k+1}^{(2)})$$

is the number of noncrossing partitions π for which the blocks of π have sizes $c_1^{(1)}, \dots, c_k^{(1)}$ and the blocks of its Kreweras complement π^c have sizes $c_1^{(2)}, \dots, c_{n-k+1}^{(2)}$. (See [41] and [24] for a discussion of noncrossing partitions and Kreweras complement.) It is stated in [41, Theorem 2.1] that this combinatorial term is in fact the top connection coefficient for the corresponding symmetric group, for which an explicit expression is given by [42, Theorem 2.2]⁵

$$\begin{aligned} &\bar{\mathcal{X}}(c_1^{(1)}, \dots, c_k^{(1)}, c_1^{(2)}, \dots, c_{n-k+1}^{(2)}) \\ &= \frac{\mathcal{X}_k^{(n)}(c_1^{(1)}, \dots, c_k^{(1)}) \mathcal{X}_{n-k+1}^{(n)}(c_1^{(2)}, \dots, c_{n-k+1}^{(2)})}{\binom{n}{k-1} \binom{n-1}{k-1}}. \end{aligned} \quad (80)$$

We can therefore write

$$\gamma_n^\infty = \sum_{k=1}^n \alpha^k \frac{k}{\binom{n}{k-1} \binom{n-1}{k-1}} \Lambda_k^{(n)} \mathcal{F}_{n-k+1}^{(n)}(\bar{L}) \quad (81)$$

⁵We originally presented this formula as a conjecture [37].

where $\Lambda_k^{(n)}$ is defined by (75) and

$$\mathcal{F}_k^{(n)} = \sum_{(c_1, \dots, c_k) \in \mathcal{S}_k^{(n)}} \mathcal{X}_k^{(n)}(c_1, \dots, c_k) \prod_{t=1}^k f_{c_t}(\bar{L}). \quad (82)$$

Theorem 4 implies that without data windowing γ_n^∞ is a function of α and $E[P^k]$, $k = 1, \dots, n$, i.e.,

$$\gamma_n^\infty = \mathcal{G}_n(\alpha, E[P], \dots, E[P^n]) \quad (83)$$

With data windowing, Theorem 6 implies that γ_n^∞ is a function of α , $E[P^k]$, and $f_k(\bar{L})$, $k = 1, \dots, n$, i.e.,

$$\gamma_n^\infty = \mathcal{G}_{n,w}(\alpha, E[P], \dots, E[P^n], f_1(\bar{L}), \dots, f_n(\bar{L})) \quad (84)$$

where $f_n(\bar{L})$ is given by (17) with exponential weighting and effective window length \bar{L} .

Fig. 3 illustrates the convergence of the finite system moment, denoted by $\gamma_n^{(M)}$, to the large system limit. The mean and standard deviation of $\gamma_n^{(M)}$ divided by the large system moment are plotted versus M . The mean and standard deviation are computed by averaging over different i.i.d. matrices \mathbf{V} . The load $\alpha = 1/2$, and the power distribution is given by $F_P(x) = 0.8u(x-10) + 0.2u(x-100)$, where $u(x)$ is the unit step function. That is, there are two groups of users with received powers 10 and 20 dB, respectively, and the probability of being a high-power user is 0.2. Also, exponential weighting is included with normalized window length $\bar{L} = 2$. As M increases, the mean converges to the large system limit, and the standard deviation converges to zero.

Returning to the computation of the large system moments of the sample covariance matrix, the approach in Section IV is used in Appendix I-C to prove the following corollary.

Corollary 3: As $(K, N, i) \rightarrow \infty, \forall n \geq 1$

$$\begin{aligned} \hat{\gamma}_n(\bar{L}, \mu) &\rightarrow \hat{\gamma}_n^\infty(\bar{L}, \mu) \\ &= \mathcal{G}_n\{1, \eta \mathcal{G}_{1,w}(1, e_1, f_1(\bar{L})), \dots, \\ &\quad \eta \mathcal{G}_{n,w}(1, e_1, \dots, e_n, f_1(\bar{L}), \dots, f_n(\bar{L}))\} \\ &= \eta \mathcal{G}_{n,w} \left\{ 1, \frac{1}{\eta} \mathcal{G}_1(1, \eta e_1), \dots, \right. \\ &\quad \left. \frac{1}{\eta} \mathcal{G}_n(1, \eta e_1, \dots, \eta e_n), f_1(\bar{L}), \dots, f_n(\bar{L}) \right\} \end{aligned}$$

where $\mathcal{G}_{n,w}(\cdot)$ and $\mathcal{G}_n(\cdot)$ are defined by (84) and (83), respectively, convergence is in probability, $e_1 = (\alpha/\eta)E[P] + (1/\eta)\sigma^2$, and $e_k = (\alpha/\eta)E[P^k]$ for $k \geq 2$.

Without data windowing, the moments can be obtained from Corollary 3 by letting the effective window length $\bar{L} \rightarrow \infty$, in which case $f_n(\bar{L}) \rightarrow 1 \forall n$. The corresponding moment $\hat{\gamma}_n^\infty(\infty, \mu)$ can be computed according to the following corollary.

Corollary 4: As $(K, N, i) \rightarrow \infty, \forall n \geq 1$

$$\begin{aligned} \hat{\gamma}_n(\infty, \mu) &\rightarrow \hat{\gamma}_n^\infty(\infty, \mu) \\ &= \mathcal{G}_n\{1, \eta \mathcal{G}_1(1, e_1), \eta \mathcal{G}_2(1, e_1, e_2), \dots, \\ &\quad \eta \mathcal{G}_n(1, e_1, \dots, e_n)\} \end{aligned}$$

where convergence is in probability.

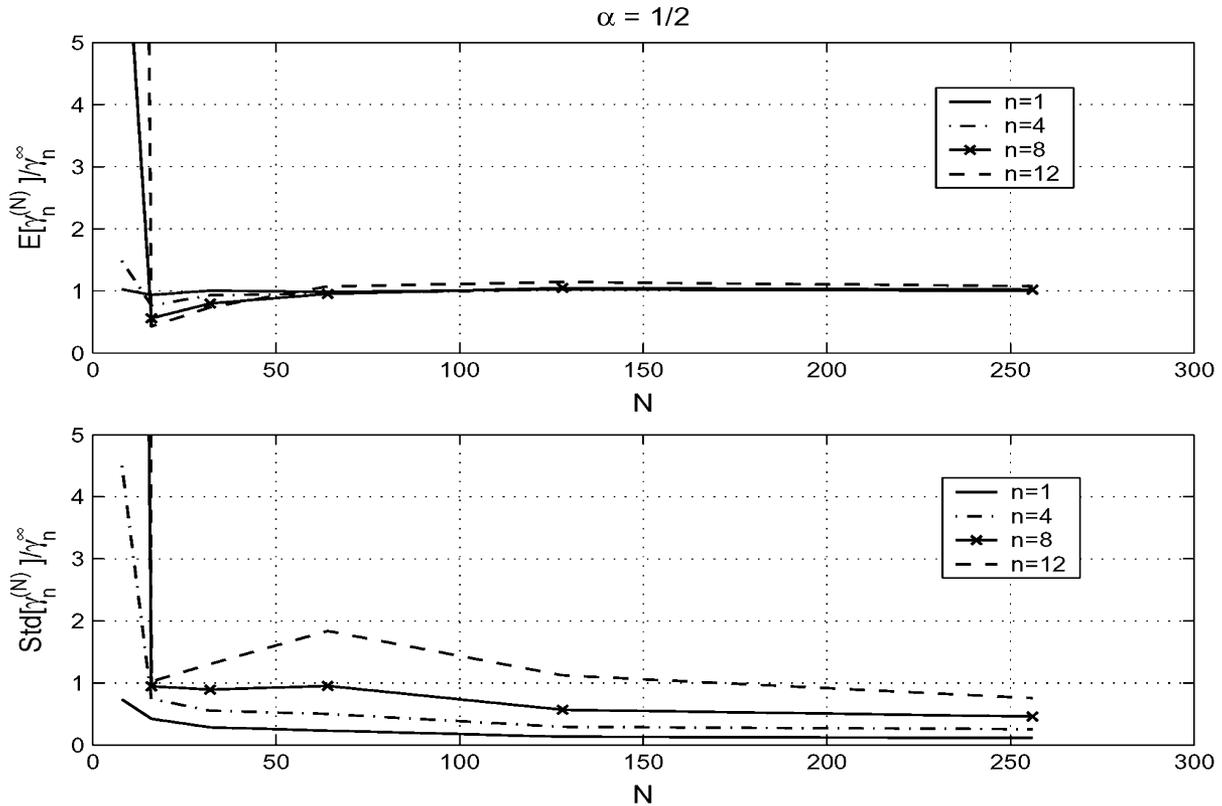


Fig. 3. Illustration of the convergence of $\gamma_n^{(M)}$ to γ_n^∞ with a nonuniform power distribution and exponential windowing.

As discussed in Section IV, the inner terms in Corollaries 3 and 4, such as $\eta\mathcal{G}_{n,w}(\cdot)$ or $1/\eta\mathcal{G}_n(\cdot)$, correspond to the moments of $\mathbf{\Omega}$ given in (37). The moments of $\mathbf{\Omega}$ serve as the moments of the “effective” power distribution, which are used to evaluate the moments of $\hat{\mathbf{R}}_T$, as stated by the corollaries.

VIII. NUMERICAL RESULTS

In this section, we present numerical results, which illustrate the transient performance of the adaptive LS and RLS filters considered. Simulation results for finite (K, N, i) are included for comparison with the large system limits. The latter results are averaged over random binary signatures and the received power distribution.

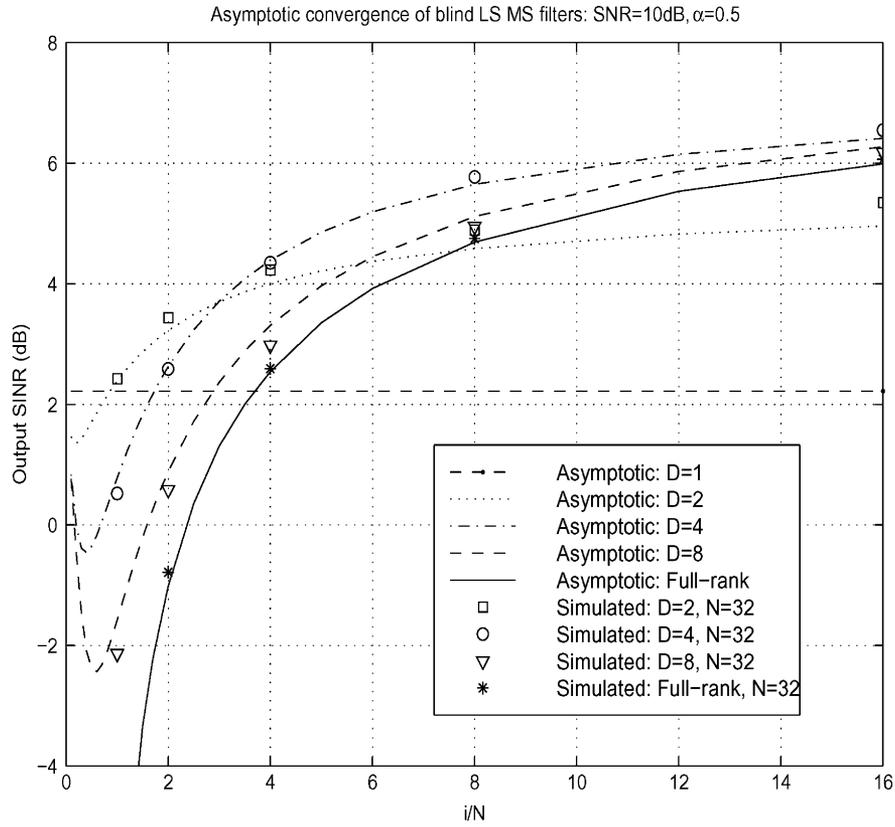
We first show results for LS filters without exponential weighting and with $\mu = 0$. Fig. 4 shows plots of output SINR versus normalized number of observations (i/N) for full- and reduced-rank LS filters with load $\alpha = 0.5$, background signal-to-noise ratio (SNR) = 10 dB, and assuming each user has the same received power. Fig. 4(a) shows results for blind LS filters, and Fig. 4(b) shows results for LS filters with training. Simulated values for $N = 32$ are shown as discrete points. The rank one filter with training estimates the matched filter (MF). These results show that the large system analysis accurately predicts the performance of the finite-size system for the cases considered. These results also show that the reduced-rank filter with appropriate rank converges significantly faster than the adaptive full-rank LS filter, both with and without training.

Fig. 4(a) shows that the optimal rank for the blind reduced-rank filter is a function of the number of observations. For $i < N$, rank $D = 1$ (the MF) is optimal, and for $N < i < 2N$, $D = 2$ is optimal. The optimal rank generally increases with the number of observations. Of course, with infinite observations ($\eta \rightarrow \infty$), the full-rank LS filter is optimal. It is shown in [19], [22] that this full-rank performance can be essentially achieved with a low-rank filter (i.e., $D = 8$ for a large range of loads and SNRs). Hence, we expect that an adaptive reduced-rank filter with $D = 8$ will converge faster than the full-rank LS filter. As shown in Corollary 1, as η increases from one to two, the output SINR for the full-rank blind LS filter increases from zero ($-\infty$ dB) to about 0 dB.

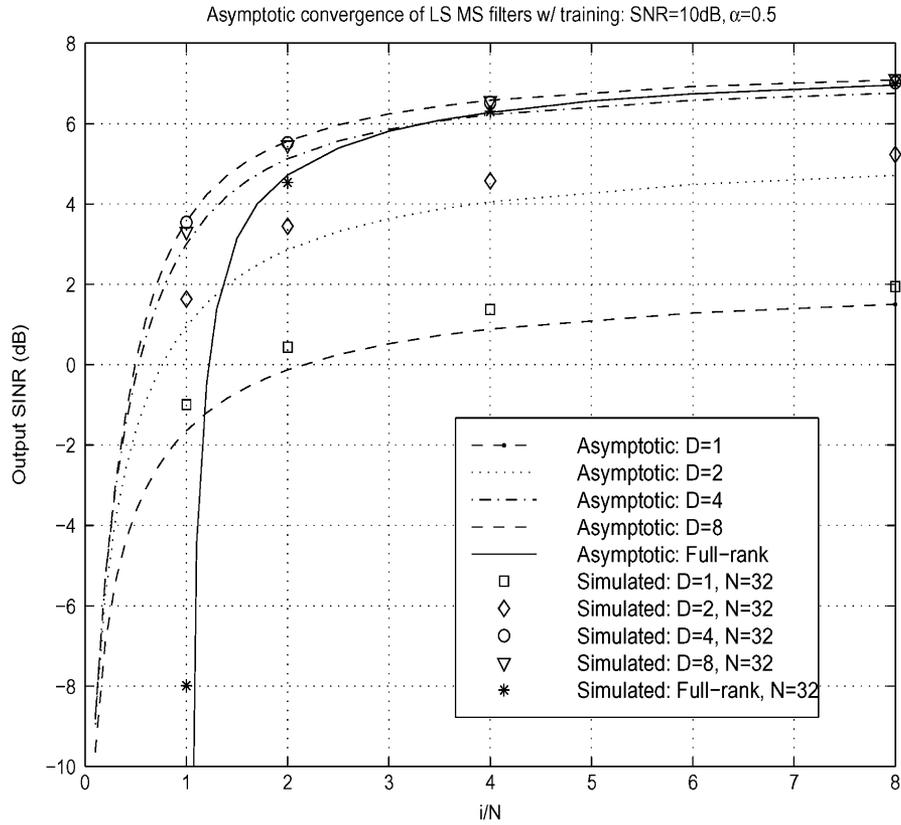
Fig. 4(a) shows that the output SINRs of the reduced-rank filters with $D > 1$ first decrease before increasing to the asymptotic SINR. This is because the blind adaptive reduced-rank filter is initialized as the MF. When the number of observations is small, the estimate of \mathbf{R} is inaccurate, which degrades the performance.

Fig. 4(b) shows that with training, $D = 8$ gives the best performance over the entire range of η . It also shows that the output SINR for the full-rank LS adaptive filter with training at $\eta = 2$ is 3 dB from MMSE performance, which is predicted by Corollary 1. In contrast, for the same target SINR, the reduced-rank filter with $D = 8$ requires approximately $1.3N$ iterations.

Fig. 5 shows the effect of rank on the performance of the training-based reduced-rank filter with different loads and SNRs. In these plots $\eta = 1.5$. With small loads, such as $\alpha = \frac{1}{16}$ and $\frac{1}{8}$, the optimal rank D is two or three.

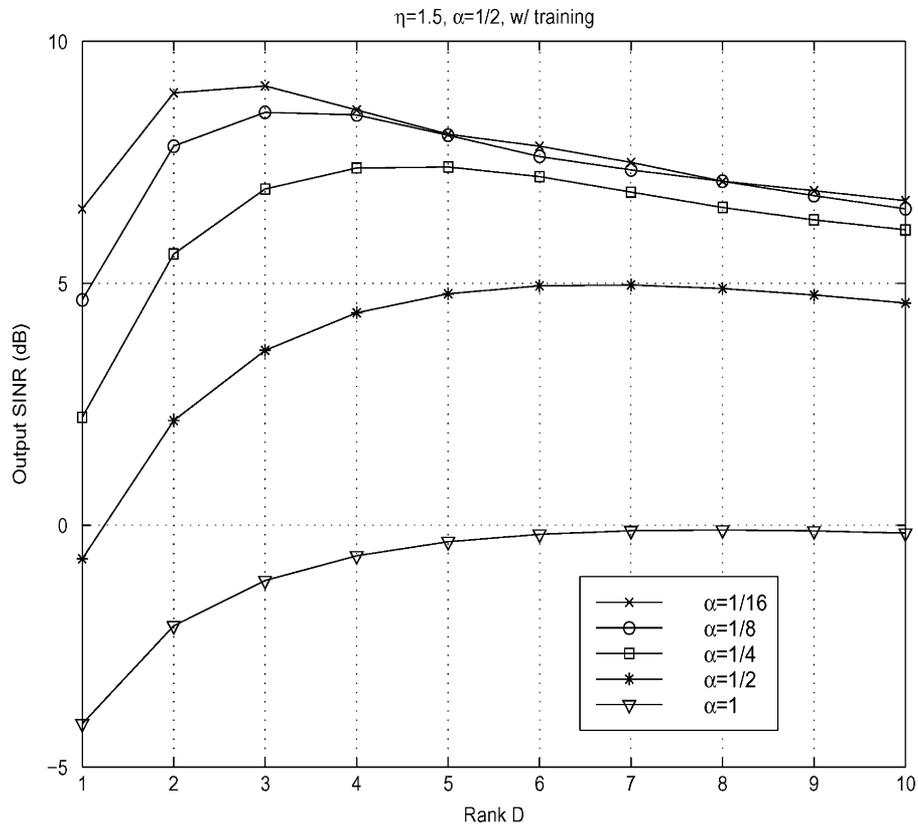


(a)

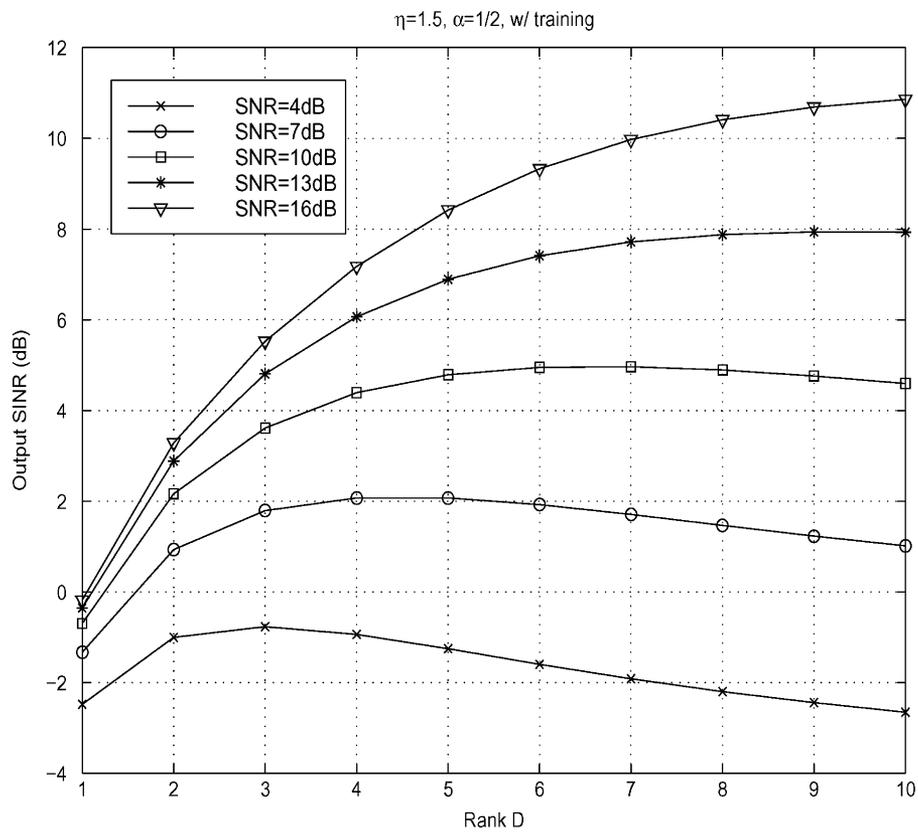


(b)

Fig. 4. Large system convergence plots for (a) blind adaptive LS filters and (b) adaptive LS filters with training.



(a)



(b)

Fig. 5. Output SINR versus rank for the reduced-rank filter with training and $\eta = 1.5$. Curves are shown for different SNRs in (a), and for different loads in (b).

Fig. 6 shows large system convergence plots for reduced- and full-rank RLS filters with exponential weighting and $\mu = 0.1$. Corresponding simulation results are also shown for $N = 64$. The system load is $\alpha = 0.5$, the SNR is 10 dB, the effective window length $\bar{L} = 10$, and all users are received with equal power. Large system analysis accurately predicts the simulation results. Comparing Fig. 6 with Fig. 4 shows that the convergence of the full-rank RLS filters is improved significantly by adding $\mu(i)\mathbf{I}$ to the sample covariance matrix. For this case, the blind reduced-rank filter with appropriate rank still converges faster than the full-rank filter. Since the effective window length is relatively long in this case, the exponential weighting does not have a significant effect. Fig. 6(b) shows that the full-rank RLS filter with $\mu = 0.1$ converges as fast as the reduced-rank RLS filter with optimized rank. Additional numerical results with nonuniform power distributions have shown similar behavior, with the exception that the performance of low-rank filters is more sensitive to power imbalances (e.g., caused by near-far effects).

In the next set of plots, shown in Figs. 7–9, we study the influence of effective window length and μ on the performance of RLS full-rank filters. Since the large system analysis accurately predicts the performance of finite systems, we only show the large system results.

Fig. 7 illustrates the effect of diagonal loading with uniform data windowing (no exponential weighting). Fig. 7(a) shows that as μ increases, the performance of the blind LS filter improves significantly for small η . For small values of μ , the SINR in each case is initially the SINR of the MF, and decreases before increasing to MMSE performance. This is due to the degradation caused by the inaccurate estimate of the sample covariance matrix, as discussed for the case $\mu = 0$. For large values of μ , the sample covariance remains close to $\mu(i)\mathbf{I}$ for small η , so that the output SINRs start at the SINR of the MF and increase monotonically.

The convergence curve for the reduced-rank blind RLS filter with $D = 8$ and $\mu = 0$ is also shown in Fig. 7(a) for comparison. The performance of the full-rank blind RLS filter with optimized μ is somewhat better than the performance of the analogous reduced-rank filter. Additional results generated for the blind reduced-rank filter with different values of μ show that diagonal loading $\mu > 0$ improves convergence, and there is generally an optimal value of μ . The performance of the reduced-rank blind RLS filter with $D = 8$ and optimized μ is nearly the same as that of the analogous full-rank filter. (Similar observations for the blind RLS filter, based on simulation, have been made in [11].)

Fig. 7(b) shows that diagonal loading also improves the convergence of the LS filter with training when η is small. In this case, large μ degrades the performance significantly for large η . When μ is small, there is a “notch” around $\eta = 1$, in which the SINR decreases to a local minimum. The performance for small η is a sensitive function of μ . The convergence curve for the reduced-rank RLS filter with training and $\mu = 0$ is also shown, and exhibits the best performance. Additional numerical results show that diagonal loading slightly *degrades* the performance of the reduced-rank RLS filter with training. That is, with small μ , the reduced-rank filter performs about the same as with $\mu = 0$, and the SINR with a fixed small η decreases as μ increases.

Fig. 8 shows convergence plots for RLS filters with training and exponential windowing. Curves are shown for different window lengths \bar{L} with $\mu = 0.1$. Exponential windowing effectively limits the training length, which prevents the adaptive filter from converging to the corresponding MMSE filter. As shown in Fig. 8, the steady-state SINR increases with \bar{L} . The output SINR starts to flatten out when $\eta \approx \bar{L}$ to $2\bar{L}$, which is consistent with the interpretation of \bar{L} as the effective window length. In addition, exponential weighting slightly improves the convergence of the blind RLS filter for small η .

Fig. 9 shows asymptotic output SINR, as $\eta \rightarrow \infty$, versus the effective window length \bar{L} for different values of μ . As expected, the asymptotic SINR increases with \bar{L} . Without exponential weighting, diagonal loading does not affect the asymptotic performance of the filters as $\eta \rightarrow \infty$. However, this is no longer true with exponential weighting. Fig. 9(a) shows that as μ increases from zero, the asymptotic SINR for the blind RLS filter increases. However, the performance degrades when μ is too large, so that there is an optimal value for μ . Performance is quite sensitive to the selection of μ when \bar{L} is small. In contrast, the asymptotic SINR with training is insensitive to changes in μ around $\mu = 0$, and $\mu = 0$ maximizes the asymptotic SINR. Additional results show that $\mu = 0$ also maximizes the asymptotic SINR for the reduced-rank RLS filter with training.

IX. CONCLUSION

Evaluating the average output SINR versus training samples for LS filters with random data is a well-known and difficult problem in adaptive filtering and estimation. The large system convergence results presented here appear to be the only available analytical results, which accurately predict performance for a wide range of system parameters and input statistics. Our model is general in the sense that it allows an arbitrary power distribution over the interfering data streams along with arbitrary data windowing. Numerical results have shown that the large system analysis is quite accurate for $N \geq 64$ for all cases considered. Additional numerical results indicate that this analysis is generally accurate for significantly smaller values of N , except for low-rank filters with data windowing.

This analysis shows that, as expected, the reduced-rank LS filters converge significantly faster than the full-rank LS filters without diagonal loading. In general, the optimal rank of the reduced-rank LS filter with finite training (or observations for blind LS with known user signature) depends on the load, background noise level, the distribution of received power across users, and the number of observations. Even so, the performance of the reduced-rank filter with training is relatively insensitive to a suboptimal choice of rank.

For the RLS filter, data windowing (e.g., exponential weighting) offers the potential of tracking a changing environment at the cost of degrading steady-state performance. Diagonal loading of the sample covariance matrix accelerates the convergence of the full-rank adaptive filter initially, but can adversely influence steady-state performance when used with exponential weighting. Still, a reduced-rank LS filter with optimized rank performs at least as well as the analogous full-

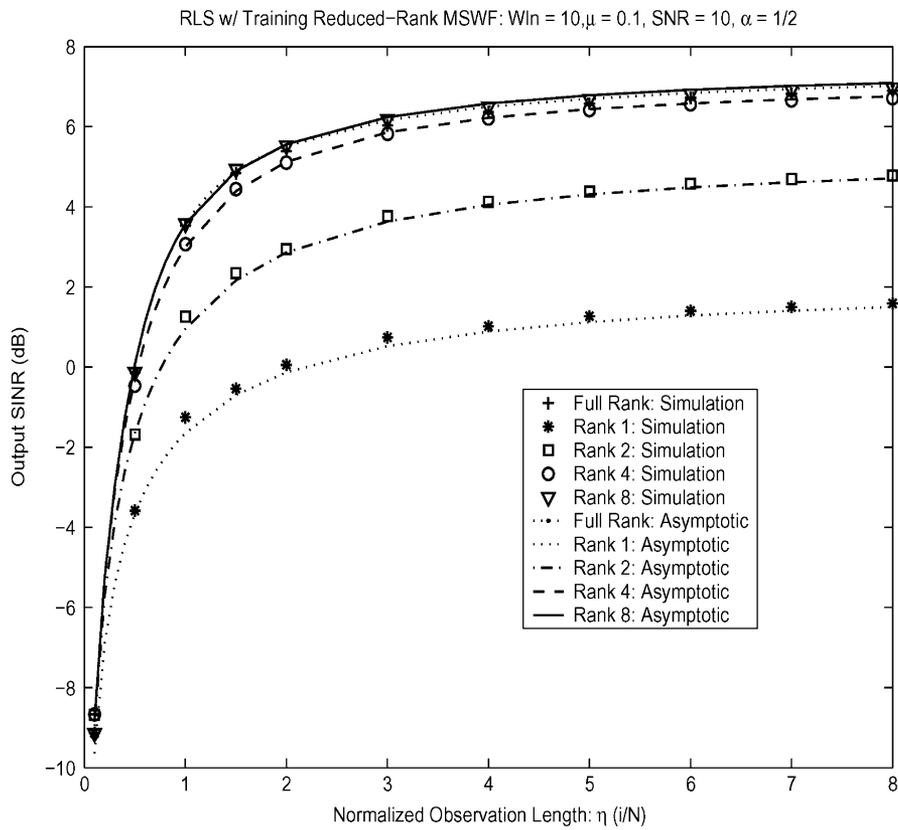
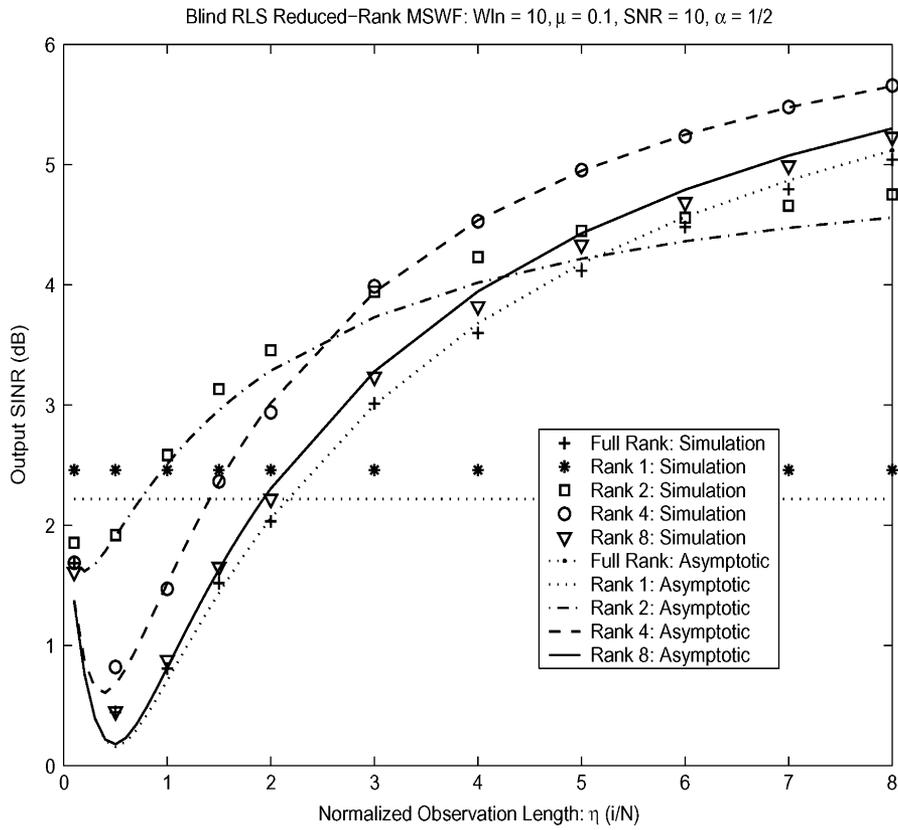


Fig. 6. Large system convergence plots with $\mu = 0.1$ and $\bar{L} = 10$ for (a) blind LS filters and (b) LS filters with training.

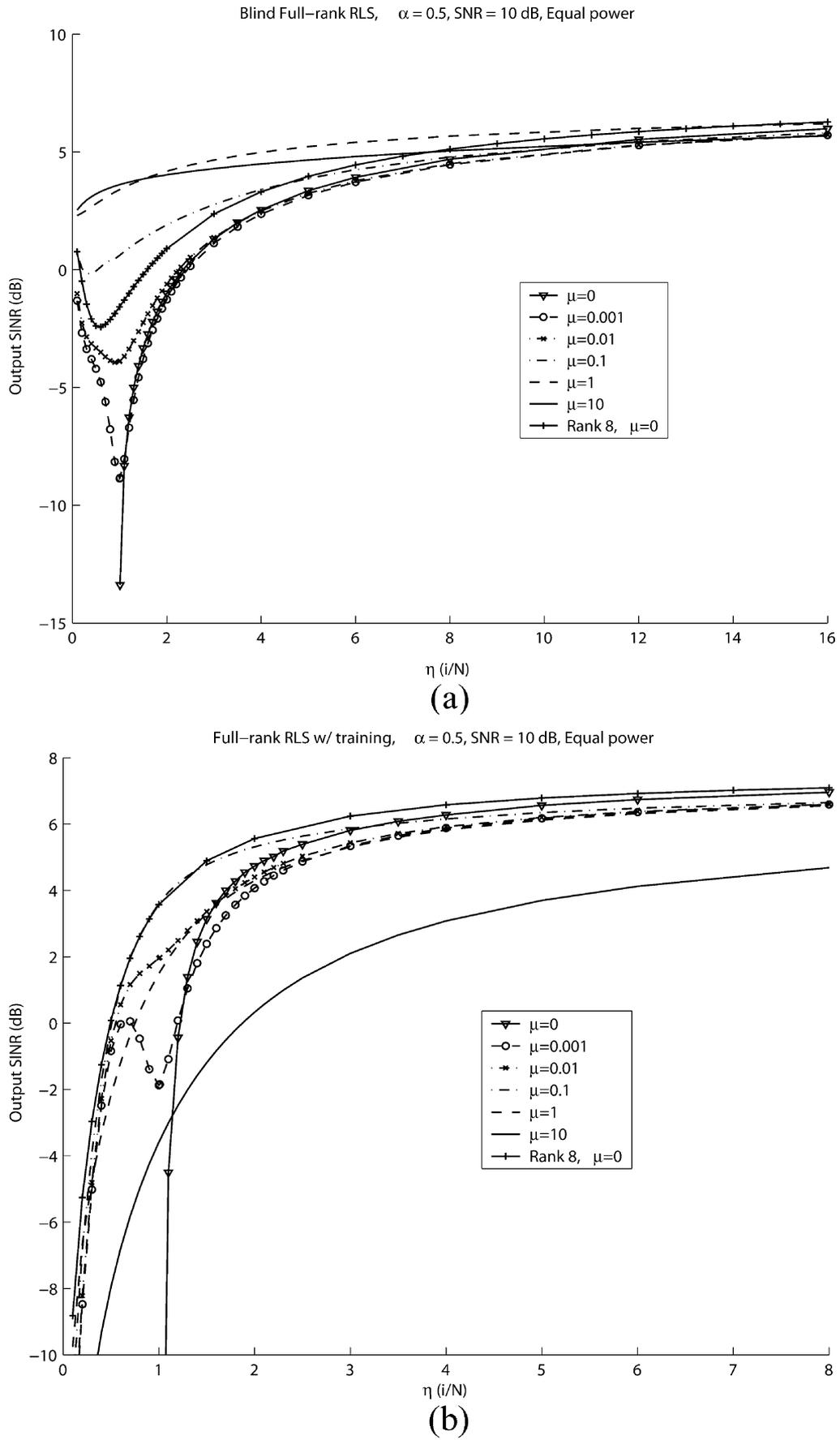
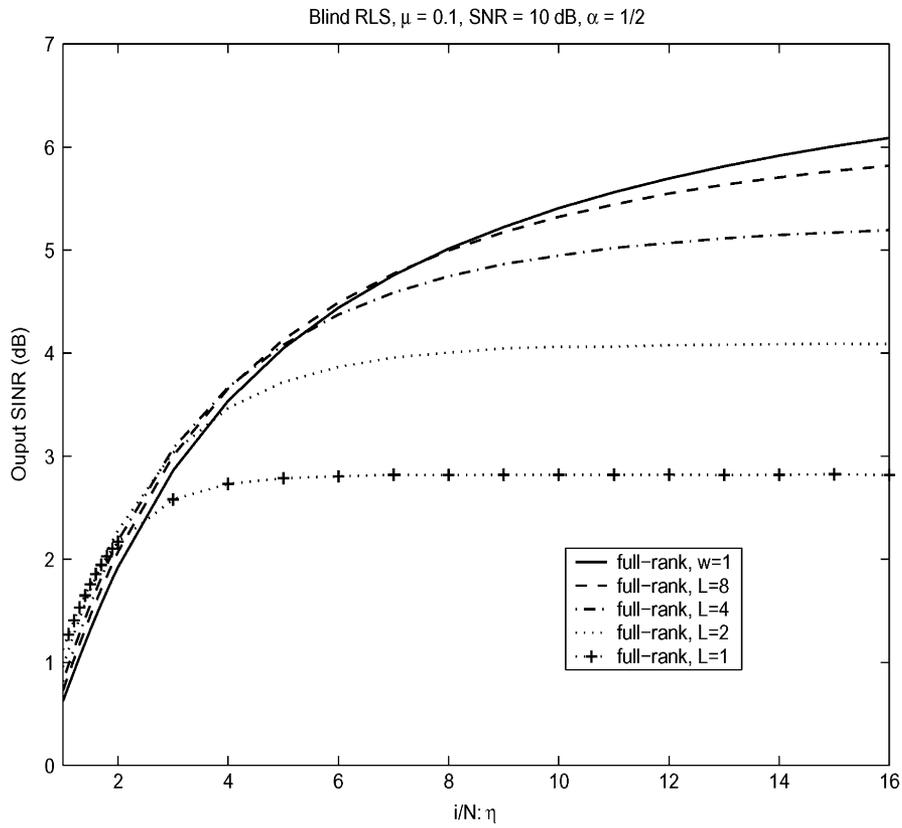
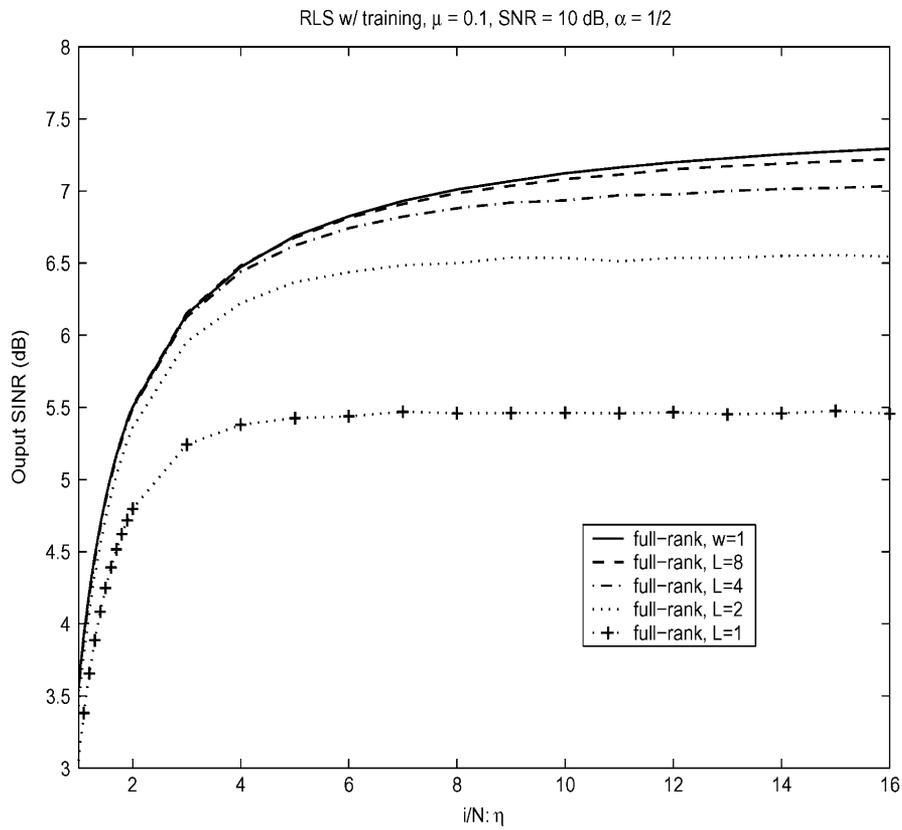


Fig. 7. Large system convergence plots for (a) blind RLS filters and (b) RLS filters with training with different values of μ .

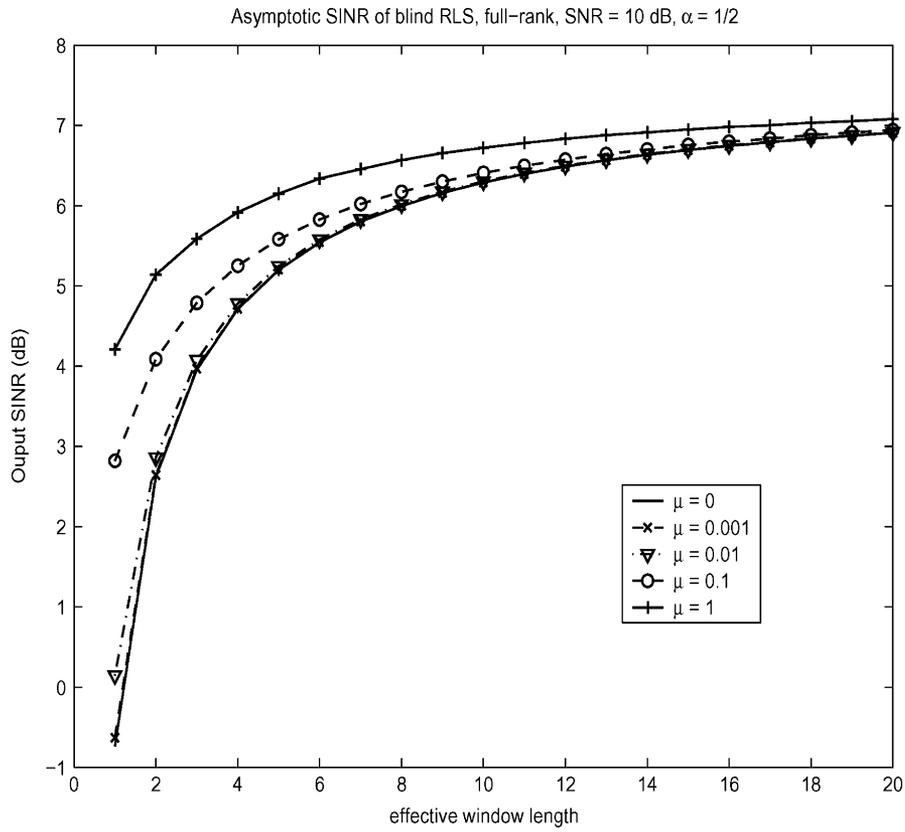


(a)

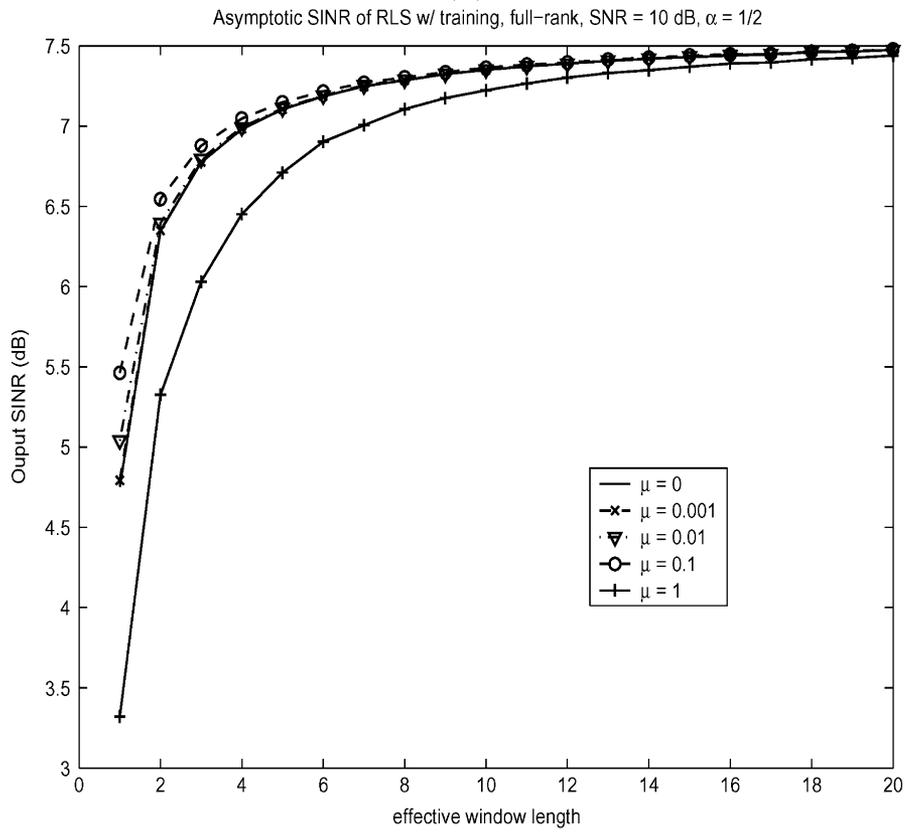


(b)

Fig. 8. Large system convergence plots for (a) blind RLS filters and (b) RLS filters with training with different values of \bar{L} (exponential windowing).



(a)



(b)

Fig. 9. Asymptotic ($\eta \rightarrow \infty$) performance of (a) the blind RLS filter and (b) the RLS filter with training versus effective window length \bar{L} .

rank filters, and is less sensitive to the selection of algorithm parameters (i.e., diagonal loading factor μ and exponential weight w).

This work can be extended in a few different directions. For example, more elaborate channel models can be considered (e.g., that include frequency selectivity). It may also be possible to obtain large system convergence results for other types of adaptive filtering algorithms, such as the LMS, or stochastic gradient algorithm [1], [2]. (Large system analyses of other reduced-rank, or subspace algorithms have appeared in [19], [20], [35].) Finally, whether or not the LS and RLS algorithms considered here are optimal, in the sense of maximizing the output SINR for a given training length, is an open question.

APPENDIX I CONVERGENCE AND COMPUTATION OF LARGE SYSTEM MOMENTS

We start by proving Theorem 5, which is used to prove the convergence of the moments of the sample covariance matrix (Theorem 1) in Appendix I-C. Theorem 4, which gives the large system moments without data windowing, follows from computations presented in the proof of Theorem 5. We then prove Theorem 6 in Appendix I-B.

A. Proof of Theorem 5: Moments of $\mathbf{WV}\mathbf{P}\mathbf{V}^\dagger$

We can expand $\gamma_n = (1/M)\text{trace}(\mathbf{WV}\mathbf{P}\mathbf{V}^\dagger)^n$ as

$$\gamma_n = \frac{1}{M} \sum_{i_1=1}^K \sum_{i_2=1}^K \cdots \sum_{i_n=1}^K P_{i_1} P_{i_2} \cdots P_{i_n} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \quad (85)$$

where

$$\zeta_n^{(M)}(i_1, i_2, \dots, i_n) = \left(\mathbf{v}_{i_1}^\dagger \mathbf{W} \mathbf{v}_{i_2} \right) \left(\mathbf{v}_{i_2}^\dagger \mathbf{W} \mathbf{v}_{i_3} \right) \cdots \left(\mathbf{v}_{i_{n-1}}^\dagger \mathbf{W} \mathbf{v}_{i_n} \right) \left(\mathbf{v}_{i_n}^\dagger \mathbf{W} \mathbf{v}_{i_1} \right) \quad (86)$$

P_i is the i th diagonal element of \mathbf{P} , and \mathbf{v}_i is the i th column of \mathbf{V} .

The method for evaluating the large system limit divides the set of indices

$$\mathcal{I}_n = \{(i_1 \cdots i_n) : 1 \leq i_j \leq K\}$$

in (85) into progressively smaller subsets. The large system moment can then be computed by partitioning the sum in (85) into smaller sums over the smaller subsets. The relation between (85) and the coloring problems in Section III is also defined in terms of these subsets.

We start by defining

$$\mathcal{I}_n^{(k)} = \{(i_1, \dots, i_n) : \exists \text{ exactly } k \text{ distinct indices}\}. \quad (87)$$

For $(i_1, \dots, i_n) \in \mathcal{I}_n^{(k)}$, let the set of distinct indices be denoted as $\mathcal{J} = \{i_{j_1}, \dots, i_{j_k}\}$. These indices are represented as different colors. Namely, let $\mathcal{A}_k = \{A_1, A_2, \dots, A_k\}$ denote a set of k distinct colors, and let $\mathcal{M} : \mathcal{J} \rightarrow \mathcal{A}_k$ be the one-to-one mapping of indices to colors. Note that $1 \leq k \leq K$, which corresponds to the range of indices. For an index vector $(i_1, \dots, i_n) \in \mathcal{I}_n^{(k)}$, we denote the corresponding *coloration* as

(a_1, a_2, \dots, a_n) where $a_j \in \mathcal{A}_k, j = 1, \dots, n$. Whereas A_k and $A_l, l \neq k$ are distinct colors, we may have $a_k = a_l$.

We further subdivide $\mathcal{I}_n^{(k)}$ into subsets in which the *multiplicity* of each distinct index (color) is specified. Let c_t be the number of indices, which satisfy $\mathcal{M}(i_j) = A_t$, where A_t denotes a specific color. In what follows, we will assume that $c_1 \leq c_2 \leq \dots \leq c_k$. We will refer to (c_1, \dots, c_k) as the *multiplicity vector*. For $(i_1, \dots, i_n) \in \mathcal{I}_n^{(k)}$, we define

$$\mathcal{I}_{(c_1, \dots, c_k)} = \{(i_1, \dots, i_n) : \text{the } k \text{ distinct indices have multiplicity vector } (c_1, \dots, c_k)\}. \quad (88)$$

Similarly, we define the set of colorations with multiplicity vector (c_1, \dots, c_k) as

$$\mathcal{C}_{(c_1, \dots, c_k)} = \{(a_1, a_2, \dots, a_n) : \exists \text{ exactly } c_t \text{ elements } a_j = A_t, t = 1, \dots, k\}. \quad (89)$$

The coloration $(\check{a}_1, \dots, \check{a}_n)$ is the same as (a_1, \dots, a_n) if there exists a permutation operation $Q(\cdot)$ on the set of colors (A_1, \dots, A_k) such that $(\check{a}_1, \dots, \check{a}_n) = (Q(a_1), \dots, Q(a_n))$.

For each $(i_1, \dots, i_n) \in \mathcal{I}_{(c_1, c_2, \dots, c_k)}$ we have

$$P_{i_1} P_{i_2} \cdots P_{i_n} = P_{j_1}^{c_1} P_{j_2}^{c_2} \cdots P_{j_k}^{c_k}$$

where j_1, \dots, j_k are the k distinct indices. Therefore, the summation in (85) over $(i_1, \dots, i_n) \in \mathcal{I}_{(c_1, c_2, \dots, c_k)}$ is $\prod_{t=1}^k E[P^{c_t}]$ times a scalar, where $E[P^m]$ is the m th moment of the power distribution.

Associated with each $(a_1, \dots, a_n) \in \mathcal{C}_{(c_1, \dots, c_k)}$ is a mapping $\mathcal{M} : \mathcal{J} \rightarrow \mathcal{A}_k$ where $\mathcal{M}(i_t) = A_t, t = 1, \dots, n$. We define a *color transformation mapping* $\mathcal{T} : \mathcal{A}'_k \rightarrow \mathcal{A}_k$, where \mathcal{A}'_k is a set of k distinct colors chosen from the set of all possible colors \mathcal{A}_K , and $\mathcal{A}'_k \neq \mathcal{A}_k$. For each $(a_1, a_2, \dots, a_n) \in \mathcal{C}_{(c_1, \dots, c_k)}$, we divide $\mathcal{C}_{(c_1, \dots, c_k)}$ into the following subsets:

$$\mathcal{C}_{(a_1, a_2, \dots, a_n)} = \{(a'_1, \dots, a'_n) : \exists \text{ a mapping } \mathcal{T} \text{ where } \mathcal{T}(a'_j) = a_j, j = 1, \dots, n\}. \quad (90)$$

That is, the colorations in $\mathcal{C}_{(a_1, a_2, \dots, a_n)}$ are obtained by applying different color transformation mappings to (a_1, a_2, \dots, a_n) . We define $\mathcal{I}_{(a_1, a_2, \dots, a_n)}$ as the corresponding set of index vectors, which subdivides $\mathcal{I}_{(c_1, \dots, c_k)}$. It is easy to show that for each $(a_1, \dots, a_n) \in \mathcal{C}_{(c_1, \dots, c_k)}$

$$|\mathcal{I}_{(a_1, \dots, a_n)}| = K(K-1) \cdots (K-k+1) = \frac{K!}{(K-k)!}. \quad (91)$$

Given a coloration $(a_1, \dots, a_n) \in \mathcal{C}_{(c_1, \dots, c_k)}$, we will need the following definition.

Segment of a Coloration: Assume that $a_{j_1} = a_{j_2} = \dots = a_{j_m} = A_m, 1 \leq m \leq k$. Then $(a_{j_t+1}, \dots, a_{j_{t+1}-1})$ is called the t -th segment of color m , denoted as $\text{Seg}_t^{(m)}$.

In this definition, the order $(a_{j_t+1}, \dots, a_{j_{t+1}-1})$ is circular. That is, if $j_{t+1} < j_t$, then a_n is followed by $(a_1, a_2, \dots, a_{j_{t+1}-1})$. Fig. 10 illustrates the definition of a segment. A segment is empty if $j_{t+1} = j_t + 1$, in which case a_{j_t} and $a_{j_{t+1}}$ are neighbors.

Recall that a *valid* coloration was defined by the first constraint in the coloring problem stated in Section VII. Colorations which do not satisfy this constraint are *invalid*. Here we redefine a valid coloration in terms of segments.

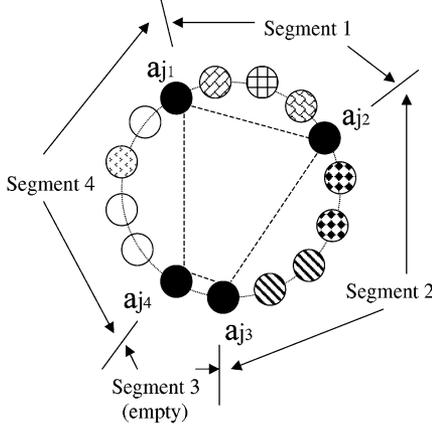


Fig. 10. Illustration of coloration segments.

Valid (Invalid) Coloration: A coloration

$$(a_1, \dots, a_n) \in \mathcal{C}_{(c_1, c_2, \dots, c_k)}$$

is valid if for all $1 \leq m \leq k$, $\text{Seg}_t^{(m)}$ and $\text{Seg}_s^{(m)}$ have no common color for $1 \leq t, s \leq c_m, t \neq s$. Otherwise, it is an invalid coloration.

For example, (A_1, A_1, A_2, A_2) and (A_1, A_2, A_2, A_1) are valid colorations in $\mathcal{C}_{(2,2)}$ whereas (A_1, A_2, A_1, A_2) is invalid. As defined in Section VII, for each $(c_1, \dots, c_k) \in \mathcal{S}_k^{(n)}$, $\mathcal{X}_k^{(n)}(c_1, \dots, c_k)$ denotes the number of valid colorations. Note that each segment of a valid coloration is itself a valid coloration.

We will show that summing over index vectors in (85) corresponding to invalid colorations in $\mathcal{I}_{(a_1, \dots, a_n)}$ gives zero in the large system limit, whereas the sum over index vectors corresponding to valid colorations converges to α^k times a scalar, which is the product of moments of the power distribution. The following two lemmas will be used to show that the sum in (85) converges in the mean-square sense.

Lemma 1: As $M \rightarrow \infty$, for each valid coloration $(a_1, \dots, a_n) \in \mathcal{C}_{(c_1, \dots, c_k)}$ we have

$$\frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)}} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \rightarrow \alpha^k F_{n-k+1}^{(n)}(\bar{L}) \quad (92)$$

where

$$F_{n-k+1}^{(n)}(\bar{L}) = \prod_{t=1}^{n-k+1} f_{c_t}(\bar{L}) \quad (93)$$

and $(c_1, \dots, c_{n-k+1}) \in \mathcal{S}_{n-k+1}^{(n)}$. For each invalid coloration of $\mathcal{C}_{(c_1, \dots, c_k)}$, denoted as (e_1, \dots, e_n) , we have

$$\frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(e_1, \dots, e_n)}} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \rightarrow 0 \quad (94)$$

where convergence is in the mean-square sense.

In the following lemma, $\mathcal{I}_{(A_1, A_2, \dots, A_k)}$ refers to the set of index vectors (i_1, \dots, i_k) , with distinct elements, i.e., $i_l \neq i_m, l \neq m$.

Lemma 2: For $(i_1, \dots, i_k) \in \mathcal{I}_{(A_1, A_2, \dots, A_k)}$, as $M \rightarrow \infty$

$$\frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}_{(A_1, \dots, A_k)}} P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k} \rightarrow \alpha^k \prod_{t=1}^k E[P^{c_t}] \quad (95)$$

The proofs are given in Appendix I-D.

We now write (85) as

$$\begin{aligned} \gamma_n &= \sum_{k=1}^n \sum_{(c_1, \dots, c_k) \in \mathcal{S}_k^{(n)}} \sum_{(a_1, \dots, a_n) \in \mathcal{C}_{(c_1, \dots, c_k)}} \frac{1}{M} \\ &\times \left\{ \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)}} \left(\prod_{t=1}^n P_{i_t} \right) \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \right\} \\ &+ \sum_{k=1}^n \sum_{(c_1, \dots, c_k) \in \mathcal{C}_k^{(n)}} \sum_{(e_1, \dots, e_n) \in \mathcal{C}_{(c_1, \dots, c_k)}} \frac{1}{M} \\ &\times \left\{ \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(e_1, \dots, e_n)}} \left(\prod_{t=1}^n P_{i_t} \right) \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \right\}. \end{aligned} \quad (96)$$

The first term on the right-hand side is a sum over index vectors corresponding to valid colorations, and the second term sums over invalid colorations. Focusing on the first term, as $(M, K) \rightarrow \infty$, we have

$$\frac{1}{M^k} \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)}} \prod_{t=1}^n P_{i_t} \rightarrow \alpha^k \prod_{t=1}^k E[P^{c_t}] \quad (97)$$

and from (92), we have

$$\begin{aligned} E \left[\frac{M^k}{K^k} \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)}} \zeta_n^{(M)}(i_1, \dots, i_n) \right] \\ = \frac{M^k K!}{K^k (K-k)!} \frac{1}{M} \\ \times E \left[\zeta_n^{(M)}(i_1, \dots, i_n) | (i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)} \right] \\ \rightarrow E \left[M^k \frac{1}{M} \zeta_n^{(M)}(i_1, \dots, i_n) | (i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)} \right] \\ \rightarrow F_{n-k+1}^{(n)} \end{aligned} \quad (98)$$

where the second step follows from (91), the third step follows from the fact that $(K!)/(K^k(K-k)!) \rightarrow 1$ as $K \rightarrow \infty$, and the last step follows from (92).

Since $\prod_{t=1}^n P_{i_t}$ and $\zeta_n^{(M)}(i_1, i_2, \dots, i_n)$ are independent random variables, we have

$$\begin{aligned} \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)}} \left(\prod_{t=1}^n P_{i_t} \right) \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \\ = \frac{1}{M^k} \sum_{(i_1, \dots, i_n) \in \mathcal{I}_{(a_1, \dots, a_n)}} \left(\prod_{t=1}^n P_{i_t} \right) \\ \times M^k \frac{1}{M} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \\ \rightarrow \alpha^k F_{n-k+1}^{(n)} \prod_{t=1}^k E[P^{c_t}]. \end{aligned} \quad (99)$$

A similar argument shows that the second sum in (96) over invalid colorations converges to zero. This establishes that the moments of \mathbf{WVPV}^\dagger converge.

To prove Theorem 4, we observe that there are $\mathcal{X}_k^{(n)}(c_1, \dots, c_k)$ valid colorations (a_1, \dots, a_n) for each multiplicity vector (c_1, \dots, c_k) . Substituting $\mathbf{W} = \mathbf{I}$, $f_k(\bar{L}) = 1$, and combining with (96) and (99) gives (74)–(76) in Theorem 4.

B. Proof of Theorem 6: Large System Moments With Windowed Data

From (96) and (99), it follows that γ_n^∞ can be written as

$$\gamma_n^\infty = \sum_{k=1}^n \alpha^k \mathcal{H}_k^{(n)} \quad (100)$$

where $\mathcal{H}_k^{(n)}$ has the form given in (79), and $\bar{\mathcal{X}}(\cdot)$ is a scalar function of $(c_1^{(1)}, \dots, c_k^{(1)})$ and $(c_1^{(2)}, \dots, c_{n-k+1}^{(2)})$. Referring to (85), the only terms in the sum, which make nonzero contributions to $\mathcal{H}_k^{(n)}$, correspond to valid colorations in $\mathcal{I}_{(a_1, \dots, a_n)}$, where $(a_1, \dots, a_n) \in \mathcal{C}_{(c_1^{(1)}, \dots, c_k^{(1)})}$. Here the set of valid colorations refers to configurations of balls in the double-coloring problem in Section VII.

The equivalence of the double-coloring problem with the evaluation of $\bar{\mathcal{X}}(\cdot)$ follows directly from the segmentation procedure used in Appendix I-D to prove Lemma 1. Namely, it is shown in the proof of the first part of Lemma 1 in Appendix I-D that we can segment a valid coloration into several smaller valid colorations, where the contribution each valid coloration makes to the sum in (92) has the form $\alpha^{k_t} F_{n_t - k_t + 1}^{(n_t)}(\bar{L})$, where $F_{n_t - k_t + 1}^{(n_t)}(\bar{L})$ is given by (93). Here t is the index for the smaller (sub)colorations, n_t is the number of balls in the t th subcoloration, and k_t is the number of distinct colors. The contribution each valid subcoloration makes to the sum (92) is then the product of the terms $\alpha^{k_t} F_{n_t - k_t + 1}^{(n_t)}$ corresponding to the different segments, as shown in (113).

The squares in the double-coloring problem correspond to the positions of the data windowing matrices (\mathbf{W}), shown in the expansion (85)–(86). For each valid double coloration, the term $f_{c_s}(\bar{L})$ (i.e., the c_s th moment of the data windowing matrix \mathbf{W}), which appears in the sum (79), corresponds to exactly c_s squares having color s . The segmentation procedure in Appendix I-D assigns one of the $c_k^{(1)}$ distinct colors to the squares within each segment. This corresponds to step 3 in the double-coloring problem. Since n and k are finite, the segmentation of valid colorations of balls (or equivalently, index vectors in (113)) accomplishes the coloration of squares in a finite number of steps in exactly the same way as stated in the double-coloring problem. This completes the proof of Theorem 6.

C. Proof of Theorem 1: Convergence of $\hat{\gamma}_n(\bar{L}, \mu)$

From (38), we only need to show that $\hat{\gamma}_n(\bar{L}, 0)$ converges. Let

$$\tilde{\mathbf{\Omega}} = \frac{1}{i} \begin{bmatrix} \mathbf{B}^\dagger & \mathbf{N}_B^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{A}^2 & 0 \\ 0 & \tilde{\sigma}^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B} \\ \mathbf{N}_B \end{bmatrix} \mathbf{W} \quad (101)$$

where \mathbf{B} is $K \times i$, \mathbf{N}_B is $L \times i$, and \mathbf{A}^2 is the diagonal matrix of user powers. The load associated with $\tilde{\mathbf{\Omega}}$, defined in Appendix I-B, is $\tilde{\alpha} = (K + L/i)$, and the n th moment of the effective power distribution is

$$E[\check{P}^n] = (KE[P^n] + L((N/L\sigma^2)^n K + L).$$

From Theorem 6 and (84), the n th large system moment for the matrix \mathbf{WVPV}^\dagger is given by

$$\begin{aligned} \gamma_n^\infty &= \mathcal{G}_{n,w}(\alpha, E[P], \dots, E[P^n], f_1(\bar{L}), \dots, f_n(\bar{L})) \\ &= \mathcal{G}_{n,w}(1, \alpha E[P], \dots, \alpha E[P^n], f_1(\bar{L}), \dots, f_n(\bar{L})) \end{aligned} \quad (102)$$

and with $\mathbf{W} = \mathbf{I}$

$$\begin{aligned} \gamma_n^\infty &= \mathcal{G}_n(\alpha, E[P], \dots, E[P^n]) \\ &= \mathcal{G}_n(1, \alpha E[P], \dots, \alpha E[P^n]). \end{aligned} \quad (103)$$

For the matrix $\tilde{\mathbf{\Omega}}$, we have

$$\begin{aligned} \tilde{\alpha} E[\check{P}^n] &= \frac{KE[P^n] + L \left(\frac{N}{L}\sigma^2\right)^n}{i} \\ &= \frac{\alpha}{\eta} E[P^n] + \frac{L}{i} \left(\frac{N}{L}\sigma^2\right)^n \\ &= \frac{\alpha}{\eta} E[P^n] + \frac{\nu}{\eta} \left(\frac{\sigma^2}{\nu}\right)^n \end{aligned}$$

where $\nu = (L/N)$. From (101) and (102), the n th large system moment of $\tilde{\mathbf{\Omega}}$ is

$$\gamma_n^\infty(\tilde{\mathbf{\Omega}}) = \mathcal{G}_{n,w}(1, \tilde{\alpha} E[\check{P}], \dots, \tilde{\alpha} E[\check{P}^n], f_1(\bar{L}), \dots, f_n(\bar{L})) \quad (104)$$

Now $\text{trace}(\mathbf{\Omega}^n) = \text{trace}(\tilde{\mathbf{\Omega}}^n)$ where $\mathbf{\Omega}$ is defined in (36). Hence, $\gamma_n(\mathbf{\Omega}) = (i)/(K + L)\gamma_n(\tilde{\mathbf{\Omega}})$. Let $\alpha = (K + L)/N$ be the effective load for $\hat{\mathbf{R}}_I(i)$. Then

$$\begin{aligned} \alpha \gamma_n^\infty(\mathbf{\Omega}) &= \eta \gamma_n^\infty(\tilde{\mathbf{\Omega}}) \\ &= \eta \mathcal{G}_{n,w}(1, \tilde{\alpha} E[\check{P}], \dots, \tilde{\alpha} E[\check{P}^n], \\ &\quad f_1(\bar{L}), \dots, f_n(\bar{L})). \end{aligned} \quad (105)$$

According to the discussion in Section IV, $\hat{\mathbf{R}}_I(i)$ has an effective power distribution with n th moment $\gamma_n(\mathbf{\Omega})$. From (102), and the discussion following (34) and (35), we have

$$\hat{\gamma}_n^\infty = \mathcal{G}_n[1, \eta \gamma_1^\infty(\mathbf{\Omega}), \dots, \eta \gamma_n^\infty(\mathbf{\Omega})]. \quad (106)$$

For each finite ν , $\hat{\gamma}_n$ therefore converges, and $\hat{\gamma}_n$ is a continuous function of ν . According to the discussion in Section IV, we must let $\nu = \frac{L}{N} \rightarrow \infty$ to obtain i.i.d. Gaussian noise, so that

$$\begin{aligned} \tilde{\alpha} E[\check{P}] &\rightarrow \frac{KE[P] + N\sigma^2}{i} \\ \tilde{\alpha} E[\check{P}^n] &\rightarrow \frac{K}{i} E[P^n], \quad \text{for } n > 1. \end{aligned} \quad (107)$$

Combining with (106) gives

$$\begin{aligned} \hat{\gamma}_n^\infty &= \mathcal{G}_n[1, \eta \mathcal{G}_{1,w}(1, e_1, f_1(\bar{L})), \dots, \\ &\quad \eta \mathcal{G}_{n,w}(1, e_1, \dots, e_n, f_1(\bar{L}), \dots, f_n(\bar{L}))] \end{aligned} \quad (108)$$

where, from (107), $e_1 = (\alpha/\eta)E[P] + (1/\eta)\sigma^2$ and $e_k = (\alpha/\eta)E[P^k]$ for $k \geq 2$. This gives Theorem 1 as well as the first equality in Corollary 3.

Let

$$\Theta = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma} \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{S}_I^\dagger \\ \mathbf{N}_S^\dagger \end{bmatrix} [\mathbf{S}_I \quad \mathbf{N}_S] \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma} \mathbf{I} \end{bmatrix}$$

so that

$$\hat{\gamma}_n(\bar{L}, \mu) = \frac{1}{N} \text{trace}[(\mathbf{B}_I^\dagger \quad \mathbf{N}_B^\dagger) \Theta (\mathbf{B}_I^\dagger \quad \mathbf{N}_B^\dagger)^\dagger \mathbf{W}(i)^n].$$

Following a similar argument as before then gives the second equality in Corollary 3.

D. Proof of Lemma 1

To prove the first part of Lemma 1, we first consider the case $k = n$, which corresponds to $c_1 = c_2 = \dots = c_n = 1$. From (86), and the fact that $E[\mathbf{v}_i \mathbf{v}_i^\dagger] = M^{-1} \mathbf{I}$, it follows that

$$\begin{aligned} E \left[\frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(A_1, \dots, A_n)} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \right] \\ &= \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(A_1, \dots, A_n)} E[\zeta_n^{(M)}(i_1, i_2, \dots, i_n)] \\ &= \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(A_n, \dots, A_n)} \text{trace}(M^{-n} \mathbf{W}^n) \\ &= \sum_{(i_1, \dots, i_n) \in \mathcal{I}(A_1, \dots, A_n)} M^{-n-1} \text{trace}(\mathbf{W}^n) \\ &\xrightarrow{M \rightarrow \infty} \alpha^n f_n(\bar{L}). \end{aligned} \quad (109)$$

Furthermore, the variance

$$\text{Var} \left[\frac{1}{M} \sum_{(i_1, \dots, i_n)} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \right] \rightarrow 0$$

as $M \rightarrow \infty$. This can be shown using an argument in [30]. (See [30, eqs. (2.14) to (2.16)].)

If $k < n$, then we have $c_k > 1$. For any valid coloration (a_1, \dots, a_n) , we may have $a_{j_1} = a_{j_2} = \dots = a_{j_{c_k}}$ as shown in Fig. 10. These points, j_1, \dots, j_{c_k} , divide the circle into c_k segments. We denote the t th segment as

$$\text{Seg}_t = (a_{j_t+1}, \dots, a_{j_{t+1}-1}).$$

If $j_{t+1} = j_t + 1$, then Seg_t is empty. According to the definition of a valid coloration, if $a_i \in \text{Seg}_t$ and $a_j \in \text{Seg}_u$ where $t \neq u$, then $a_i \neq a_j$. Hence, we have

$$\begin{aligned} \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(a_1, \dots, a_n)} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \\ &= \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(a_1, \dots, a_n)} \prod_{m=1}^{c_k} \left[\left(\sum_{k=1}^M w_k v_{k i_{j_m}}^* v_{k i_{(j_m+1)}} \right) \dots \right. \\ &\quad \left. \left(\sum_{k=1}^M w_k v_{k i_{(j_{m+1}-1)}} v_{k i_{j_{m+1}}} \right) \right] \\ &= \frac{1}{M} \sum_{(i_{j_1} = \dots = i_{j_{c_k}})} \prod_{m=1}^{c_k} \left[\sum_{(i_{j_{m+1}}, \dots, i_{j_{m+1}-1}) \in \mathcal{I}(a_{j_{m+1}}, \dots, a_{j_{m+1}-1})} \right. \\ &\quad \left. \times \zeta_{(i_{j_{m+1}} - i_{j_m})}^{(M)}(i_{j_m}, \dots, i_{j_{m+1}-1}) \right]. \end{aligned} \quad (110)$$

Now we consider different cases for (c_1, \dots, c_{k-1}) .

Case 1: $c_1 = \dots = c_{k-1} = 1$: This corresponds to the situation where the A_i 's are distinct, except for $A_{j_1}, \dots, A_{j_{c_k}}$. According to (109), we have

$$\begin{aligned} \sum_{(i_{j_{n+1}}, \dots, i_{j_{n+1}-1}) \in \mathcal{I}(1, \dots, 1)} \zeta_{(i_{j_{n+1}} - i_{j_n})}^{(M)}(i_{j_n}, \dots, i_{j_{n+1}-1}) \\ \rightarrow \alpha^{i_{j_{n+1}} - i_{j_n} - 1} f_{i_{j_{n+1}} - i_{j_n}}(\bar{L}) \end{aligned} \quad (111)$$

Hence, (110) becomes

$$\begin{aligned} \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(A_1, \dots, A_n)} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \\ \rightarrow \alpha \times \alpha^{(i_{j_2} - i_{j_1} - 1)} f_{i_{j_2} - i_{j_1}}(\bar{L}) \times \dots \\ \times \alpha^{(i_{j_1} - i_{j_{c_k}} + n - 1)} f_{i_{j_1} - i_{c_k}}(\bar{L}) \\ = \alpha^{n - c_k + 1} f_{i_{j_2} - i_{j_1}}(\bar{L}) \dots f_{i_{j_1} - i_{c_k}}(\bar{L}) \\ = \alpha^k F_{n-k+1}^{(n)}(\bar{L}) \end{aligned} \quad (112)$$

where the sequence of indices of f in the first and second steps is $\{i_{j_2} - i_{j_1}, i_{j_3} - i_{j_2}, \dots, i_{j_1} - i_{c_k}\}$, and all negative subscripts are modulo n , since the subscripts are cyclically ordered. Also,

$$\sum_{t=1}^{c_k} (i_{j_{t+1}} - i_{j_t}) = n$$

and $c_k = n - k + 1$. This establishes the first part of Lemma 1 for Case 1.

Case 2: $c_{k-1} > 1$: In this case, the valid coloration constraint implies that there is no common element in different segments $\text{Seg}_t = (a_{j_t+1}, \dots, a_{j_{t+1}-1})$ and $\text{Seg}_u, t \neq u$. Each segment is also a valid coloration in the coloration set $\mathcal{C}_{(c_1^{(t)}, \dots, c_{k_t}^{(t)})}^{(t)}$, which implies that

$$(c_1^{(1)}, \dots, c_{k_1}^{(1)}, \dots, c_1^{(c_k)}, \dots, c_{k_{c_k}}^{(c_k)}, c_k)$$

is a permutation of (c_1, \dots, c_k) , and some segments can be empty, e.g., if A_{j_t} and $A_{j_{t+1}}$ are neighbors. Therefore, if $c_{k_t}^{(t)} > 1$, the same argument used for Case 1 can be applied to $(c_1^{(t)}, \dots, c_{k_t}^{(t)})$.

Since n and k are finite, by considering analogous cases for (c_1, \dots, c_{k-2}) , we can show in a finite number of steps that

$$\begin{aligned} \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(A_1, \dots, A_n)} \zeta_n^{(M)}(i_1, i_2, \dots, i_n) \\ \rightarrow \alpha \times \alpha^{k_1 + \dots + k_{c_k}} F_{n_1 - k_{n+1}}^{(n_1)} \dots F_{n_{c_k} - k_{c_k} + 1}^{(n_{c_k})} \\ = \alpha^k F_{n-k+1}^{(n)}(\bar{L}) \end{aligned} \quad (113)$$

This establishes the first part of Lemma 3.

For the second part of Lemma 1, consider the invalid coloration $(e_1, \dots, e_n) \in \mathcal{C}_{(c_1, \dots, c_k)}$ shown in Fig. 11. That is, $e_{t_1} = e_{t_2} = A_t$ and $e_{s_1} = e_{s_2} = A_s$ where $t \neq s$, and the

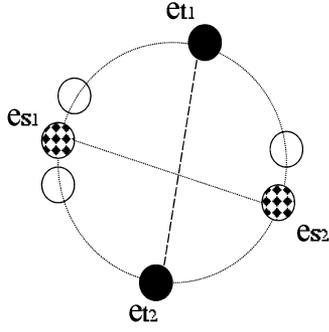


Fig. 11. Illustration of an invalid coloring.

line $e_{t_1}e_{t_2}$ crosses the line $e_{s_1}e_{s_2}$. The sum in (110) is then replaced by the sum

$$\begin{aligned} & \frac{1}{M} \sum_{(i_1, \dots, i_n) \in \mathcal{I}(e_1, \dots, e_n)} \tilde{\zeta}_n^{(M)}(i_1, i_2, \dots, i_n) \\ &= \frac{1}{M} \sum_{(i_{j_1} = \dots = i_{j_{c_k}})} \prod_{m=1}^{c_k} \left[\sum_{(i_{j_{m+1}}, \dots, i_{j_{m+1}-1}) \in \mathcal{I}(e_{j_{m+1}}, \dots, e_{j_{m+1}-1})} \right. \\ & \quad \left. \times \zeta_{(i_{j_{m+1}} - i_{j_m})}^{(M)}(i_{j_m}, \dots, i_{j_{m+1}-1}) \right] \end{aligned} \quad (114)$$

where the terms $\tilde{\zeta}_n^{(M)}$ differ from the analogous terms on the left-hand side of (110) in that there are at least two terms $\zeta_{(i_{j_t} - i_{j_t})}^{(M)}(i_{j_{t+1}}, \dots, i_{j_{t+1}-1})$ on the right-hand side of (114) which have at least two common indices as arguments. These common indices, say, i_p and i_q , where $i_p = i_q$, correspond to balls $e_p = e_q = A_p$ in an invalid coloring, as illustrated in Fig. 11. When summing over (i_1, \dots, i_n) , both of these indices range from 1 to K . This effectively reduces the dimension of the space of possible index vectors (i_1, \dots, i_n) by one, which causes the corresponding sum over the set of constrained index vectors to converge to zero.

We consider the following cases.

Case 1: Each segment $\text{Seg}_{(e_{j_t+1}, \dots, e_{j_t+1}-1)}$ is a valid coloring and

$$(e_{j_t+1}, \dots, e_{j_t+1}-1) \in \mathcal{C}_{(c_1^{(t)}, \dots, c_{k_t}^{(t)})}^{n_t}$$

where $n_t = j_{t+1} - j_t$, and k_t is the number of distinct colors. For each of these colorations, we have

$$\begin{aligned} & \sum_{(i_{j_t+1}, \dots, i_{j_t+1}-1) \in \mathcal{I}(e_{j_t+1}, \dots, e_{j_t+1}-1)} \tilde{\zeta}_{(i_{j_t+1} - i_{j_t})}^{(M)}(i_{j_t}, \dots, i_{j_t+1}-1) \\ & \rightarrow \alpha^{k_t-1} F_{k_t}^{(n_t)}. \end{aligned}$$

Assume that there are $z > 1$ segments $\text{Seg}_{t_m}, m = 1, \dots, z$, and that each of these segments has at least one $e_{p_m}, m =$

$1, \dots, z$, which has color A_0 . The corresponding indices in the sum (114) satisfy $i_{p_1} = i_{p_2} = \dots = i_{p_z} = i_p$, so that

$$\begin{aligned} & \sum_{(i_{j_t+1}, \dots, i_{j_t+1}-1) \in \mathcal{I}(e_{j_t+1}, \dots, e_{j_t+1}-1)} \tilde{\zeta}_{(i_{j_t+1} - i_{j_t})}^{(M)}(i_{j_t}, \dots, i_{j_t+1}-1) \\ &= \sum_{i_p} \sum_{i_p \notin (i_{j_t+1}, \dots, i_{j_t+1}-1)} \tilde{\zeta}_{(i_{j_t+1} - i_{j_t})}^{(M)}(i_{j_t}, \dots, i_{j_t+1}-1) \\ &= \sum_{i_p} X(i_{j_t}, i_p) \\ & \rightarrow \alpha^{k_t-1} F_{k_t}^{(n_t)} \end{aligned}$$

where

$$X(i_{j_t}, i_p) = \sum_{i_p \notin (i_{j_t+1}, \dots, i_{j_t+1}-1)} \tilde{\zeta}_{(i_{j_t+1} - i_{j_t})}^{(M)}(i_{j_t}, \dots, i_{j_t+1}-1)$$

and is $O(1/M)$. Since $z > 1$, the sum in (114) becomes $\frac{1}{M} \sum_{(i_{j_t}, i_p)} (X(i_{j_t}, i_p))^z$, which is

$$O\left(M^{-1} \times M^2 \times \frac{1}{M^z}\right) = O\left(\frac{1}{M^{z-1}}\right).$$

This establishes the second part of Lemma 1 for this case.

Case 2: The segments $(e_{j_t+1}, \dots, e_{j_t+1}-1)$ may not all be valid colorations. The same argument used for Case 1 can be applied to these invalid colorations, which gives

$$\begin{aligned} & \sum_{(i_{j_t+1}, \dots, i_{j_t+1}-1) \in \mathcal{I}(i_{j_t+1}, \dots, i_{j_t+1}-1)} \tilde{\zeta}_{(i_{j_t+1} - i_{j_t})}^{(M)}(i_{j_t}, \dots, i_{j_t+1}-1) \\ & \rightarrow 0 \end{aligned}$$

Therefore, Lemma 1 is also valid for this case.

E. Proof of Lemma 2

For $1 \leq i_p, i_q \leq K, i_p \neq i_q, \forall p \neq q, 1 \leq p, q \leq k$

$$\begin{aligned} & E \left[\frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(A_1, \dots, A_k)} P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k} \right] \\ &= \frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(A_1, \dots, A_k)} E [P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k}] \\ &= \frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(A_1, \dots, A_k)} \prod_{t=1}^k E [P^{c_t}] \\ & \quad \times \xrightarrow{M \rightarrow \infty} \alpha^k \prod_{t=1}^k E [P^{c_t}]. \end{aligned} \quad (115)$$

The second moment can be computed as

$$\begin{aligned} & E \left[\frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(A_1, \dots, A_k)} P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k} \right]^2 \\ &= \frac{1}{M^{2k}} \sum_{(i_1, \dots, i_k)} \sum_{(l_1, \dots, l_k)} E [P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k} P_{l_1}^{c_1} P_{l_2}^{c_2} \dots P_{l_k}^{c_k}] \\ &= \frac{1}{M^2 k} \sum_{(i_1, \dots, i_k, l_1, \dots, l_k) \in \mathcal{I}(A_1, \dots, A_{2k})} E [P_{i_1}^{c_1} \dots P_{i_k}^{c_k} P_{l_1}^{c_1} \dots P_{l_k}^{c_k}] \\ & \quad + \frac{1}{M^2 k} \sum_{(i_1, \dots, i_k, l_1, \dots, l_k) \notin \mathcal{I}(A_1, \dots, A_{2k})} E [P_{i_1}^{c_1} \dots P_{i_k}^{c_k} P_{l_1}^{c_1} \dots P_{l_k}^{c_k}]. \end{aligned}$$

Following the same sequence as in (115), the first term, which sums over those $(i_1, \dots, i_k, l_1, \dots, l_k)$ with distinct elements, converges to

$$\alpha^{2k} \prod_{t=1}^k (E[P_{c_t}])^2.$$

If $i_p = l_q$, then the number of corresponding vectors $(i_1, \dots, i_k, l_1, \dots, l_k)$ grows as $o(M^{2k})$. Therefore, as $M \rightarrow \infty$, the second term goes to zero, and we have

$$\begin{aligned} & \text{Var} \left[\frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(A_1, \dots, A_k)} P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k} \right] \\ &= E \left[\left(\frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(A_1, \dots, A_k)} P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k} \right)^2 \right] \\ & \quad - \left[E \left(\frac{1}{M^k} \sum_{(i_1, \dots, i_k) \in \mathcal{I}(A_1, \dots, A_k)} P_{i_1}^{c_1} P_{i_2}^{c_2} \dots P_{i_k}^{c_k} \right) \right]^2 \\ & \rightarrow \alpha^{2k} \prod_{t=1}^k (E[P_{c_t}])^2 - \left(\alpha^k \prod_{t=1}^k E[P_{c_t}] \right)^2 \\ &= 0. \end{aligned}$$

This proves Lemma 2.

APPENDIX II

PROOFS OF THEOREMS 2 AND 3: LARGE SYSTEM SINR FOR RLS FILTERS

It will be convenient to use the following notation. Two vectors \mathbf{v}_1 and \mathbf{v}_2 are *asymptotically equivalent* if

$$\lim_{(K, N, i) \rightarrow \infty} E(\|\mathbf{v}_1 - \mathbf{v}_2\|^2) = 0$$

in which case we will write $\mathbf{v}_1 \asymp \mathbf{v}_2$.

A. Theorem 2: Full-Rank SINR

To prove Theorem 2, we use the following asymptotic forms for the LS filter \mathbf{c}_1 , which are derived in Appendix II-D. Namely

$$\mathbf{c}_1 \asymp \frac{1}{1 + P_1 \hat{\gamma}_{-1}^\infty (f_1(\bar{L}) - \hat{\psi}_{-1}^\infty)} \left(a_1 \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 + a_2 \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \right)$$

where $\hat{\psi}_{-1}^\infty$, a_1 , and a_2 are given in Theorem 2

$$\mathbf{q}_1 = \frac{1}{i} \mathbf{J} \mathbf{W} \mathbf{b}_1^\dagger \quad (116)$$

where $\mathbf{b}_1 = [b_1(1) \dots b_1(i)]$ and

$$\mathbf{J} = \mathbf{S}_I \mathbf{A}_I \mathbf{B}_I(i) + \mathbf{N}(i) \quad (117)$$

is the $N \times i$ received interference-plus-noise matrix. The output SINR does not depend on the scaling of \mathbf{c}_1 . Hence, for both LS filters (with or without training) we assume that

$$\mathbf{c}_1 = \kappa_1 \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 + \kappa_2 \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \quad (118)$$

where κ_1 and κ_2 are constants. This is exact in the large system limit. To simplify the notation, we drop the time index i .

We also need the following limits, which are derived in Appendix II-C

$$\mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^m \mathbf{q}_1 \asymp 0 \quad (119)$$

$$\mathbf{q}_1^\dagger \hat{\mathbf{R}}_I^{m-1} \mathbf{q}_1 \asymp \hat{\psi}_{m-1}^\infty \quad (120)$$

for $m = 0, -1$.

Computing the large system SINR for the \mathbf{c}_1 in (118) gives

$$\begin{aligned} \beta &= P_1 a_1^2 (\hat{\gamma}_{-1})^2 \times \left\{ \sigma^2 a_1^2 \hat{\gamma}_{-2} + \sigma^2 a_2^2 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_I^{-2} \mathbf{q}_1 \right. \\ & \quad + a_1^2 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 \\ & \quad \left. + a_2^2 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \right\}^{-1}. \quad (121) \end{aligned}$$

Now $\hat{\gamma}_{-1}$ and $\hat{\gamma}_{-2}$ converge to $\hat{\gamma}_{-1}^\infty$ and $\hat{\gamma}_{-2}^\infty$, respectively, and

$$\mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 = \sum_{k=2}^K P_k \mathbf{s}_k^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_k. \quad (122)$$

Let $\hat{\mathbf{R}}_{I;(1,k)}$ denote the sample interference-plus-noise covariance matrix without the terms corresponding to \mathbf{s}_1 and \mathbf{s}_k , i.e.,

$$\hat{\mathbf{R}}_{I;(1,k)} = \hat{\mathbf{R}}_I - A_k \mathbf{q}_k \mathbf{s}_k^\dagger - A_k \mathbf{s}_k \mathbf{q}_k^\dagger - P_k \hat{f}_1(\bar{L}) \mathbf{s}_k \mathbf{s}_k^\dagger. \quad (123)$$

In analogy with the form for \mathbf{c}_1 in (118), we can write

$$\hat{\mathbf{R}}_I^{-1} \mathbf{s}_k = \kappa'_1 \hat{\mathbf{R}}_{I;(1,k)}^{-1} \mathbf{s}_k + \kappa'_2 \hat{\mathbf{R}}_{I;(1,k)}^{-1} \mathbf{q}_k$$

where κ'_1 and κ'_2 are constants. Substituting in (122) gives

$$\begin{aligned} & \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 \\ & \asymp \sum_{k=2}^K \frac{P_k}{\left[1 + P_k \hat{\gamma}_{-1}^\infty (f_1(\bar{L}) - \hat{\psi}_{-1}^\infty) \right]^2} \\ & \quad \times \left[\mathbf{s}_k^\dagger \hat{\mathbf{R}}_{I;(1,k)}^{-1} \mathbf{s}_1 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_{I;(1,k)}^{-1} \mathbf{s}_k \right. \\ & \quad \left. + P_k (\hat{\gamma}_{-1}^\infty)^2 \mathbf{q}_k^\dagger \hat{\mathbf{R}}_{I;(1,k)}^{-1} \mathbf{s}_1 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_{I;(1,k)}^{-1} \mathbf{q}_k \right] \quad (124) \end{aligned}$$

where we have used (119) and the fact that $\mathbf{s}_k^\dagger \hat{\mathbf{R}}_{I;(1,k)}^{-1} \mathbf{s}_1 \asymp 0$. Evaluating the large system limit gives

$$\begin{aligned} & \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 \\ & \asymp \frac{1}{N} \sum_{k=2}^K \frac{P_k \hat{\gamma}_{-2}^\infty + P_k^2 (\hat{\gamma}_{-1}^\infty)^2 \hat{\psi}_{-2}^\infty}{\left(1 + P_k \hat{\gamma}_{-1}^\infty [f_1(\bar{L}) - \hat{\psi}_{-1}^\infty] \right)^2} \\ & \rightarrow \alpha E_P \left[\frac{P \hat{\gamma}_{-2}^\infty + P^2 (\hat{\gamma}_{-1}^\infty)^2 \hat{\psi}_{-2}^\infty}{\left(1 + P \hat{\gamma}_{-1}^\infty [f_1(\bar{L}) - \hat{\psi}_{-1}^\infty] \right)^2} \right] \quad (125) \end{aligned}$$

where we have used (120).

Finally, we again apply the preceding argument, and the argument in Appendix II-C to evaluate

$$\begin{aligned}
& \mathbf{q}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \\
& \asymp \frac{1}{i^2} \mathbf{b}_1 \mathbf{W} \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{J} \mathbf{W} \mathbf{b}_1^\dagger \\
& = \text{trace} \left[\frac{1}{i^2} \mathbf{W} \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{J} \mathbf{W} \right] \\
& \asymp \frac{\mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\mathcal{W}_{1,2}(\phi_{-1}^\infty)} \text{trace} \left[\frac{1}{i} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-1} - \frac{\mu}{i} \mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_I^{-2} \right] \\
& \asymp \frac{\mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\mathcal{W}_{1,2}(\phi_{-1}^\infty)} \frac{1}{i} \sum_{k=2}^K \left[P_k \mathbf{s}_k^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_k - \mu P_k \mathbf{s}_k^\dagger \hat{\mathbf{R}}_I^{-2} \mathbf{s}_k \right] \\
& \asymp \frac{\mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\mathcal{W}_{1,2}(\phi_{-1}^\infty)} \frac{1}{i} \sum_{k=2}^K \left[\frac{P_k \hat{\gamma}_{-1}^\infty}{1 + P_k \hat{\gamma}_{-1}^\infty [f_1(\bar{L}) - \hat{\psi}_{-1}^\infty]} \right. \\
& \quad \left. - \mu \frac{P_k \hat{\gamma}_{-2}^\infty + P_k^2 (\hat{\gamma}_{-1}^\infty)^2 \hat{\psi}_{-2}^\infty}{\left(1 + P_k \hat{\gamma}_{-1}^\infty [f_1(\bar{L}) - \hat{\psi}_{-1}^\infty]\right)^2} \right] \\
& \rightarrow \frac{\mathcal{W}_{2,2}(\phi_{-1}^\infty)}{\mathcal{W}_{1,2}(\phi_{-1}^\infty)} \frac{\alpha}{\eta} E_P \left[\frac{P \hat{\gamma}_{-1}^\infty}{1 + P \hat{\gamma}_{-1}^\infty [f_1(\bar{L}) - \hat{\psi}_{-1}^\infty]} \right. \\
& \quad \left. - \mu \frac{P \hat{\gamma}_{-2}^\infty + P^2 (\hat{\gamma}_{-1}^\infty)^2 \hat{\psi}_{-2}^\infty}{\left(1 + P \hat{\gamma}_{-1}^\infty [f_1(\bar{L}) - \hat{\psi}_{-1}^\infty]\right)^2} \right]. \tag{126}
\end{aligned}$$

Combining (121)–(126) gives Theorem 2.

B. Corollary 2: Asymptotic Performance of RLS Filters

We need to evaluate (21) as $\eta \rightarrow \infty$. First, we observe that

$$\hat{\gamma}_k^\infty(\hat{\mathbf{R}}_I) = \lim_{(K,N,i) \rightarrow \infty} \text{trace}(\hat{\mathbf{R}}_I^k) = \eta^{-k} \hat{\gamma}_k^\infty(\eta \hat{\mathbf{R}}_I). \tag{127}$$

Now Corollary 3 implies that

$$\begin{aligned}
& \hat{\gamma}_n^\infty(\eta \hat{\mathbf{R}}_I) \\
& = \eta^{n+1} \mathcal{G}_{n,w} \left\{ 1, \frac{1}{\eta} \mathcal{G}_1(1, \eta e_1), \dots, \frac{1}{\eta} \mathcal{G}_n(1, \eta e_1, \dots, \eta e_n), \right. \\
& \quad \left. f_1(\bar{L}), \dots, f_n(\bar{L}) \right\} \\
& = \mathcal{G}_{n,w} \left\{ 1, \mathcal{G}_1(1, \tilde{e}_1), \dots, \mathcal{G}_n(1, \tilde{e}_1, \dots, \tilde{e}_n), \bar{L}, \frac{\bar{L}}{2}, \dots, \frac{\bar{L}}{n} \right\}
\end{aligned}$$

where $\tilde{e}_k = \eta e_k$, $\tilde{e}_1 = \alpha E[P] + \sigma^2$, and $\tilde{e}_k = \alpha E[P^k]$ for $k \geq 2$. This shows that $\hat{\gamma}_n^\infty(\eta \hat{\mathbf{R}}_I)$ is a continuous and bounded function of the \tilde{e} 's and \bar{L} , and converges to a finite constant as $\eta \rightarrow \infty$. Hence, from (127), as $\eta \rightarrow \infty$, $\hat{\gamma}_k^\infty(\hat{\mathbf{R}}_I)$ is $O(\eta^{-k})$.

We also observe that $f_k(\bar{L}) = (\bar{L}/k\eta)(1 - e^{-(k\eta/\bar{L})})$ is $O(\eta^{-1})$, and from (24) and (25), it is easy to show that $\hat{\psi}_{-1}^\infty$ is $O(\eta^{-1})$ and $\hat{\psi}_{-2}^\infty$ is $O(\eta^{-2})$. It is then straightforward to derive Corollary 2 from (21).

C. Derivation of (119) and (120)

From (116) and (117), we have $\mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^m \mathbf{q}_1 = 1/i \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^m \mathbf{J} \mathbf{W} \mathbf{b}_1^\dagger$. Note that $\hat{\mathbf{R}}_I = (1/i) \mathbf{J} \mathbf{W} \mathbf{J}^\dagger + (\mu/\eta) \mathbf{I}$, and \mathbf{s}_1 , \mathbf{q}_1 , and \mathbf{J} are independent and zero mean, so that

$$E(\mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^m \mathbf{q}_1) \rightarrow 0. \tag{128}$$

The variance is given by

$$\begin{aligned}
& E(\mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^m \mathbf{q}_1 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_I^m \mathbf{s}_1) \\
& = \frac{1}{i^2} \text{trace} \left\{ E(\mathbf{s}_1 \mathbf{s}_1^\dagger) E \left[\hat{\mathbf{R}}_I^m \mathbf{J} \mathbf{W} E(\mathbf{b}_1^\dagger \mathbf{b}_1) \mathbf{W} \mathbf{J}^\dagger \hat{\mathbf{R}}_I^m \right] \right\} \\
& = \frac{1}{i^2 N} \text{trace} \left(E(\hat{\mathbf{R}}_I^m \mathbf{J} \mathbf{W}^2 \mathbf{J}^\dagger \hat{\mathbf{R}}_I^m) \right) \\
& \rightarrow 0
\end{aligned}$$

where the last step follows from the fact that

$$\frac{1}{iN} \text{trace} E[\hat{\mathbf{R}}_I^m \mathbf{J} \mathbf{W}^2 \mathbf{J}^\dagger \hat{\mathbf{R}}_I^m]$$

converges to a finite limit. This establishes convergence in the mean square sense.

To prove (120), we first observe that

$$\mathbf{q}_1^\dagger \hat{\mathbf{R}}_I^m \mathbf{q}_1 = \frac{1}{i} \text{trace} \left(\frac{1}{i} \mathbf{W}^2 \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-m} \mathbf{J} \right). \tag{129}$$

Let $\hat{\mathbf{R}}_{It} = \hat{\mathbf{R}}_I - \frac{1}{i} w_t \mathbf{J}_t \mathbf{J}_t^\dagger$, where \mathbf{J}_t is the t th column of \mathbf{J} . We then have⁶

$$\begin{aligned}
& \frac{1}{i} \text{trace} \left(\frac{1}{i} \mathbf{W}^k \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-m} \mathbf{J} \right) = \frac{1}{i} \sum_{t=1}^i \frac{\frac{1}{i} w_t^k \mathbf{J}_t^\dagger \hat{\mathbf{R}}_{It}^{-m} \mathbf{b}_{Jt}}{\left(1 + \frac{1}{i} w_t \mathbf{J}_t^\dagger \hat{\mathbf{R}}_{It}^{-1} \mathbf{J}_t\right)^m} \\
& \asymp \phi_{-m}^\infty \frac{1}{i} \sum_{t=1}^i \frac{w_t^k}{\left(1 + w_t \phi_{-1}^\infty\right)^m} \\
& \asymp \phi_{-m}^\infty \frac{1}{\eta} \tag{130}
\end{aligned}$$

$\mathcal{W}_{k,m}(\phi_{-1}^\infty)$ where

$$\phi_{-m}^\infty = \lim_{(K,N,i) \rightarrow \infty} \frac{1}{i} \mathbf{J}_t^\dagger \hat{\mathbf{R}}_{It}^{-m} \mathbf{J}_t, \quad m = 1, 2, \mathcal{W}_{k,m}(x)$$

is defined by (18), and

$$\begin{aligned}
& \frac{1}{i} \mathbf{J}_t^\dagger \hat{\mathbf{R}}_{It}^{-1} \mathbf{J}_t = \frac{1}{i} (\mathbf{S}_I \mathbf{A}_I \mathbf{b}_{It} + \mathbf{N}_t)^\dagger \hat{\mathbf{R}}_{It}^{-1} (\mathbf{S}_I \mathbf{A}_I \mathbf{b}_{It} + \mathbf{N}_t) \\
& \asymp \frac{1}{i} \mathbf{b}_{It}^\dagger \mathbf{A}_I^\dagger \mathbf{S}_I^\dagger \hat{\mathbf{R}}_{It}^{-1} \mathbf{S}_I \mathbf{A}_I \mathbf{b}_{It} + \frac{1}{i} \mathbf{N}_t^\dagger \hat{\mathbf{R}}_{It}^{-1} \mathbf{N}_t \\
& \asymp \frac{1}{i} \text{trace}[\mathbf{S}_I \mathbf{P}_I \mathbf{S}_I^\dagger \hat{\mathbf{R}}_{It}^{-1}] + \frac{\sigma^2}{\eta N} \text{trace}[\hat{\mathbf{R}}_{It}^{-1}]. \tag{131}
\end{aligned}$$

⁶The following calculation was provided by M. Peacock, who pointed out that our original proof of Theorem 2 had incorrectly assumed that the matrices \mathbf{W}^k and $\mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-m} \mathbf{J}$ are asymptotically free.

Applying the same manipulations as in (126) gives ϕ_{-1}^∞ in (26). Moreover

$$\begin{aligned} \frac{1}{i^2} \text{trace} \left[\mathbf{W}^2 \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-m} \mathbf{J} \right] \\ \asymp \frac{\mathcal{W}_{2,m}(\phi_{-1}^\infty)}{\mathcal{W}_{1,m}(\phi_{-1}^\infty)} \frac{1}{i^2} \text{trace} \left[\mathbf{W} \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-m} \mathbf{J} \right] \end{aligned} \quad (132)$$

and substituting $\frac{1}{i} \mathbf{J} \mathbf{W} \mathbf{J}^\dagger = \hat{\mathbf{R}}_I - (\mu/\eta) \mathbf{I}$, gives

$$\begin{aligned} \frac{1}{i^2} \text{trace} \left[\mathbf{W} \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{J} \right] &= \frac{1}{i} \text{trace} \left[\mathbf{I} - \frac{\mu}{\eta} \hat{\mathbf{R}}_I^{-1} \right] \\ &\rightarrow \frac{1}{\eta} - \frac{\mu}{\eta^2} \hat{\gamma}_{-1}^\infty \end{aligned} \quad (133)$$

$$\begin{aligned} \frac{1}{i^2} \text{trace} \left[\mathbf{W} \mathbf{J}^\dagger \hat{\mathbf{R}}_I^{-2} \mathbf{J} \right] &= \frac{1}{i} \text{trace} \left[\hat{\mathbf{R}}_I^{-1} - \frac{\mu}{\eta} \hat{\mathbf{R}}_I^{-2} \right] \\ &\rightarrow \frac{1}{\eta} \hat{\gamma}_{-1}^\infty - \frac{\mu}{\eta^2} \hat{\gamma}_{-2}^\infty. \end{aligned} \quad (134)$$

Combining with (132) gives the expressions for $\hat{\psi}_{-1}^\infty$ and $\hat{\psi}_{-2}^\infty$ in (24) and (25).

D. Derivation of (116)

We first observe that

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}_I + A_1 \mathbf{q}_1 \mathbf{s}_1^\dagger + A_1 \mathbf{s}_1 \mathbf{q}_1^\dagger + P_1 \hat{f}_1(\bar{L}) \mathbf{s}_1 \mathbf{s}_1^\dagger \quad (135)$$

where \mathbf{q}_1 is given by (116) and

$$\hat{f}_1(\bar{L}) = \frac{1}{i} \text{trace}[\mathbf{W}] \rightarrow f_1(\bar{L}), \quad \text{as } (K, N, i) \rightarrow \infty.$$

Let

$$\hat{\mathbf{R}}_2 = \hat{\mathbf{R}}_I + A_1 \mathbf{q}_1 \mathbf{s}_1^\dagger \quad (136)$$

$$\hat{\mathbf{R}}_1 = \hat{\mathbf{R}}_2 + A_1 \mathbf{s}_1 \mathbf{q}_1^\dagger \quad (137)$$

so that

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}_1 + P_1 \hat{f}_1(\bar{L}) \mathbf{s}_1 \mathbf{s}_1^\dagger. \quad (138)$$

Applying the matrix inversion lemma in each case gives

$$\begin{aligned} \hat{\mathbf{R}}^{-1} \mathbf{s}_1 &= \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1 - \frac{\hat{\mathbf{R}}_1^{-1} \mathbf{s}_1 \times P_1 \hat{f}_1(\bar{L}) \mathbf{s}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1}{1 + P_1 \hat{f}_1(\bar{L}) \mathbf{s}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1} \\ &= \frac{\hat{\mathbf{R}}_1^{-1} \mathbf{s}_1}{1 + P_1 \hat{f}_1(\bar{L}) \mathbf{s}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1} \end{aligned} \quad (139)$$

$$\hat{\mathbf{R}}^{-1} \mathbf{q}_1 = \hat{\mathbf{R}}_1^{-1} \mathbf{q}_1 - \frac{\hat{\mathbf{R}}_1^{-1} \mathbf{s}_1 \times P_1 \hat{f}_1(\bar{L}) \mathbf{s}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{q}_1}{1 + P_1 \hat{f}_1(\bar{L}) \mathbf{s}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1} \quad (140)$$

$$\begin{aligned} \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1 &= \hat{\mathbf{R}}_2^{-1} \mathbf{s}_1 \left(1 - \frac{A_1 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1}{1 + A_1 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1} \right) \\ &= \frac{\hat{\mathbf{R}}_2^{-1} \mathbf{s}_1}{1 + A_1 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1} \end{aligned} \quad (141)$$

$$\hat{\mathbf{R}}_1^{-1} \mathbf{q}_1 = \hat{\mathbf{R}}_2^{-1} \mathbf{q}_1 - \frac{A_1 \hat{\mathbf{R}}_2^{-1} \mathbf{s}_1 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_2^{-1} \mathbf{q}_1}{1 + A_1 \mathbf{q}_1^\dagger \hat{\mathbf{R}}_1^{-1} \mathbf{s}_1} \quad (142)$$

$$\begin{aligned} \hat{\mathbf{R}}_2^{-1} \mathbf{s}_1 &= \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 - \frac{A_1 \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1}{1 + A_1 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1} \\ &\asymp \hat{\mathbf{R}}_I^{-1} \mathbf{s}_1 - A_1 \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \hat{\gamma}_{-1}^\infty \end{aligned} \quad (143)$$

$$\begin{aligned} \hat{\mathbf{R}}_2^{-1} \mathbf{q}_1 &= \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 - \frac{A_1 \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1}{1 + A_1 \mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1} \\ &\asymp \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \end{aligned} \quad (144)$$

where the last two limits follow from the fact that $\mathbf{s}_1^\dagger \hat{\mathbf{R}}_I^{-1} \mathbf{q}_1 \asymp 0$. Finally, combining (5) and (117), we have

$$\hat{\mathbf{s}}_1(i) = A_1 \hat{f}_1(\bar{L}) \mathbf{s}_1 + \mathbf{q}_1. \quad (145)$$

The expression (116) can be derived by combining (136)–(145) with the expressions $\mathbf{c}_1 = \hat{\mathbf{R}}^{-1} \hat{\mathbf{s}}_1$ (with training) and $\mathbf{c}_1 = \hat{\mathbf{R}}^{-1} \mathbf{s}_1$ (blind).

APPENDIX III DERIVATION OF (41)–(45)

We start by deriving (41). From (40) and the representation (35) and (36), we have

$$m_{\hat{\mathbf{R}}}(z) = \frac{1}{-z + \alpha_{\hat{\mathbf{R}}} \int \frac{\lambda}{1 + \lambda m_{\hat{\mathbf{R}}}(z)} dG_{\hat{\mathbf{R}}}(\lambda)} \quad (146)$$

where $G_{\hat{\mathbf{R}}}$ is the large system eigenvalue distribution of $\hat{\mathbf{R}}$ defined in (36), and $\alpha_{\hat{\mathbf{R}}}$ denotes the load associated with a covariance matrix $\hat{\mathbf{Q}} = \mathbf{V} \mathbf{P} \mathbf{V}^\dagger$, as defined in Appendix I-B. That is, $\alpha_{\hat{\mathbf{Q}}}$ is the number of columns of \mathbf{V} divided by the number of rows of \mathbf{V} , so that from (35) we have $\alpha_{\hat{\mathbf{R}}} = (K + L)/N$. From (36), we observe that $\hat{\mathbf{R}}$ has rank i , so that

$$dG_{\hat{\mathbf{R}}}(x) = \frac{i}{K + L} dG_{\hat{\mathbf{R}}}^{\setminus i}(x) + \frac{K + L - i}{K + L} \delta(x) dx \quad (147)$$

where $G_{\hat{\mathbf{R}}}^{\setminus i}(\cdot)$ is the eigenvalue distribution of $\hat{\mathbf{R}}$ without the $K + L - i$ zero eigenvalues, and $\delta(x)$ is the unit singular function. Substituting into (146) gives

$$m_{\hat{\mathbf{R}}}(z) = \frac{1}{-z + \eta \int \frac{\lambda}{1 + \lambda m_{\hat{\mathbf{R}}}(z)} dG_{\hat{\mathbf{R}}}(\lambda)}. \quad (148)$$

Now

$$\begin{aligned} &\int \frac{\lambda}{1 + \lambda m_{\hat{\mathbf{R}}}(z)} dG_{\hat{\mathbf{R}}}(\lambda) \\ &= \frac{1}{m_{\hat{\mathbf{R}}}(z)} \left(1 - \int \frac{dG_{\hat{\mathbf{R}}}(\lambda)}{1 + \lambda m_{\hat{\mathbf{R}}}(z)} \right) \\ &= \frac{1}{m_{\hat{\mathbf{R}}}(z)} \left(1 - \frac{1}{m_{\hat{\mathbf{R}}}(z)} \int \frac{dG_{\hat{\mathbf{R}}}(\lambda)}{\lambda + \frac{1}{m_{\hat{\mathbf{R}}}(z)}} \right) \\ &= \frac{1}{m_{\hat{\mathbf{R}}}(z)} \left(1 - \frac{1}{m_{\hat{\mathbf{R}}}(z)} m_{\hat{\mathbf{R}}}^{-1}[-m_{\hat{\mathbf{R}}}(z)] \right) \end{aligned} \quad (149)$$

and combining (149) with (148) gives

$$m_{\hat{\mathbf{R}}}[-m_{\hat{\mathbf{R}}}^{-1}(z)] = m_{\hat{\mathbf{R}}}(z) \left(1 - \frac{1}{\eta} [1 + zm_{\hat{\mathbf{R}}}(z)]\right). \quad (150)$$

From (40) we also have

$$m_{\hat{\mathbf{\Omega}}}[-m_{\hat{\mathbf{R}}}^{-1}(z)] = \frac{1}{\frac{1}{m_{\hat{\mathbf{R}}}(z)} + \alpha_{\hat{\mathbf{\Omega}}} \int \frac{P}{1+Pm_{\hat{\mathbf{\Omega}}}[-m_{\hat{\mathbf{R}}}^{-1}(z)]} dF_{\hat{\mathbf{\Omega}}}(P)} \quad (151)$$

where $\alpha_{\hat{\mathbf{\Omega}}} = (K + L)/i$ is the effective load associated with covariance matrix $\hat{\mathbf{\Omega}}$, and $F_{\hat{\mathbf{\Omega}}}(\cdot)$ is the large system power distribution associated with $\hat{\mathbf{\Omega}}$. From (36), we have

$$dF_{\hat{\mathbf{\Omega}}}(P) = \frac{K}{K+L} dF(P) + \frac{L}{K+L} \delta\left(P - \frac{N\sigma^2}{L}\right) \quad (152)$$

where $F(\cdot)$ is the distribution of the user powers (i.e., diagonal elements of $|\mathbf{A}_I|^2$). Substituting into (151) gives

$$m_{\hat{\mathbf{\Omega}}}[-m_{\hat{\mathbf{R}}}^{-1}(z)] = \frac{1}{\frac{1}{m_{\hat{\mathbf{R}}}(z)} + \frac{\sigma^2}{\eta} + \frac{\alpha}{\eta} \int \frac{P}{1+Pm_{\hat{\mathbf{\Omega}}}[-m_{\hat{\mathbf{R}}}^{-1}(z)]} dF(P)} \quad (153)$$

where the second term in the denominator is obtained by letting $L \rightarrow \infty$ so that $N/L \rightarrow 0$. This accounts for the additive Gaussian noise, as explained in Section IV. The relation (41) can now be obtained after some further manipulations by combining (153) with (150).

The relation (42) is obtained by taking the limit $z \rightarrow 0$ in (41). To derive (45), we use the analogous definition of $\gamma_n^\infty[x]$, as that in (52), and note that

$$\gamma_{-2}^\infty = -\frac{\partial \gamma_{-1}^\infty}{\partial x} = -\frac{\partial}{\partial x} \left(\frac{1}{x + \alpha \int \frac{PdF(P)}{1+P\gamma_{-1}^\infty[x]}} \right) \quad (154)$$

where we have used (44). Evaluating the derivative and combining again with (44) gives (45).

To derive (43), we first apply (45) to the sample covariance matrix $\hat{\mathbf{R}}$. As shown in Section IV, the effective power distribution is the eigenvalue distribution of $\hat{\mathbf{\Omega}}$. Hence, we have

$$\begin{aligned} \hat{\gamma}_{-2}^\infty &= \frac{\hat{\gamma}_{-1}^\infty}{\alpha_{\hat{\mathbf{R}}} \int \frac{\lambda dG_{\hat{\mathbf{\Omega}}}(\lambda)}{(1+\lambda\hat{\gamma}_{-1}^\infty)^2}} \\ &= \frac{\hat{\gamma}_{-1}^\infty}{\eta \int \frac{\lambda dG_{\hat{\mathbf{\Omega}}}(\lambda)}{(1+\lambda\hat{\gamma}_{-1}^\infty)^2}} \\ &= \frac{\hat{\gamma}_{-1}^\infty}{\frac{\eta}{\hat{\gamma}_{-1}^\infty} \left(\int \frac{dG_{\hat{\mathbf{\Omega}}}(\lambda)}{1+\lambda\hat{\gamma}_{-1}^\infty} - \int \frac{dG_{\hat{\mathbf{\Omega}}}(\lambda)}{(1+\lambda\hat{\gamma}_{-1}^\infty)^2} \right)} \end{aligned} \quad (155)$$

where $\alpha_{\hat{\mathbf{R}}} = (K + L)/i$ and we have used (147).

We now evaluate each term in the denominator. Taking the limit $z \rightarrow 0$ in (149) and (150) gives, after some manipulation

$$\int \frac{dG_{\hat{\mathbf{\Omega}}}(\lambda)}{1+\lambda\hat{\gamma}_{-1}^\infty} = 1 - \frac{1}{\eta}. \quad (156)$$

Combining with (45) gives

$$\begin{aligned} \int \frac{dG_{\hat{\mathbf{\Omega}}}(\lambda)}{(1+\lambda\hat{\gamma}_{-1}^\infty)^2} &= \frac{1}{(\hat{\gamma}_{-1}^\infty)^2} \int \frac{dG_{\hat{\mathbf{\Omega}}}(\lambda)}{\left(\frac{1}{\hat{\gamma}_{-1}^\infty} + \lambda\right)^2} \\ &= \frac{1}{(\hat{\gamma}_{-1}^\infty)^2} \frac{\int \frac{dG_{\hat{\mathbf{\Omega}}}(\lambda)}{\frac{1}{\hat{\gamma}_{-1}^\infty} + \lambda}}{\frac{1}{\hat{\gamma}_{-1}^\infty} + \alpha_{\hat{\mathbf{\Omega}}} \int \frac{PdF_{\hat{\mathbf{\Omega}}}(P)}{(1+P\hat{\gamma}_{-1}^\infty(1-\frac{1}{\eta}))^2}} \\ &= \frac{1}{\hat{\gamma}_{-1}^\infty} \frac{1 - \frac{1}{\eta}}{\frac{1}{\hat{\gamma}_{-1}^\infty} + \frac{\sigma^2}{\eta} + \frac{\alpha}{\eta} \int \frac{PdF(P)}{(1+P\hat{\gamma}_{-1}^\infty(1-\frac{1}{\eta}))^2}} \\ &= \frac{1}{\hat{\gamma}_{-1}^\infty} \frac{1 - \frac{1}{\eta}}{\frac{1}{\hat{\gamma}_{-1}^\infty} + \frac{\sigma^2}{\eta} + \frac{\alpha}{\eta} \int \frac{PdF(P)}{(1+P\hat{\gamma}_{-1}^\infty)^2}} \\ &= \frac{1}{\hat{\gamma}_{-1}^\infty} \frac{1 - \frac{1}{\eta}}{\frac{1}{\hat{\gamma}_{-1}^\infty} + \frac{1}{\eta} \left(\frac{\gamma_{-1}^\infty}{\gamma_{-2}^\infty}\right)} \end{aligned} \quad (157)$$

and combining (155)–(157) with (42) gives (43).

We now give a simple, direct derivation of (44) and (45), which does not rely on the Stieltjes transform. Specifically

$$\begin{aligned} 1 &= \lim_{N \rightarrow \infty} \frac{1}{N} \text{trace}(\mathbf{R}^{-1}\mathbf{R}) \\ &= \lim_{N \rightarrow \infty} \text{trace}[\mathbf{R}^{-1}(\sigma^2\mathbf{I} + \mathbf{SPS}^\dagger)] \\ &= \sigma^2 \lim_{N \rightarrow \infty} \frac{1}{N} \text{trace}\mathbf{R}^{-1} \\ &\quad + \lim_{N \rightarrow \infty} \sum_{k=2}^N \frac{1}{N} \frac{P_k (\mathbf{s}_k \mathbf{R}_{I,k}^{-1} \mathbf{s}_k)}{1 + P_k \mathbf{s}_k \mathbf{R}_{I,k}^{-1} \mathbf{s}_k} \\ &= \sigma^2 \gamma_{-1}^\infty + \alpha \int \frac{P\gamma_{-1}^\infty}{1 + P\gamma_{-1}^\infty} dF(P) \end{aligned} \quad (158)$$

and

$$\begin{aligned} \gamma_{-1}^\infty &= \lim_{N \rightarrow \infty} \frac{1}{N} \text{trace}(\mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1}) \\ &= \lim_{N \rightarrow \infty} \text{trace}[\mathbf{R}^{-1}(\sigma^2\mathbf{I} + \mathbf{SPS}^\dagger)\mathbf{R}^{-1}] \\ &= \sigma^2 \lim_{N \rightarrow \infty} \frac{1}{N} \text{trace}(\mathbf{R}^{-2}) \\ &\quad + \lim_{N \rightarrow \infty} \sum_{k=2}^N \frac{1}{N} \frac{P_k \mathbf{s}_k \mathbf{R}_{I,k}^{-2} \mathbf{s}_k}{(1 + P_k \mathbf{s}_k \mathbf{R}_{I,k}^{-1} \mathbf{s}_k)^2} \\ &= \sigma^2 \gamma_{-2}^\infty + \alpha \int \frac{P\gamma_{-2}^\infty}{(1 + P\gamma_{-1}^\infty)^2} dF(P) \end{aligned} \quad (159)$$

where $\mathbf{R}_{I,k}$ is the interference-plus-noise covariance matrix for user k , and we have used the relations

$$\mathbf{R}^{-1}\mathbf{s}_k = (\mathbf{R}_{I,k}^{-1}\mathbf{s}_k)/(1 + P_k(\mathbf{s}_k \mathbf{R}_{I,k}^{-1} \mathbf{s}_k))$$

$$\lim_{N \rightarrow \infty} \mathbf{s}_k^\dagger \mathbf{R}_{I,k}^{-n} \mathbf{s}_k = \gamma_{-n}^\infty, \text{ and (29).}$$

ACKNOWLEDGMENT

The authors wish to thank M. Peacock and the anonymous reviewers for their comments and corrections to the original manuscript, and R. Speicher for referring the authors to [42] and [41] along with helpful explanations.

REFERENCES

- [1] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [2] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications*. Boston, MA: Kluwer Academic, 1985.
- [3] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [4] J. S. Goldstein and I. S. Reed, "Reduced rank adaptive filtering," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 492–496, Feb. 1997.
- [5] M. L. Honig, "A comparison of subspace adaptive filtering techniques for DS-CDMA interference suppression," in *Proc. IEEE MILCOM*, Monterey, CA, Nov. 1997, pp. 836–840.
- [6] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 677–690, Mar. 1998.
- [7] G. Woodward, M. Honig, and B. Vucetic, "Lattice-based subspace decomposition for DS-CDMA detection," in *Proc. Int. Symp. Spread Spectrum Techniques and Applications (ISSSTA)*, Sun City, South Africa, Sep. 1998.
- [8] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.
- [9] Y. Song and S. Roy, "Blind adaptive reduced-rank detection for DS-CDMA signals in multipath channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1960–1970, Nov. 1999.
- [10] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1406–1415, Sep. 1999.
- [11] D. A. Pados and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 290–300, Feb. 2001.
- [12] M. L. Honig and J. S. Goldstein, "Adaptive reduced-rank interference suppression based on the multistage Wiener filter," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 986–994, Jun. 2002.
- [13] M. Joham, Y. Sun, M. D. Zoltowski, M. Honig, and S. Goldstein, "A new backward recursion for the multi-stage nested Wiener filter employing Krylov subspace methods," in *Proc. MILCOM 2001*, McLean, VA, Oct. 2001, pp. 1210–1213.
- [14] G. Dietl, M. D. Zoltowski, and M. Joham, "Reduced-rank equalization for EDGE via conjugate gradient implementation of multi-stage nested Wiener filter," in *Proc. Fall Vehicular Technology Conf. (VTC 2001)*, Atlantic City, NJ, Oct. 2001, pp. 1912–1916.
- [15] I. S. Reed, J. D. Mallet, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, pp. 853–863, Nov. 1974.
- [16] B. D. Van Veen, "Adaptive convergence of linearly constrained beamformers based on the sample covariance matrix," *IEEE Trans. Signal Process.*, vol. 39, no. 6, pp. 1470–1473, Jun. 1991.
- [17] J. L. Krolik and D. N. Swingler, "On the mean-square error performance of adaptive minimum variance beamformers based on the sample covariance matrix," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 445–448, Feb. 1994.
- [18] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.
- [19] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1928–1946, Jul. 2001.
- [20] J. Zhang and X. Wang, "Large-system performance analysis of blind and group-blind multiuser receivers," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2507–2523, Sep. 2002.
- [21] W. Chen, U. Mitra, and P. Schniter, "On the equivalence of three reduced rank linear estimators with applications to DS-CDMA," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2609–2614, Sep. 2002.
- [22] P. Loubaton and W. Hachem, "Asymptotic analysis of reduced rank Wiener filters," in *Proc. Information Theory Workshop*, Paris, France, Apr. 2003, pp. 328–331.
- [23] L. Li, A. M. Tulino, and S. Verdú, "Design of reduced-rank MMSE multiuser detectors using random matrix methods," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 986–1008, Jun. 2004.
- [24] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, ser. Foundations and Trends in Communication and Information Theory, 2004, vol. 1, pp. 1–182.
- [25] S. Moshavi, E. G. Kanterakis, and D. L. Schilling, "Multistage linear receivers for ds-cdma systems," *Int. J. Wireless Inf.*, vol. 3, pp. 1–17, Jan. 1996.
- [26] D. Guo, L. K. Rasmussen, and T.-J. Lim, "Linear parallel interference cancellation in long-code CDMA multiuser detection," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 12, pp. 2074–2081, Dec. 1999.
- [27] A. Grant and C. Schlegel, "Convergence of linear interference cancellation multiuser receivers," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1824–1834, Oct. 2001.
- [28] R. R. Müller and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1429–1441, Dec. 2001.
- [29] L. Trichard, "Large system analysis of linear multistage receivers," Ph.D. dissertation, Univ. Sydney, Sydney, Australia, 2001.
- [30] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivariate Anal.*, vol. 12, pp. 1–38, 1982.
- [31] Y. Q. Yin, "Limiting spectral distribution for a class of random matrix," *J. Multivariate Anal.*, vol. 20, pp. 50–68, 1986.
- [32] L. Li, A. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection," *Commun. in Inf. and Syst.*, vol. 1, no. 3, pp. 273–304, Fall 2001.
- [33] R. Singh and L. B. Milstein, "Interference suppression for ds/cdma," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 446–453, Mar. 1999.
- [34] W. Xiao, "Large system performance of linear interference suppression for DS-CDMA," Ph.D. dissertation, Northwestern Univ., Evanston, IL, 2001.
- [35] W. Xiao and M. L. Honig, "Performance of adaptive reduced-rank interference suppression in the presence of dynamic fading," in *Proc. 2000 Asilomar Conf. Signals, Systems, and Computers*, vol. 2, Pacific Grove, CA, Nov. 2000, pp. 1143–1147.
- [36] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review," *Statist. Sinica*, vol. 9, pp. 611–677, 1999.
- [37] W. Xiao and M. L. Honig, "Large system convergence analysis of adaptive recursive least squares algorithms," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2001, pp. 1142–1151.
- [38] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.
- [39] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [40] P. Biane, "Minimal factorizations of cycle and central multiplicative functions on the infinite symmetric group," *J. Combin. Theory*, ser. A, vol. 76, no. 2, pp. 197–212, 1996.
- [41] I. P. Goulden and D. M. Jackson, "Connexion coefficients for the symmetric group, free products in operator algebras and random matrices," in *Fields Institute Communications: Free Probability Theory*, D.-V. Voiculescu, Ed. Providence, RI: Amer. Math. Soc., 1997, vol. 12, pp. 105–125.
- [42] ———, "The combinatorial relationship between trees, cacti and certain connection coefficients for the symmetric group," *Europ. J. Combin.*, vol. 13, pp. 357–365, 1992.