Accuracy of the Wiener-Hopf Solution When Based on Sample Statistics

Stephan Weiss, *Senior Member, IEEE*, Ian K. Proudler, and Yoann Altmann, *Member, IEEE*,

***Abstract*—This paper addresses the problem of an adaptive linear combiner when using sample statistics. In this case, the Wiener-Hopf solution can be based on sample covariance matrices and cross-correlation vectors. Despite these estimates being inaccurate, the calculated, approximate solution can be very precise. We explore why this is so by interpreting the Wiener-Hopf solution as coming from a least squares problem. We compare this solution to the case where we exploit knowledge about some of the statistics. Surprisingly this has very limited benefits and often is detrimental. We also show why it is disadvantageous to separately estimate the two statistics.**

***Index Terms*—adaptive systems, least squares approximation, systems identification**

I. INTRODUCTION

FOR linear combining, the Wiener filter [1], [2] provides the optimum solution both in terms of mean square and least squares errors when the signal statistics are stationary and known [3], [4]. In the discrete time case, the linear combiner problem is stated as $\mathbf{Rw} = \mathbf{p}$ [5], where $\mathbf{w}$ linearly combines an input data vector with covariance $\mathbf{R}$, whose cross-correlation with a desired signal yields a cross-correlation vector $\mathbf{p}$. If $\mathbf{R}$ is full rank, the Wiener-Hopf solution is $\mathbf{w}_\mathrm{o} = \mathbf{R}^{-1}\mathbf{p}$, where $\mathbf{w}_\mathrm{o}$ is the optimum linear combiner. Many practical methods, such as least mean squares and recursive least squares algorithms [6]–[8], should, in the limit, converge to $\mathbf{w}_\mathrm{o}$ [9], [10].

If exact signal statistics are not available, we can approximate $\hat{\mathbf{w}}_\mathrm{o} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{p}}$ via estimated quantities $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$. This for example forms the starting point of the recursive least squares algorithm [6], [7]. The estimates $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ may be imprecise as their variances only reduce inversely proportional to the sample size. However, the result for $\hat{\mathbf{w}}_\mathrm{o}$ can be surprisingly accurate. In contrast, if we exploit knowing $\mathbf{R}$, $\hat{\mathbf{w}}_\mathrm{o}$ can be rather imprecise: e.g. substituting the exact $\mathbf{R} = \mathbf{I}$ produces $\hat{\mathbf{w}}_\mathrm{o} = \hat{\mathbf{p}}$.

In order to explore the above discrepancy in precision, in this paper we investigate three questions: (i) why can $\hat{\mathbf{R}}^{-1}\hat{\mathbf{p}}$ be accurate despite imprecise statistical estimates? (ii) Is it possible to improve the solution $\hat{\mathbf{w}}_0$ if the input signal statistics — often under a user's control e.g. in a system identification setup — are known? (iii) Related to the last point, we also want to see how estimating $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ from different data sets will impact on the solution.

S. Weiss and I.K. Proudler are with Dept. of EEE, University of Strathclyde, Glasgow, Scotland (e-mail: {stephan.weiss;ian.proudler}@strath.ac.uk).
Y. Altmann is with School of Eng. & Physical Sciences, Heriot Watt University, Edinburgh, Scotland (e-mail: y.altmann@hw.ac.uk)

In this letter, we review the Wiener-Hopf solution and restate the problem as a least squares minimisation. Compared to earlier work [11], we present corrected variance terms, include a theoretical analysis for the case of statistics drawn from different data sets, and, based on simulations[1], and provide an assessment of the statistical significance of these results. In our analysis, we assume that all signals are zero-mean circularly-symmetric complex Gaussian. In order to track covariance estimates via Wishart distributions, all signals are assumed to be spatially but not temporally correlated. This excludes that data is derived from a tap-delay-line structure as in a temporal filter. Instead, we focus on the more general linear combiner, but we briefly comment on temporal correlation in Sec. VII.

II. SYSTEM IDENTIFICATION AND WIENER-HOPF

System Identification Setup and Assumptions. As a representative use of the Wiener-Hopf solution, we consider an adaptive system identification problem [6], [7] as described in Fig. 1. An unknown linear system $\boldsymbol{\omega} \in \mathbb{C}^M$ is excited by a signal $\mathbf{x}_n \in \mathbb{C}^M$. The same signal $\mathbf{x}_n$ feeds into an adaptive combiner with coefficient vector $\mathbf{w} \in \mathbb{C}^M$, whose dimension for simplicity matches the size M of the unknown system. The output $y[n] = \mathbf{w}^\mathrm{H}\mathbf{x}_n$ of the adaptive combiner $\mathbf{w}$ is compared to the signal $d[n] = \boldsymbol{\omega}^\mathrm{H}\mathbf{x}_n + v[n]$, with $v[n]$ representing observation noise. The adaptive system aims to adjust $\mathbf{w}$ in order to minimise the error $e[n] = d[n] - y[n]$ in the mean square sense.

We assume that the observation noise $v[n]$ is independent from the input signal $\mathbf{x}_n$. For tractability of the subsequent analysis, we further assume that both input and observation noise are temporally uncorrelated. Thus, with the expectation operator $\mathcal{E}\{\cdot\}$ and $\delta[\tau]$ denoting the Kronecker delta function, we have $\mathcal{E}\{v[n]v^*[n-\tau]\} = \sigma_v^2\delta[\tau]$, where σ_v^2 is the observation noise power, and $\mathcal{E}\big\{\mathbf{x}_n\mathbf{x}_{n-\tau}^\mathrm{H}\big\} = \mathbf{R}\delta[\tau]$.

Wiener-Hopf Solution. To minimise the mean square value $\chi = \mathcal{E}\big\{|e[n]|^2\big\}$ of the error $e[n]$ in Fig. 1, we formulate

$$\mathbf{w}_\mathrm{o} = \arg\min_{\mathbf{w}} \mathcal{E}\big\{|e[n]|^2\big\} \ . \tag{1}$$

[1]Code available at https://github.com/StephanWeiss5/WienerHopfAccuracy.

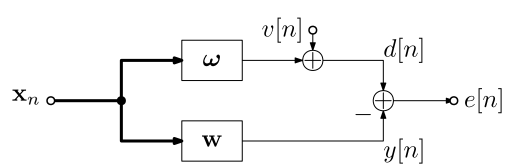

Fig. 1. System identification of an unknown system $\boldsymbol{\omega}$ by an adaptive combiner $\mathbf{w}$.

With $e[n] = d[n] - \mathbf{w}^{\mathrm{H}}\mathbf{x}_n$, the mean square error depends on the covariance matrix of the input, $\mathbf{R} = \mathcal{E}\{\mathbf{x}_n\mathbf{x}_n^{\mathrm{H}}\}$, as well as the cross-correlation vector $\boldsymbol{p} = \mathcal{E}\{d^*[n]\mathbf{x}_n\}$ between the desired signal and the input. Provided that $\mathbf{R}$ has full rank, the so-called Wiener-Hopf solution [6]–[8],

$$\mathbf{w}_{\mathrm{o}} = \mathbf{R}^{-1}\mathbf{p} \;, \tag{2}$$

solves the problem in (1).

System Identification Metrics. Since we are interested in the accuracy of system identification, we define a system error $\boldsymbol{\epsilon} = \mathbf{w} - \boldsymbol{\omega}$ between the unknown system and our model. We define a mean square value ξ as

$$\xi = \mathcal{E}\big\{\|\mathbf{w} - \boldsymbol{\omega}\|_2^2\big\} = \mathcal{E}\big\{\boldsymbol{\epsilon}^{\mathrm{H}}\boldsymbol{\epsilon}\big\} = \mathrm{tr}\big\{\mathcal{E}\big\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{H}}\big\}\big\} \;, \tag{3}$$

with $\mathrm{tr}\{\cdot\}$ the trace operator. We have already defined the mean square error $\chi = \mathcal{E}\big\{|e[n]|^2\big\} = \mathcal{E}\big\{|d[n] - \mathbf{w}^{\mathrm{H}}\mathbf{x}_n|^2\big\}$, which represents the cost term in (1). It can be written as

$$\chi = \mathcal{E}\big\{v[n] + \boldsymbol{\omega}^{\mathrm{H}}\mathbf{x}_n - \mathbf{w}^{\mathrm{H}}\mathbf{x}_n|^2\big\} = \sigma_v^2 + \mathcal{E}\big\{\boldsymbol{\epsilon}^{\mathrm{H}}\mathbf{R}\boldsymbol{\epsilon}\big\} \;. \tag{4}$$

Thus, compared to the mean square system error ξ in (3), the mean square error χ in (4) involves a weighted inner product of $\boldsymbol{\epsilon}$. Thus, it is only for $\mathbf{R} = c\mathbf{I}$, for some $c \in \mathbb{R}$, that the optimisations with respect to ξ and χ are equivalent.

III. Second Order Sample Statistics

Auto- and Cross-Correlation Estimators. The covariance matrix $\mathbf{R}$ and the cross-correlation vector $\mathbf{p}$ in (2) must in practice be estimated from a finite sample size, say N snapshots for time indices $n = 0, \ldots, (N-1)$. With

$$\begin{aligned}\mathbf{X} &= [\mathbf{x}_0,\; \mathbf{x}_1,\; \ldots,\; \mathbf{x}_{N-1}] \\ \mathbf{d} &= [d[0],\; d[1],\; \ldots,\; d[N-1]]^{\mathrm{H}} \;.\end{aligned}$$

we can formulate estimates

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{H}} \;, \qquad \hat{\mathbf{p}} = \frac{1}{N}\mathbf{X}\mathbf{d} \;. \tag{5}$$

Since $\mathbf{x}_n$ is zero mean, we find that

$$\mathcal{E}\big\{\hat{\mathbf{R}}\big\} = \frac{1}{N}\mathcal{E}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\} = \frac{1}{N}\sum_{n=0}^{N-1}\mathcal{E}\{\mathbf{x}_n\mathbf{x}_n^{\mathrm{H}}\} = \mathbf{R} \tag{6}$$

and likewise $\mathcal{E}\{\hat{\mathbf{p}}\} = \mathbf{p}$. So the estimators in (5) are unbiased.

Variance of Sample Covariance Estimator. To judge how the estimates in (5) improve with N, we investigate the variance of the covariance estimator,

$$\begin{aligned}\mathrm{var}\{\hat{\mathbf{R}}\} &= \mathcal{E}\Big\{(\hat{\mathbf{R}} - \mathbf{R})(\hat{\mathbf{R}} - \mathbf{R})^{\mathrm{H}}\Big\} \\ &= \frac{1}{N^2}\mathcal{E}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\} - \mathbf{R}\mathbf{R}^{\mathrm{H}} \;, \end{aligned} \tag{7}$$

where the simplification step exploited (6). By resolving the first term in (7) using the Appendix (with $\boldsymbol{\Psi} = \mathbf{I}$), we obtain

$$\mathrm{var}\{\hat{\mathbf{R}}\} = \frac{1}{N}\mathrm{tr}\{\mathbf{R}\}\,\mathbf{R} \;. \tag{8}$$

This is consistent with results in e.g. [12]–[14]. A similar dependency can be shown for the cross-correlation estimator [15], i.e. the estimators in (5) only converge inversely proportional to the sample size N.

Illustrative Example. As an example for using sample statistics for the Wiener-Hopf solution, we consider the system $\boldsymbol{\omega}^{\mathrm{H}} = [1, \frac{1}{2}, \frac{1}{4}]$ and $N = 100$ input vectors $\mathbf{x}_n$ with $\mathbf{R} = \mathbf{I}$. With an observation noise power $\sigma_v^2 = 10^{-3}$, for one realisation of the sample statistics we obtain $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ such that

$$\|\mathbf{R} - \hat{\mathbf{R}}\|_{\mathrm{F}}^2 = 0.1235 \;, \qquad \|\mathbf{p} - \hat{\mathbf{p}}\|_2^2 = 0.0151 \;. \tag{9}$$

For comparison, note that $\mathcal{E}\Big\{\|\mathbf{R} - \hat{\mathbf{R}}\|_{\mathrm{F}}^2\Big\} = \mathrm{tr}\Big\{\mathrm{var}\{\hat{\mathbf{R}}\}\Big\}$ and thus with (8),

$$\mathcal{E}\Big\{\|\mathbf{R} - \hat{\mathbf{R}}\|_{\mathrm{F}}^2\Big\} = \frac{1}{N}\mathrm{tr}\{\mathbf{R}\}^2 = \frac{M^2}{N} = 0.09$$

and $\mathcal{E}\big\{\|\mathbf{p} - \hat{\mathbf{p}}\|_2^2\big\} = \frac{M}{N}\|\omega\|_2^2 = 0.0413$ [11], which are of the same order as the results in (9). Despite these inaccuracies of the estimates, for the approximate Wiener-Hopf solution, $\hat{\mathbf{w}}_{\mathrm{o}} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{p}}$, we obtain a system error

$$\|\hat{\mathbf{w}}_{\mathrm{o}} - \boldsymbol{\omega}\|_2^2 = 3.69 \cdot 10^{-5} \;. \tag{10}$$

Hence the Wiener-Hopf solution is more accurate than the estimates on which it is based by orders of magnitude. In absence of observation noise, i.e. $\sigma_v^2 = 0$ but otherwise with the same data, the system error in (10) drops to $\approx 10^{-31}$, i.e. close to machine accuracy in a double floating point format.

IV. Least Squares Analysis

Questions. The example in Sec. III raises some questions which we want to investigate as specific cases using a least squares analysis further below:

(C1) Why can the Wiener-Hopf solution be accurate despite using inaccurate sample statistics?

(C2) Since in a system identification setting we may know the statistics of the input, is there a benefit if we utilise knowledge of $\mathbf{R}$ to replace $\hat{\mathbf{R}}$?

The last point triggers a more hypothetical question:

(C3) What happens if we calculate $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ from two independent measurements?

We will begin by investigating case (C3) as it is the most general scenario. Consider two different data matrices $\mathbf{X}_i$, $i = 1, 2$, such that $\hat{\mathbf{R}} = \mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}/N$ and $\hat{\mathbf{p}} = \mathbf{X}_2\mathbf{d}/N$. From this, we derive case (C2) by replacing $\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}/N$ with $\mathbf{R}$, and case (C1) by setting $\mathbf{X}_1 = \mathbf{X}_2$. In addition to the assumptions in Sec. I, we consider the data matrices to have full row rank.

Least Squares Formulation. For the most general case (C3), we replace (5) by

$$\begin{aligned}\hat{\mathbf{R}} &= \frac{1}{N}\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}} \;, \\ \hat{\mathbf{p}} &= \frac{1}{N}\mathbf{X}_2\mathbf{d} = \frac{1}{N}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\boldsymbol{\omega} + \frac{1}{N}\mathbf{X}_2\mathbf{v} \;.\end{aligned}$$

Inserting this into the Wiener-Hopf solution, we have

$$\hat{\mathbf{w}}_{\mathrm{o}} = \left(\mathbf{X}_1^{\mathrm{H}}\mathbf{X}_1\right)^{-1}\mathbf{X}_2^{\mathrm{H}}\mathbf{X}_2\boldsymbol{\omega} + \left(\mathbf{X}_1^{\mathrm{H}}\mathbf{X}_1\right)^{-1}\mathbf{X}_2^{\mathrm{H}}\mathbf{v} \;.$$

For the system error $\boldsymbol{\epsilon} = \hat{\mathbf{w}}_{\mathrm{o}} - \boldsymbol{\omega}$,

$$\begin{aligned}\boldsymbol{\epsilon} = &\left(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)^{-1}\left(\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}} - \mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)\boldsymbol{\omega} \\ &+ \left(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)^{-1}\mathbf{X}_2\mathbf{v} \;, \end{aligned} \tag{11}$$

we find a system-related term depending on $\boldsymbol{\omega}$, and another term due to the observation noise $\mathbf{v}$.

Least Squares Interpretation. For case (C1) with $\mathbf{X}_1 = \mathbf{X}_2$, in (11) the system-related term vanishes, and only $\boldsymbol{\epsilon} = \mathbf{X}_1^{\dagger}\mathbf{v}$ remains, where $\{\cdot\}^{\dagger}$ denotes the pseudo-inverse, providing the smallest possible system error in the least squares sense, given $\mathbf{X}_1$ and $\mathbf{v}$. Thus, the solution does not depend on the accuracy of $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$, as long as $\mathbf{X}_1$ has full row rank.

For cases (C2) and (C3), the system-related term in (11) does not drop out, and its impact requires closer investigation of the mean square system error ξ.

V. Mean Square System Error Analysis

By inserting (11) into (3), the result can be organised as

$$\xi = \xi_\omega + \xi_v \tag{12}$$

with a system-related term ξ_ω and an observation noise-related term ξ_v. Using trace rules, $\mathcal{E}\{\mathbf{v}\mathbf{v}^{\mathrm{H}}\} = \sigma_v^2\mathbf{I}$, and abbreviating $\mathbf{W} = \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{H}}$, we find

$$\xi_\omega = \mathrm{tr}\Big\{\mathbf{W}\Big(\mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\} + \mathbf{I} - \mathcal{E}\{(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\} - \mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\}\Big)\Big\} , \tag{13}$$

$$\xi_v = \sigma_v^2\mathrm{tr}\{\mathcal{E}\{\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\}\} . \tag{14}$$

We investigate these terms ξ_v and ξ_ω in turn.

A. Observation Noise-Related Error Term

For case (C1), (14) simplifies to an inverse Wishart-distributed term $(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}$ which has the expected value [16], [17]

$$\mathcal{E}\{(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\} = \frac{1}{N-M}\mathbf{R}^{-1} . \tag{15}$$

Therefore ξ_v simplifies to

$$\xi_{v,1} = \sigma_v^2\mathrm{tr}\{\mathcal{E}\{(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\}\} = \frac{\sigma^2}{N-M}\mathrm{tr}\{\mathbf{R}^{-1}\} . \tag{16}$$

In case of (C2), we replace $(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}$ with $\mathbf{R}^{-2}/N^2$ and obtain

$$\xi_{v,2} = \frac{\sigma_v^2}{N}\mathrm{tr}\{\mathbf{R}^{-1}\} . \tag{17}$$

For case (C3) with independent $\mathbf{X}_1$ and $\mathbf{X}_2$,

$$\xi_{v,3} = \sigma_v^2\mathrm{tr}\{\mathcal{E}\{(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}\}\,\mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\}\} ,$$

we additionally exploit the fact that [16], [17]

$$\mathcal{E}\{(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}\} = C\Big(\mathbf{R}^{-2} + \frac{1}{(N-M)}\mathbf{R}^{-1}\mathrm{tr}\{\mathbf{R}^{-1}\}\Big)$$

with $C = [(N-M)^2-1]^{-1}$ to yield

$$\xi_{v,3} = \frac{N^2\sigma_v^2}{(N-M)[(N-M)^2-1]}\mathrm{tr}\{\mathbf{R}^{-1}\} . \tag{18}$$

Thus, assuming that $N > M+1$, we find that $\xi_{v,2} < \xi_{v,1} < \xi_{v,3}$, while for $N \gg M$, $\xi_{v,1}$ and $\xi_{v,3}$ approach $\xi_{v,2}$.

B. System-Related Error Term

For case (C1), it is easy to see that for (13) and in line with the results in Sec IV, we have $\xi_{\omega,1} = 0$. For case (C2) with $\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}$ replaced by $N\mathbf{R}$, (13) simplifies to

$$\xi_{\omega,2} = \frac{1}{N^2}\mathrm{tr}\Big\{\mathbf{W}\Big(\mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\mathbf{R}^2\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\} - \mathbf{I}\Big)\Big\} .$$

Resolving the expected term with the help of (22) in the Appendix by setting $\boldsymbol{\Psi} = \mathbf{R}$, we obtain

$$\xi_{\omega,2} = \frac{1}{N}\mathrm{tr}\{\mathbf{R}^{-1}\}\,\mathrm{tr}\{\mathbf{R}\mathbf{W}\} . \tag{19}$$

For case (C3), as a first step we address the term $\boldsymbol{\Theta} = \mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\}$ in (13) using (22) in the Appendix with $\boldsymbol{\Psi} = (\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}$:

$$\boldsymbol{\Theta} = N^2C\mathbf{I} + \frac{2N^2C}{N-M}\mathrm{tr}\{\mathbf{R}^{-1}\}\,\mathbf{R} .$$

Inserting into (13) and with (15), we obtain

$$\xi_{\omega,3} = \frac{N(M^2+1) - M(M^2-1)}{(N-M)[(N-M)^2-1]}\mathrm{tr}\{\mathbf{W}\} + \frac{2N^2}{(N-M)[(N-M)^2-1]}\mathrm{tr}\{\mathbf{R}^{-1}\}\,\mathrm{tr}\{\mathbf{R}\mathbf{W}\} , \tag{20}$$

provided that $N > M+1$. Under this provision, we have $\xi_{\omega,3} > \xi_{\omega,2} > \xi_{\omega,1}$.

VI. Simulations and Results

For an unknown system $\boldsymbol{\omega} \in \mathbb{C}^5$ with coefficients drawn from a complex Gaussian distribution, we calculate the Wiener-Hopf solution for different signal to noise ratios (SNRs), defined as $1/\sigma_v^2$, between -20 dB and 40 dB. For three different samples sizes $N \in \{30, 300, 3000\}$ and with $M = 5$ coefficients in $\mathbf{w}$, an ensemble of 10^6 experiments is performed for each parameter combination and for the three different cases. For each ensemble probe, the data is generated via $\mathbf{x}_n = \mathbf{A}\mathbf{u}_n$ with $\mathbf{u}_n \sim \mathcal{CN}(\underline{0}, \mathbf{I})$, where a matrix $\mathbf{A}$ has singular values $\sigma_m = (\frac{1}{2})^{m-1}$. Thus, $\mathbf{x}_n \sim \mathcal{CN}(\underline{0}, \mathbf{R})$, where $\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathrm{H}}$ has eigenvalues $\lambda_m = \sigma_m^2$, $m = 1, \ldots, M$, and describes the spatial correlation of the data.

Fig. 2 shows both theoretical and simulated results versus the SNR. Theoretical mean square system errors $\xi_i = \xi_{\omega,i} + \xi_{v,i}$, $i = 1, 2, 3$, are given by (16), (17), (18), (19), and (20). Note $\xi_{\omega,1} = 0$. For the simulations, individual squared system errors $\|\boldsymbol{\epsilon}\|_2^2$ are averaged to give ensemble means $\hat{\xi}_i$. Additionally, Fig. 3 characterises the estimated probability density function $\hat{p}(\|\boldsymbol{\epsilon}\|_2^2)$ of the system error at two separate SNR values, noting that as a result of the heavy-tailed distributions and the logarithmic axis, the means appear off-centre.

The theoretical and simulated results agree very closely. At low SNR, the error decays approximately with $\frac{1}{N}$, as expected for $N \gg M$ according to Sec. V-A. Particularly for low N, here noticeable for $N = 30$ in Fig. 2, on average the best performance is achieved if knowledge of $\mathbf{R}$ is exploited. However, the differences of the means are not statistically significant given the distributions in Fig. 4. For cases (C2) and (C3), at high SNR ξ_i saturates due to the system errors

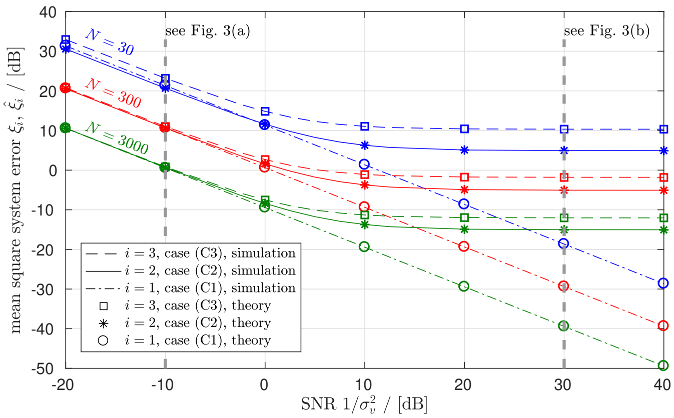

Fig. 2. Mean squared system error ξ versus signal to noise ratio $1/\sigma_v^2$ for sample sizes $\{30, 300, 3000\}$ for the three cases, compared to ensemble averages $\hat{\xi}_i$, $i = 1, 2, 3$.

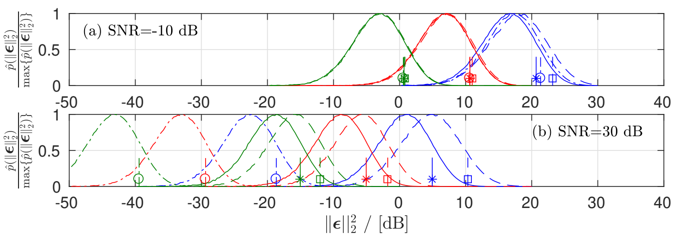

Fig. 3. Measured normalised distributions of squared system error $\|\boldsymbol{\epsilon}\|^2$ for (a) - 10 dB and (b) 30 dB SNR; markers and vertical lines indicate theoretical and ensemble means, respectively, as shown in the legend of Fig. 2.

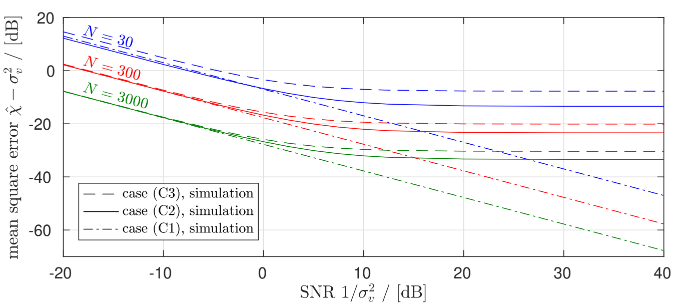

Fig. 4. Ensemble mean squared error $\hat{\chi}$ versus SNR $1/\sigma_v^2$.

$\xi_{\omega,i}$, $i = 2, 3$. Here the differences are statistically significant. Thus at higher SNR, case (C1) is always preferable.

In the above simulations, the mean squared error χ in (4) minimised by the Wiener-Hopf solution is not equivalent to minimising ξ due to the choice of $\mathbf{R}$, as seen in (4). Therefore, Fig. 4 shows the ensemble mean square error $\hat{\chi}$, corrected for the offset due to the observation noise power σ_v^2. Even though $\mathbf{R}$ is not particularly well conditioned, $(\hat{\chi}-\sigma_v^2)$ behaves similarly to ξ in Fig. 2, and the relative performance between the three cases remains unaltered.

VII. Discussion and Conclusions

For a linear combiner, we have interpreted the approximate Wiener-Hopf solution based on sample statistics as a least squares problem, and separated the error into a system-related and an observation noise-related term. The former is absent if both $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ are calculated — here with the best linear unbiased estimator — from the same data set as in case (C1). Replacing $\hat{\mathbf{R}}$ by the exact statistics, if known, for case (C2) in principle offers a lower observation-noise related mean squared error term. But their mean values do not differ significantly, and (C2) only offers benefits for low SNR and for low sample size, i.e. in situations where the performance w.r.t. the error metric is poor anyway. The hypothetical case (C3) of calculating $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ from different, independent data sets unsurprisingly gave the poorest results. Hence it appears generally advantageous to operate on one data set, and neglect knowledge of $\mathbf{R}$ in order to exploit the least squares minimisation at the heart of the Wiener-Hopf solution when approximated using sample statistics. The conditioning of $\mathbf{R}$ impacts proportionally on the error metric, which motivates near-orthogonal input sequences [20]–[22].

For temporally correlated data, e.g. if $\mathbf{x}_n$ is a tap delay line vector, the effective sample size decreases, and the variance of estimators increases. With $\mathbf{X}\mathbf{X}^{\mathrm{H}}$ being no longer Wishart distributed, the analysis becomes difficult; yet our results present upper bounds on the performance which are reached for the temporally uncorrelated case.

Appendix: Resolving $\mathcal{E}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\boldsymbol{\Psi}\mathbf{X}\mathbf{X}^{\mathrm{H}}\}$

For a data matrix $\mathbf{X} \in \mathbb{C}^{M\times N}$, we want to simplify

$$\boldsymbol{\Phi} = \mathcal{E}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\boldsymbol{\Psi}\mathbf{X}\mathbf{X}^{\mathrm{H}}\} \ , \tag{21}$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{M\times M}$ is independent of $\mathbf{X}$. Let $\Phi_{m,n}$ be the element in the mth row and nth column of $\boldsymbol{\Phi}$, $X_{m,\alpha}$ an element of $\mathbf{X}$, and $\Psi_{k,\ell}$ an element of $\boldsymbol{\Psi}$. We have

$$\begin{aligned}\Phi_{m,n} &= \mathcal{E}\Bigg\{\sum_{\alpha,\beta=1}^{N}\sum_{k,\ell=1}^{M} X_{m,\alpha}X^*_{k,\alpha}\Psi_{k,\ell}X_{\ell,\beta}X^*_{n,\beta}\Bigg\}\\ &= \sum_{\alpha,\beta=1}^{N}\sum_{k,\ell=1}^{M}\mathcal{E}\{\Psi_{k,\ell}\}\,\mathcal{E}\{X_{m,\alpha}X^*_{k,\alpha}X_{\ell,\beta}X^*_{n,\beta}\} \ .\end{aligned}$$

Using Isserlis' formula [18] to simplify the 4th order moment, we obtain

$$\begin{aligned}\Phi_{m,n} = \sum_{\alpha,\beta=1}^{N}\sum_{k,\ell=1}^{M}\Big(&\mathcal{E}\{X_{m,\alpha}X^*_{k,\alpha}\}\,\mathcal{E}\{X_{\ell,\beta}X^*_{n,\beta}\}\\ &+\mathcal{E}\{X_{m,\alpha}X_{\ell,\beta}\}\,\mathcal{E}\{X^*_{k,\alpha}X^*_{n,\beta}\}\\ &+\mathcal{E}\{X_{m,\alpha}X^*_{n,\beta}\}\,\mathcal{E}\{X^*_{k,\alpha}X_{\ell,\beta}\}\Big)\mathcal{E}\{\Psi_{k,\ell}\} \ .\end{aligned}$$

For circularly symmetric Gaussian data, $\mathcal{E}\{\mathbf{X}\mathbf{X}^{\mathrm{T}}\} = \mathbf{0}$ [19]. With $R_{m,n}$ an element of $\mathbf{R} = \frac{1}{N}\mathcal{E}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\}$, this simplifies to

$$\begin{aligned}\Phi_{m,n} = \sum_{k,\ell=1}^{M}\big(&N^2R_{m,k}\mathcal{E}\{\Psi_{k,\ell}\}\,R_{\ell,n}\\ &+NR_{m,n}\mathcal{E}\{\Psi_{k,\ell}\}\,R_{\ell,k}\big) \ ,\end{aligned}$$

Thus, overall (21) becomes

$$\boldsymbol{\Phi} = N^2\mathbf{R}\mathcal{E}\{\boldsymbol{\Psi}\}\,\mathbf{R} + N\mathbf{R}\mathrm{tr}\{\mathcal{E}\{\boldsymbol{\Psi}\}\,\mathbf{R}\} \ . \tag{22}$$

References

[1] A. N. Kolmogorov, "Interpolation und Extrapolation von stationären zufälligen Folgen," *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, vol. 5, no. 1, pp. 3–14, 1941.

[2] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. New York: Wiley, 1949.

[3] H. Bode and C. Shannon, "A simplified derivation of linear least square smoothing and prediction theory," *Proceedings of the IRE*, vol. 38, no. 4, pp. 417–425, 1950.

[4] T. Kailath, *Lectures on Wiener and Kalman Filtering*, 2nd ed. Wien: Springer Verlag, 1981.

[5] A. Oppenheim and G. Verghese, *Signals, Systems and Inference*. Pearson Education, 2015.

[6] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, New York: Prentice Hall, 1985.

[7] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs: Prentice Hall, 1991.

[8] A. H. Sayed, *Adaptive Filter*. John Wiley & Sons, 2011.

[9] B. Widrow and E. Walach, *Adaptive Inverse Control: A Signal Processing Approach*. Wiley – IEEE Press, 2008, ch. 2. Wiener Filters, pp. 40–58.

[10] H. Boche, V. Pohl, and H.V. Poor, "The Wiener theory of causal linear prediction is not effective," in *62nd IEEE Conference on Decision and Control*, 2023, pp. 8229–8234.

[11] S. Weiss and I.K. Proudler, "Why can Wiener-Hopf reach accurate solutions with poor statistical estimates?" in *23rd IEEE Statistical Signal Processing Workshop*, Edinburgh, UK, June 2025, pp. 146-150.

[12] W. Bär and F. Dittrich, "Useful formula for moment computation of normal random variables with nonzero means," *IEEE Transactions on Automatic Control*, vol. 16, no. 3, pp. 263–265, June 1971.

[13] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4775–4789, Sept 2012.

[14] C. Delaosa, J. Pestana, N. J. Goddard, S. Somasundaram, and S. Weiss, "Sample space-time covariance matrix estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, pp. 8033–8037, May 2019.

[15] C. Delaosa, J. Pestana, I. K. Proudler, and S. Weiss, "Impact of space-time covariance matrix estimation on bin-wise eigenvalue and eigenspace perturbations," *Signal Processing*, vol. 233, 109946, Aug. 2025.

[16] D. Maiwald and D. Kraus, "On moments of complex Wishart and complex inverse Wishart distributed matrices," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, Germany, vol. 5, pp. 3817–3820, Apr. 1997.

[17] ——, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 147, pp. 162–168, 2000.

[18] D. R. Brillinger, *Time series: data analysis and theory*. Society for Industrial and Applied Mathematics, 2001.

[19] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data. The Theory of Improper and Non-Circular Signals*. Cambridge University Press, 2010.

[20] C. Antweiler and M. Dörbecker, "Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control," *Annales des Télécommunications*, vol. 49, pp. 386–397, Jul. 1994.

[21] R. Martin, U. Heute, and C. Antweiler, *Advances in Digital Speech Transmission*. Wiley, 2008, ch. Multi-Channel System Identification with Perfect Sequences – Theory and Applications –, pp. 171–198.

[22] B. Wahlberg, H. Hjalmarsson, and M. Annergren, "On optimal input design in system identification for control," in *49th IEEE Conference on Decision and Control*, 2010, pp. 5548–5553.