Why Can Wiener-Hopf Reach Accurate Solutions With Poor Statistical Estimates?

Stephan Weiss and Ian K. Proudler
Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, Scotland
{stephan.weiss, ian.proudler}@strath.ac.uk

***Abstract*—The Wiener-Hopf solution is based on estimates of the covariance matrix and the cross-correlation vector, which will likely be poor for small sample sizes. Yet the solution can be very accurate. We explore why this is, and why it is disadvantageous to separately estimate the two statistics, or try to include explicit knowledge of the input process to the adaptive system. The results are based on a least squares interpretation of Wiener-Hopf and an exploration of the variances of estimates, which we underpin by simulations.**

I. Introduction

For linear filtering, in the stationary case, the Wiener filter [1], [2] represents the optimum solution both in terms of mean square and least squares errors when the signal statistics are known [3], [4]. While originally formulated for the continuous time case, the results also apply to discrete time signals and their statistics [5], [6]. The Wiener filter is required to linearly combine an input to match a desired signal in the mean squared error sense [7], [8]. In the discrete domain the Wiener filter is given by the Wiener-Hopf equation $\mathbf{w}_{\rm o} = \mathbf{R}^{-1}\mathbf{p}$. Here $\mathbf{R}$ is the — assumed to be full rank — covariance matrix of the input and $\mathbf{p}$ a vector of cross-correlations between the input and desired signal. This ideal filter then presents a limit to which practical solutions should converge [9], [10], such as least mean squares and recursive least squares algorithms [7], [8].

If the signal statistics are not known or only partially available, it is tempting to calculate an approximation of $\mathbf{w}_{\rm o}$ via the Wiener-Hopf equation using estimates of $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$, or from assumed statistics. After all, the Wiener-Hopf equation applied to estimates is, for example, the starting point of the recursive least squares algorithm [7], [8]. While such estimates for $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ may be poor, the result for $\mathbf{w}_{\rm o}$ can be surpringly accurate. In contrast, if the input signal is uncorrelated Gaussian with unit variance, then the cross-correlation vector $\mathbf{p}$ also converges to $\mathbf{w}_{\rm o}$ but the accuracy increases only with $1/N$, where N is the sample size on which the best linear unbiased estimator operates.

In this paper, we aim to explore the above discrepancy in precision. We therefore want to firstly explore

- (Case 1) why the Wiener-Hopf solution $\hat{\mathbf{R}}^{-1}\hat{\mathbf{p}}$ can be accurate despite inaccurate statistical estimates.

Since in excellent signals and systems text books such as [5], [6], the Wiener solution is rightfully based on the expected, accurate statistics, it may even be tempting to ask if it is sensible to separately estimate $\mathbf{R}$ and $\mathbf{p}$ from different data sets, or to include knowledge, of say $\mathbf{R}$, into the estimate. Therefore, we will address two further questions of

- (Case 2) whether it is advantageous to use precise knowledge of $\mathbf{R}$ instead of its estimate $\hat{\mathbf{R}}$; and
- (Case 3) whether $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$ can be obtained from different data sets.

In order explore and answer these questions, we will interprete the Wiener-Hopf solution as a least squares problem.

In the following, Sec. II reviews the system identification problem as one application of the Wiener-Hopf solution, and states some useful metrices for the identification task. The estimation of the second order statistics required for the Wiener-Hopf solution is outlined in Sec.III. A least squares interpretation in Sec. IV leads an analysis of the variance of errors in Sec. V for the different cases above. These theoretical results are compared to experiments in Sec. VI, followed by conclusions in Sec. VII.

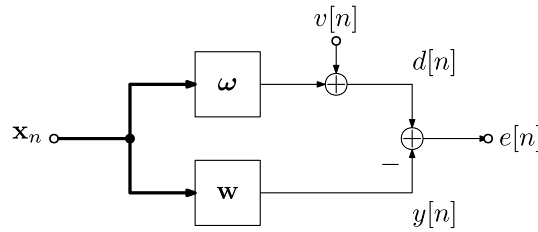

Fig. 1. Model for adaptive system identification.

II. System Identification and Wiener-Hopf Solution

A. System Identification Setup

A standard challenge in adaptive filtering [7], [8] is the system identification problem outlined in Fig. 1. A zero-mean complex symmetric Gaussian input $\mathbf{x}_n \in \mathbb{C}^M$ excites an unknown system $\boldsymbol{\omega} \in \mathbb{C}^M$, generating an output $d[n] = \boldsymbol{\omega}^{\rm H}\mathbf{x}_n + v[n]$, where $v[n]$ is observation noise that is a zero-mean uncorrelated complex symmetric Gaussian process that is assumed to be statistically independent from $\mathbf{x}_n$. An adaptive filter, $\mathbf{w} \in \mathbb{C}^M$, which we assume to be of the same length M as the unknown system, produces an output $y[n] = \mathbf{w}^{\rm H}\mathbf{x}_n$. The aim of the adaptive filter is to adjust $\mathbf{w}$ such that the error $e[n] = d[n] - y[n]$ is minimised in the mean squared error sense.

B. Wiener-Hopf Solution

The problem of minimising the mean square error $e[n]$ in Fig. 1 is formulated as

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

where $\mathcal{E}\{\cdot\}$ is the expectation operator. With the error $e[n] = d[n] - \mathbf{w}^\mathrm{H}\mathbf{x}_n$, the mean square error depends on the covariance matrix of the input, $\mathbf{R} = \mathcal{E}\{\mathbf{x}_n\mathbf{x}_n^\mathrm{H}\}$, and the cross-correlation vector between the desired signal and the input, $\boldsymbol{p} = \mathcal{E}\{d^*[n]\mathbf{x}_n\}$. If $\mathbf{R}$ is full rank, then the so called Wiener-Hopf equation [7], [8]

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

solves the problem in (1).

C. System Identification Metrics

In order to assess the accuracy of system identification, we define two metrics. First, $\boldsymbol{\epsilon} = \mathbf{w} - \boldsymbol{\omega}$ measures the mismatch between the unknown system and its model. We define its mean square value ξ as

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

which in the optimum case $\mathbf{w} = \mathbf{w}_\mathrm{o}$ will be zero.

As a second metric, the mean square error χ,

$$\chi = \mathcal{E}\{|e[n]]|^2\} = \mathcal{E}\{|d[n] - \mathbf{w}^\mathrm{H}\mathbf{x}_n|^2\} \tag{4}$$

directly represents the cost to be minimised in (1). It can be related to the mean square system mismatch via

$$\begin{aligned}\chi &= \mathcal{E}\{v[n] + \boldsymbol{\omega}^\mathrm{H}\mathbf{x}_n - \mathbf{w}^\mathrm{H}\mathbf{x}_n|^2\} && (5)\\ &= \sigma_v^2 + \mathcal{E}\{\boldsymbol{\epsilon}^\mathrm{H}\mathbf{R}\boldsymbol{\epsilon}\} \ . && (6)\end{aligned}$$

Note that (6) minimises a weighted inner product of $\boldsymbol{\epsilon}$ compared to ξ in (3).

III. Estimation of Covariance Matrix and Cross-Correlation Vector

A. Auto- and Crosscorrelation Estimators

In practice, the covariance matrix and the cross-correlation vector in (2) must be estimated from finite data, say N snapshots $n = 0, \ldots, (N-1)$. As unbiased estimators of $\mathbf{R}$ and $\mathbf{p}$, we utilise

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{X}\mathbf{X}^\mathrm{H} \tag{7}$$

$$\hat{\mathbf{p}} = \frac{1}{N}\mathbf{X}\mathbf{d} \tag{8}$$

where

$$\mathbf{X} = [\mathbf{x}_0,\ \mathbf{x}_1,\ \ldots,\ \mathbf{x}_{N-1}] \tag{9}$$

$$\mathbf{d} = [d[0],\ d[1],\ \ldots,\ d[N-1]]^\mathrm{H} \ . \tag{10}$$

It is easy to show that $\mathcal{E}\{\hat{\mathbf{R}}\} = \mathbf{R}$ and $\mathcal{E}\{\hat{\mathbf{p}}\} = p$.

To analyse the variance of the above estimators, for simplicity we assume that data is temporally uncorrelated. Defining $\hat{r}_{su} = \mathbf{u}^\mathrm{H}\mathbf{s}/N$ as the estimate of the cross-correlation r_{su} between two arbitrary signals $s[n]$ and $u[n]$ based on N samples in vectors $\mathbf{s}$ and $\mathbf{u}$, we can simplify results from e.g. [11]–[14] to obtain

$$\mathrm{var}\{\hat{r}_{su}\} = \frac{1}{N}r_{ss}r_{uu} \tag{11}$$

for the variance of the estimator. Therefore, with $\mathbf{r} \in \mathbb{R}^M$ holding the variances of the elements of $\mathbf{x}_n$, we obtain

$$\mathrm{var}\{\hat{\mathbf{R}}\} = \frac{1}{N}\mathbf{r}\mathbf{r}^\mathrm{H} \ , \tag{12}$$

$$\mathrm{var}\{\hat{\mathbf{p}}\} = \frac{1}{N}\mathbf{r}\sigma_d^2 \ . \tag{13}$$

Thus the variance of the estimates reduces like $1/N$ as the sample size N increases, but also depends on the power of the signals involved. We illustrate the impact on the Wiener-Hopf solution based on estimates in the following example.

Example 1: We want to identify the system $\omega^\mathrm{H} = [1, 0.5, .25]^\mathrm{H}$ with $M = 3$ using an uncorrelated sequence of input vectors $\mathbf{x}_n$ with $\mathbf{R} = \mathbf{I}$. We have $N = 100$ snapshots, and no observation noise, i.e. $\sigma_v^2 = 0$. For one particular instance of data, we measure $\|\mathbf{R} - \hat{\mathbf{R}}\|_\mathrm{F}^2 = 0.1235$ and $\|\mathbf{p} - \hat{\mathbf{p}}\|_2^2 = 0.0154$. In comparison, the expected variances are $\mathcal{E}\{\|\mathbf{R} - \hat{\mathbf{R}}\|_\mathrm{F}^2\} = \frac{M^2}{N} = 0.09$ and $\mathcal{E}\{\|\mathbf{p} - \hat{\mathbf{p}}\|_2^2\} = \frac{M}{N}\|\boldsymbol{\omega}\|_2^2 = 0.0413$ based on (12) and (13). Despite these inaccuracies of the estimates, for the Wiener-Hopf solution, $\hat{\mathbf{w}}_\mathrm{o} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{p}}$, we measure a system mismatch of only $\|\hat{\mathbf{w}}_\mathrm{o} - \boldsymbol{\omega}\|_2^2 < 0.5 \cdot 10^{-30}$. △

From Example 1, we take away two observations:

1) Despite poor statistical estimates, the Wiener-Hopf solution is very accurate and estimation errors appear to cancel.
2) For $\mathbf{R} = \mathbf{I}$, we know that $\mathbf{p} = \boldsymbol{\omega}$, but clearly the cross-correlation estimate $\hat{\mathbf{p}}$ represents a much poorer approximation of the system than the Wiener-Hopf solution.

IV. Least Squares Interpretation

A. Preliminary Considerations

The observations following Example 1 suggest the investigation of the three cases raised in Sec. I: (Case 1) the calculation based on estimates as in (12) and (13); (Case 2) the inclusion of knowledge about the precise covariance matrix $\mathbf{R}$; and (Case 3) the use of two different data sets to determine $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$. For the analysis, we will be utilising a least squares interpretation of the Wiener-Hopf solution; note that for Case 1 with (7) and (8), we have $\hat{\mathbf{w}}_\mathrm{o} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{p}} = (\mathbf{X}\mathbf{X}^\mathrm{H})^{-1}\mathbf{X}\mathbf{d} = (\mathbf{X}^\mathrm{H})^\dagger\mathbf{d}$, where $\{\cdot\}^\dagger$ denotes the pseudo-inverse as the classic minimum norm solution to a least squares problem [15]. As it is the most general scenario, we first address Case 3, and then derive from its analysis the results for Cases 2 and 1.

B. Case of Two Different Data Sets

Assume that the estimates are taken from two different data sets $\{\mathbf{X}_1\}$ and $\{\mathbf{X}_2, \mathbf{d}\}$, where $\mathbf{d} = \mathbf{X}_2\boldsymbol{\omega} + \mathbf{v}$ with $\mathbf{v}^\mathrm{T} = [v[0], \ldots, v[N-1]]$. Thus,

$$\hat{\mathbf{R}} = \mathbf{X}_1\mathbf{X}_1^\mathrm{H} \tag{14}$$

$$\hat{\mathbf{p}} = \mathbf{X}_2\mathbf{d} = \mathbf{X}_2\mathbf{X}_2^\mathrm{H}\boldsymbol{\omega} + \mathbf{X}_2\mathbf{v} \ . \tag{15}$$

Therefore for the Wiener-Hopf solution, we have

$$\hat{\mathbf{w}}_{\mathrm{o}} = \left(\mathbf{X}_1^{\mathrm{H}}\mathbf{X}_1\right)^{-1}\mathbf{X}_2^{\mathrm{H}}\mathbf{X}_2\boldsymbol{\omega} + \left(\mathbf{X}_1^{\mathrm{H}}\mathbf{X}_1\right)^{-1}\mathbf{X}_2^{\mathrm{H}}\mathbf{v} \, . \tag{16}$$

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

$$\boldsymbol{\epsilon} = \left\{\left(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)^{-1}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}} - \mathbf{I}\right\}\boldsymbol{\omega} + \left(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)^{-1}\mathbf{X}_2\mathbf{v} \, ,$$

Expanding the first term on the r.h.s.,

$$\begin{aligned}\boldsymbol{\epsilon} = &\left(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)^{-1}\left(\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}} - \mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)\boldsymbol{\omega} \\ &+ \left(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}\right)^{-1}\mathbf{X}_2\mathbf{v} \, ,\end{aligned} \tag{17}$$

we find a perturbation term due to the difference in covariance estimates, $(\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}} - \mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})$, and another one due to the observation noise $\mathbf{v}$.

C. Case of Incorporating Knowledge of Input Covariance

We assume knowledge of $\mathbf{R}$, and replace $\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}$ in (16) and (17) by $N\mathbf{R}$. Thus, for the Wiener-Hopf solution

$$\hat{\mathbf{w}}_{\mathrm{o}} = \frac{1}{N}\left(\mathbf{R}^{-1}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\boldsymbol{\omega} + \mathbf{R}^{-1}\mathbf{X}_2\mathbf{v}\right) \, , \tag{18}$$

and for the system mismatch we find

$$\boldsymbol{\epsilon} = \left\{\frac{1}{N}\mathbf{X}_2^{\mathrm{H}}\mathbf{X}_2 - \mathbf{I}\right\}\boldsymbol{\omega} + \frac{1}{N}\mathbf{X}_2^{\mathrm{H}}\mathbf{v} \, . \tag{19}$$

The mismatch in (19) still depends on an error in the covariance estimate, and on a noise-related term.

D. Case of Consistent Data

Assuming $\mathbf{X}_2 = \mathbf{X}_1 = \mathbf{X}$, we have for the Wiener-Hopf solution

$$\mathbf{w}_{\mathrm{o}} = \boldsymbol{\omega} + (\mathbf{X}\mathbf{X}^{\mathrm{H}})^{-1}\mathbf{X}\mathbf{v} \, . \tag{20}$$

Therefore the system mismatch simply becomes

$$\boldsymbol{\epsilon} = (\mathbf{X}^{\mathrm{H}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{H}}\mathbf{v} \, . \tag{21}$$

Note that $\boldsymbol{\epsilon}$ now only has an observation-noise related term. Further, (21) uses the pseudo-inverse $\boldsymbol{\epsilon} = (\mathbf{X}^{\mathrm{H}})^{\dagger}\mathbf{v}$, and thus provides the smallest possible system mismatch in the least squares sense, given $\mathbf{X}$ and $\mathbf{v}$.

V. Mean Square System Mismatch Analysis

A. Mean Square System Mismatch

Exploiting the independence of the observation noise $\mathbf{v}$ from all other terms, for the mean square system mismatch with two data sets, we find that for $\xi = \mathcal{E}\{\boldsymbol{\epsilon}^{\mathrm{H}}\boldsymbol{\epsilon}\}$,

$$\xi = \xi_x + \xi_v \, , \tag{22}$$

where ξ_x is a data-related term, which by using the trace operator $\mathrm{tr}\{\cdot\}$ and trace rules, can be brought into the form

$$\begin{aligned}\xi_x = \mathrm{tr}\,\{\mathbf{W}\,(\mathbf{I} + &\mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-2}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\} \\ - \mathcal{E}\{(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}&\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\} - \mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\} \, ,\end{aligned} \tag{23}$$

with $\mathbf{W} = \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{H}}$. The second term ξ_v in (22) can be resolved as

$$\xi_v = \sigma_v^2\mathrm{tr}\{\mathcal{E}\{\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\}\} \tag{24}$$

We investigate these terms in turn.

B. Observation Noise-Related Variance of System Mismatch

Assuming that $\mathbf{X}_1$ has full row rank, and that $\mathcal{E}\{(\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}})^{-1}\} = \mathbf{R}^{-1}$, we can resolving the 4th order cumulant in (24) by Isserlis' theorem [16]. Further noting that complementary correlations of the type $\mathcal{E}\{\mathbf{X}_1\mathbf{X}_1^{\mathrm{T}}\} = \mathbf{0}$ [17], we can obtain

$$\xi_v = \frac{\sigma_v^2}{N}\mathrm{tr}\{\mathbf{R}^{-1}\} \tag{25}$$

irrespective which of three cases we are investigating.

C. Data Mismatch-Related Variance of System Mismatch

The data-related term ξ_x from (22) is only non-zero for Cases 2 and 3 in Secs. IV-C and IV-B, respectively. Case 3 for separated data sets is difficult to analyse, but note the mismatch (17) contains two noisy estimates compared to the one in (19). For Case 2 with the replacement of the term $\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}}$ by $N\mathbf{R}$, (23) simplifies to

$$\xi_x = \frac{1}{N^2}\mathrm{tr}\{\left(\mathcal{E}\{\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\mathbf{R}^{-2}\mathbf{X}_2\mathbf{X}_2^{\mathrm{H}}\} - \mathbf{I}\right)\mathbf{W}\} \, . \tag{26}$$

Using again Isserlis' theorem, the cancellation of complementary correlations, and trace rules, after some rearrangements

$$\xi_x = \frac{1}{N}\mathrm{tr}\{\mathbf{R}^{-1}\}\,\mathrm{tr}\{\mathbf{R}\mathbf{W}\} \tag{27}$$

emerges for Case 2, where the input covariance matrix is known to be $\mathbf{R}$ and embedded in the Wiener-Hopf solution.

VI. Simulations and Results

Below, simulations explore the system mismatch of the three cases for both uncorrelated and correlated inputs, and compared the results to the theoretical values of Sec. V.

A. Uncorrelated Input

For an unknown system $\boldsymbol{\omega} \in \mathbb{C}^5$, we calculate the Wiener-Hopf solution for different signal to noise ratios, defined as $1/\sigma_v^2$, between -30 dB and 40 dB, and three different samples sizes $N \in \{30, 300, 3000\}$ with filter length $M = 5$. An ensemble of 1000 experiments is performed for each parameter combination and for the three different cases of (Case 1) determining the statistics from a single data set, (Case 2) embedding knowledge of $\mathbf{R} = \mathbf{I}$ into the solution, and (Case 3) using separate data sets for the estimation of $\hat{\mathbf{R}}$ and $\hat{\mathbf{p}}$.

The ensemble mean square system mismatch $\hat{\xi}$, i.e. the averaged value across the experiments, is presented in Fig. 2. For Case 1, experimental values closely follow the theoretical curves of ξ_v according to (25). For Case 2, $\hat{\xi}$ for low SNR also follows the limits of ξ_v, while for higher SNRs above 0 dB starts to levels out at ξ_x, determined according to (27). For Case 3, we do not have a theoretical expression for ξ_x, but for higher values of the sample size N, the curves level out approximately 3 dB above those of Case 2, suggesting that the inclusion of two noisy estimates rather than a single noisy estimate approximately doubles the mean square system mismatch. Note that for uncorrelated input with $\mathbf{R} = \mathbf{I}$, minimising the mean square system mismatch ξ in (3) is equivalent to minimising the mean square error χ in (6), which is targeted by the Wiener-Hopf solution.

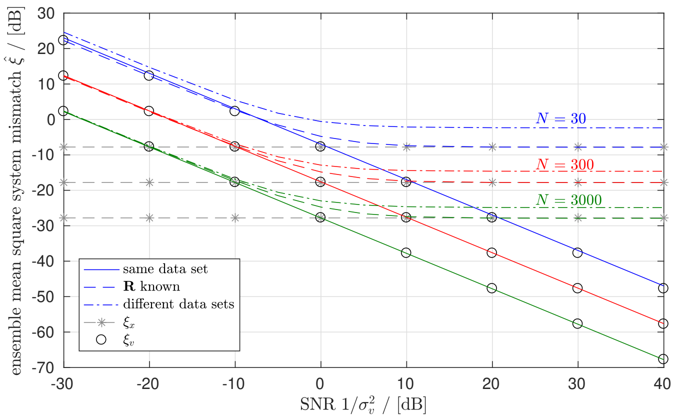

Fig. 2. Ensemble mean squared system mismatch $\hat{\xi}$ versus signal to noise ratio $1/\sigma_v^2$, the sample size N, and the realisation of the Wiener-Hopf solution via the three considered cases for an uncorrelated input. Also shown are the theoretical curves for ξ_v (an assympote for all cases) and ξ_x (an assymptote for Case 2).

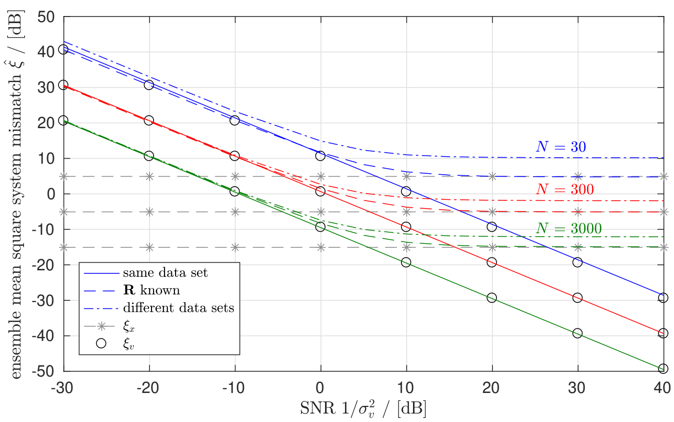

Fig. 3. Ensemble mean squared system mismatch $\hat{\xi}$ versus SNR $1/\sigma_v^2$ with parameters as in Fig. 2 but for correlated input.

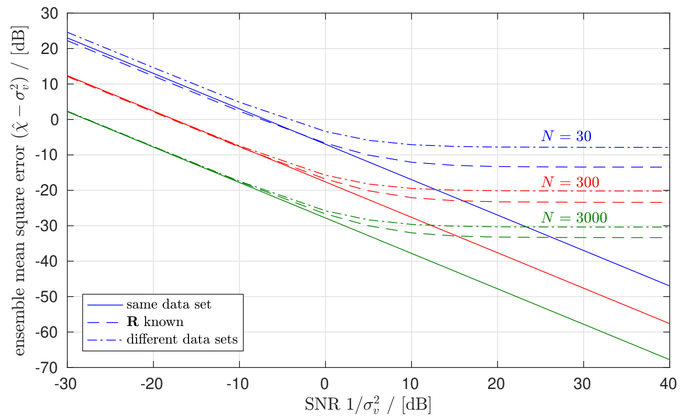

Fig. 4. Ensemble mean squared error $\hat{\chi}$ versus SNR $1/\sigma_v^2$ for the experiment with correlated input.

B. Correlated Input

To explore the impact of a correlated input, data $\mathbf{x}_n$ is generated from $\mathbf{u}_n \in \mathcal{CN}(\underline{0}, \mathbf{I})$ via $\mathbf{x}_n = \mathbf{A}\mathbf{u}_n$, where the innovation filter $\mathbf{A}$ possess singular values $\sigma_m = 2^{-(m-1)}$, $m = 1, \ldots, M$. Therefore, the covariance matrix is $\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathrm{H}}$ with eigenvalues $\lambda_m = 2^{-2(m-1)}$, $m = 1, \ldots, M$. The remaining parameters are the same as for Sec. VI-A.

Fig. 3 shows the ensemble results, where again performance increases as the SNR increases, and levels out for Cases 2 and 3. For all cases, the assumpotes for $\sigma_v^2 \ll 1$ are given by ξ_v. Again for Case 2, for $\sigma_v^2 \gg 1$ the mean square system mismatch levels out at the assympotes determined by (27). For Case 3, where we have not derived theoretical limits, for larger values of N performance degrades by approximately 3 dB or a factor of two due to doubling of estimation errors via both $\mathbf{X}_1$ and $\mathbf{X}_2$ in the case of two separate data sets. Generally error levels are significantly increased in comparison to Fig. 2 due to the term $\mathrm{tr}\{\mathbf{R}^{-1}\} = \sum_m \lambda_m^{-1} = \sum_{m=0}^{4} 4^m = 341$ in both (25) and (27).

Since the Wiener-Hopf solution aims to minimise the mean squared error χ, with $\mathbf{R} \neq \mathbf{I}$ this cost function now differs from the mean square system mismatch. For a better assessment how close the experimental MSE $\hat{\chi}$ approaches its minimum value, we consider $(\hat{\chi} - \sigma_v^2)$ which, different from χ in (6), can tend to zero for $\sigma_v^2 \neq 0$. Note that the metric $(\hat{\chi} - \sigma_v^2)$ represents an inner product of ϵ weighted by the covariance matrix $\mathbf{R}$, and hence, if not well-conditioned, $(\hat{\chi} - \sigma_v^2)$ might have behaved differently from ξ. Nonetheless, while no theoretical expression for χ have been derived, the ensemble statistics in Fig. 4 indicate that the relative performance for the cases and parameter sets compared to the ensemble mean square system mismatch Fig. 2 remains unaltered.

VII. Discussion and Conclusions

We have investigated the ability of the Wiener-Hopf solution to reach fairly accurate results from imprecise estimates. This is due to the fact that Wiener-Hopf can be recast as a least squares problem, whose answer, as long as the measurement matrix is full rank, will only be impeded by observation noise and the conditioning of the covariance matrix, but not by the precision of auto- or cross-correlation estimates. In principle therefore, short sequences suffice for estimation, whereby near-orthonality of input data vectors — and hence good conditioning or even near-diagonality of the covariance estimates — is advantageous, motivating e.g. 'perfect sequences' as input for system identification purposes [18]–[20]. This answers Case 1.

If one might be tempted to draw estimates from different data sets, i.e. to measure $\mathbf{R}$ and $\mathbf{p}$ from separate, uncorrelated data, or to embed knowledge about the scenario into the solution in case $\mathbf{R}$ is known, then such attempts result generally in a worse performance compared to Case 1. This is particularly at high SNR, where the performance now becomes limited by estimation errors that only tail off with $1/N$, where N is the sample size of the estimates. Case 3, where estimation errors from two different data sets aggregate, is thereby worse than embedding precise knowledge of $\mathbf{R}$ as part of Case 2.

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] B. Girod, R. Rabenstein, and A. Stenger, *Signals and Systems*. Chichester: J. Wiley & Sons, 2001.

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

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

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

[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] 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.

[12] 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.

[13] 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, May 2019, pp. 8033–8037.

[14] 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*, 2025, to appears.

[15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, Maryland: John Hopkins University Press, 1996.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.