Efficient Estimation and Tracking of Response Statistics for (Randomly) Time-Variant Systems

A Dissertation Presented

by

Lang Peng

to

Electrical and Computer Engineering Department

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the field of

Electrical and Computer Engineering

Northeastern University
Boston, Massachusetts

April 2014
To my family, and Buddha
Abstract

Efficient Estimation and Tracking of Response Statistics
for (Randomly) Time-Variant Systems by

Lang Peng

Doctor of Philosophy in Electrical and Computer Engineering
Northeastern University, April 2014
Hanoch Lev-Ari, Adviser

Time-variant systems are often encountered in engineering applications, ranging from underwater acoustic communications to array processing. The inability of conventional (i.e., stationary-based) system identification techniques to cope with rapid time variation of the system’s impulse response has motivated the development of novel approaches to identify linear, arbitrarily time-variant systems. Most of these techniques rely on a known statistical characterization of the time-variation of the system’s response.

We present a technique for determining the autocorrelation of an arbitrarily time-variant system, based on estimation of the first and second order moments of the system’s randomly time-varying impulse response. Our approach relies on a (known) key relation between the system response autocorrelation function and certain 2nd and 4th order moments of the system input and (noisy) output signals, with no other prior information about the dynamics of the system response required. We introduce a Wiener identification problem (WIP) interpretation of this key relation, which enables us to benefit from the wealth of existing results about the dynamics and performance of standard adaptive filters. In particular, we develop efficient
time-recursive alternatives to the existing non-recursive procedure for estimating the system’s response autocorrelation. In addition, we exploit the structure and sparsity of the covariance matrix associated with our WIP interpretation to develop lag-recursive solutions that achieve a further reduction of the overall computational cost. In particular, we explore the possibility of using the maximum entropy method (MEM) to obtain a computationally efficient direct estimate of the system’s (multi-channel) Doppler-spread spectrum.
Acknowledgements

Destiny ties people together.

Every encounter is a lot.

Thus, I deeply acknowledge those who spend a lot of time with me,
especially my family, my advisor, and my classmates.
## Contents

1 Introduction ................................................................. 1
   1.1 Characterization of time varying channels ....................... 4
   1.2 Estimation of $W(t)$ .................................................. 8
   1.3 Estimation of $C_W(m)$ ............................................. 11
   1.4 Prior-information-free estimation of $C_W(m)$ .................... 13

2 Estimating the Response Autocorrelation of a Linear Time Varying System 16
   2.1 Prior-info-free estimation of $C_W(m)$ ......................... 17
   2.2 Uncorrelated scattering: diagonal $C_W(m)$ .................... 21
   2.3 Wiener (MMSE) problem interpretation .......................... 24
   2.4 The linear regression assumption ................................. 28
   2.5 Computational cost ................................................ 32

3 Time Recursive Implementation ........................................ 34
   3.1 Time-recursive implementation ..................................... 35
   3.2 Performance Analysis ............................................... 43
      3.2.1 Stationary (=Static) case .................................... 45
      3.2.2 Dynamic case ................................................ 46

4 Lag Recursive Implementation ......................................... 47
   4.1 Lag-recursive implementation ..................................... 48
      4.1.1 Inter-block shift structure ................................. 48
List of Figures

1.1 Doppler spectrum for the Jakes model ................................................. 8
1.2 System set up ....................................................................................... 9
1.3 Optimized moments estimation ............................................................. 11

3.1 WIP interpretation .............................................................................. 35
3.2 Time-recursive estimation of $E \Psi_m(\cdot)$ ......................................... 36
3.3 Time-recursive estimation of $\langle ED_m(t) \rangle_t$ .................................. 36
3.4 LMS Learning curve for $c_w(0)$ for three different $\mu$ values ............... 38
3.5 Tradeoff curve for LMS: $D_S$ vs. the time constant for $c_w(0)$ .............. 38
3.6 RLS Learning curve for $c_w(0)$ for three different $\lambda$ values ............ 40
3.7 Tradeoff curve for RLS: $D_S$ vs. the time-constant for $c_w(0)$ .............. 40
3.8 NLMS simulation learning curve of $c_w(0)$ with time varying step size algorithm 1 42
3.9 NLMS simulation learning curve of $c_w(0)$ with time varying step size algorithm 2 42

4.1 Illustration of inter-block shift structure of $\mathcal{R}_U(m - 1)$ (left) and $\mathcal{R}_U(m)$ (right): the overlap between these two matrices is shown in dark blue, while the non-overlapping portions are shown in grey and light blue. ......................................................... 50
4.2 Illustration of intra-block shift structure of $\mathcal{R}_U(m - 1)$ (left) and $\mathcal{R}_U(m)$ (right): yellow area covers the same elements while deep blue covers the different elements. ................................................................. 54
4.3 Illustration of combination of inter- and intra-block shift structure of $\mathcal{R}_U(m-1)$ (left) and $\mathcal{R}_U(m)$ (right): yellow area are elements shared by two matrix and blue area are elements need to be estimated. .................................................. 54

4.4 Power spectrum estimation comparison: FFT vs. ME ........................................ 57

6.1 spectrum illustration of $S_w(f)$ and $S_u(f)$ .................................................. 66

C.1 Bandpass to baseband conversion ................................................................. 84
List of Tables

1.1 Computational cost of fundamental equation .......................... 14
2.1 Computational cost of static (non-iterative) solution ............... 33
3.1 Steady state performance of LMS and RLS .............................. 45
3.2 Tracking performance of LMS and RLS ................................. 45
Chapter 1

Introduction

Time varying systems, both natural and man-made, are all around us. Examples include astronomical and cosmic systems [1, 2], geophysics systems [3, 4], earth’s climate and weather system [5], ionosphere communication channels [6, 7], wireless channels [8, 9], underwater acoustic channels [10, 11], biological systems [12, 13], vehicle systems [14, 15, 16], and manufacturing plants [17, 18]. The study and characterization of linear time variant systems (LTVS) has attracted increasing attention in the past half century. In particular, numerous techniques have been proposed for extracting information about the system response from the measurements of a system’s input and output signals.

The input-output relation of a causal discrete-time LTVS with input $u(t)$ and output $y(t)$ is 

$$y(t) = \sum_{k=0}^{\infty} w_k(t)u(t - k),$$

so that the sequence of time varying response coefficients $W(t) = [w_0(t) w_1(t) w_2(t) \cdots]$ provides a complete characterization of the dynamic behavior of the LTVS. Some control and signal processing applications rely on a statistical characterization of the underlying LTVS, such as the autocorrelation of $W(t)$. Examples include blind equalization[9], adaptive power control [19], adaptive beamforming [20], cellular system deployment [19], and blind deconvolution [21]. The LTVS response $W(t)$ is often modeled as an ergodic stationary process whose autocorrelation or spectrum are used to customize the design of control and signal processing algorithms.

Numerous other engineering applications rely on the specific system response $W(t)$ to op-
timize their performance. Examples include system identification [22], linear predictive coding of speech signals [22, 23], adaptive noise cancellation [9, 24], outlier elimination [22, 25, 26], spectral analysis [22, 27], adaptive control [28, 16], moving target tracking [29], and random parameter tracking [30]. In this case, an adaptive algorithm is used to determine and track the instantaneous values of the response coefficients \{w_k(t)\}.

When the time variation of \{w_k(t)\} is relatively slow, conventional adaptive algorithms, such as LMS or RLS [21, 18], can be used to obtain estimates of the instantaneous \{w_k(t)\}. However, these classical algorithms perform poorly in the presence of a rapidly varying response \(W(t)\). Several improvements have been proposed to cope with rapid time variation, often relying on statistical information about the LTVS, such as:

- Optimized averaging [31], which uses the autocorrelation of \(W(t)\) to optimize the construction of an estimate for \(W(t)\)
- Extended RLS algorithm [21, 32], which relies on a state space model for \(W(t)\), combined with the well-known Kalman filter, to construct an RLS-like algorithm for estimation of a rapidly varying \(W(t)\).

In both cases one has to rely on statistical information about the dynamics of the response \(W(t)\): optimized averaging makes direct use of the autocorrelation of \(W(t)\), while the extended RLS algorithm does so indirectly, by relying on the equivalent characterization in terms of a state space model for \(W(t)\).

The autocorrelation and higher order moments of an observed fluctuating signal usually can be estimated by time averaging. However, when the signal is the impulse response of a time varying system, it is not directly observable: usually one can only measure the system’s input and output signals. Nevertheless, the LTVS input-output relation \(y(t) = \sum_k w_k(t)u(t - k)\) induces relations between the statistics of the input and output signals and the statistics of the time variant impulse response \(W(t)\). Such relations can be exploited, for instance, to estimate the autocorrelation of \(W(t)\) [33, 34]. The average autocorrelation \(C_W(m)\) of the impulse response \(W(t)\) satisfies a set of linear matrix equations that involve certain 2nd and
4th order moments of the observed signals \( y(t) \) and \( u(t) \).

For wireless fading channels, WSSUS (wide sense stationary uncorrelated scatterers) is normally assumed. This means that the autocorrelation function 
\[
    r_{kl}(t, s) \equiv E\{w_k(t)w_l^*(s)\}
\]
depends only on the time difference \( t - s = m \) and \( E\{w_k(t)w_l(s)\} = 0 \) for every \( k \neq l \) and all \( (t, s) \), i.e., different propagation paths. In this case, the autocorrelation matrix 
\[
    C_W(m) = [r_{kl}(m)]_{k,l=0}^M
\]
is diagonal because its off-diagonal entries vanish. It is often assumed, in addition, that all paths share the same distribution, so that 
\[
    r_{kk}(m) = r_{00}(m) \quad \text{independent of} \quad k,
\]
so that 
\[
    C_W(m) = r_{00}(m)I.
\] The well known Jake’s model for fading mobile communication channels [9, 35] is a specific WSSUS example with 
\[
    r_{00}(m) = J_0(w_Dm)
\]
where \( w_D \) is the discrete-time version of the normalized maximum Doppler frequency of the channel. However, in other cases such as the underwater acoustic communication channel, the WSSUS assumption usually does not hold. Consequently we will not make any a priori assumptions about the structure of the average autocorrelation matrix \( C_W(m) \). An accurate and reliable estimate of \( C_W(m) \) will allow us to: (1) determine channel capacity [36]; (2) implement adaptive power control [37]; (3) optimize the detection scheme [35]; (4) improve channel equalization with further availability of \( W(t) [9]; (5) improve the coding scheme [9]; and (6) determine channel’s Doppler spectrum.

A prior-information-free approach for estimation of the autocorrelation of the time variant system response \( W(t) \) was introduced in [33, 38], based on a novel explicit relation between the autocorrelation of \( W(t) \) and certain 2nd and 4th order moments of the signals \( u(t) \) and \( d(t) \), a noisy version of the output signal \( y(t) \). Such moments can be estimated via long-term averaging directly from measurements of the system input and output signals. This results in a non-recursive (off-line) solution that requires a signal record of considerable length. The same (non-recursive) solution was also obtained in [39] via an approximate modeling argument that led to a deterministic least squares problems, whose solution provides an estimate for the autocorrelation of \( W(t) \).

In this dissertation we develop efficient (adaptive) alternatives to the static (non-recursive) procedures for estimation of \( C_W(m) \) that were proposed in [39, 33]. Our approach relies on a new interpretation of the relation between \( u(\cdot), d(\cdot) \) and \( C_W(m) \) as a Wiener identification
problem (WIP), which enables us to exploit the well-developed theory of linear adaptive filters to obtain time-recursive estimates for the autocorrelation $C_W(m)$ (see Sec. 3.1), and to provide explicit performance metrics for such estimates (see Sec. 3.2). Our approach significantly reduces the length of the signal record required for a prescribed level of estimation error.

Furthermore, we exploit the structure of the covariance matrix associated with our WIP interpretation to develop time- and lag-recursive solutions that significantly reduce the overall computational cost (see Sec. 4.1). We also explore the possibility of using the maximum entropy method (MEM) to obtain a computationally efficient direct estimate of the Doppler-spread spectrum ($=$Fourier transform of $C_W(m)$) from $\{C_W(m); 0 \leq m \leq L\}$ for small values of $L$ (see Sec. 4.2).

Another opportunity for reducing computational cost is via exploring the sparsity of the impulse response $W(t)$ that occurs in many multipath propagation scenarios [37]. We demonstrate in Ch. 5 that such sparsity can be used to reduce the size of vectors and matrices in our Wiener problem interpretation, thus leading to a sparse version of our time- and lag-recursive solutions.

Finally, in Ch. 6 we provide a baseband version of our WIP interpretation, which can be used when the channel input signal $u(t)$ occupies a well-defined passband with a sufficiently narrow bandwidth.

1.1 Characterization of time varying channels

The starting point of much of the earlier research on identification of a linear time-variant system (LTVS) is a discrete-time FIR representation [33, 34, 35, 40], viz.

$$y(t) = \sum_{k=0}^{M-1} w_k(t)u(t-k) = W(t)U(t)$$

where $W(t) \triangleq [w_0(t) \; w_1(t) \; \cdots \; w_{M-1}(t)]$ is a row vector of length $M$ that represents the time-varying impulse response $\{w_k(t)\}$, $U(t) = [u(t) \; u(t-1) \; \cdots \; u(t-M+1)]^T$ is a column vector consisting of the most recent $M$ samples of the input signal sequence, and $y(t)$ is the noise-free
output signal. We shall assume in the sequel that the input signal $u(t)$ has stationary moments up to the 4th order. This is consistent, for instance, with all digital communication channels, in which the input signal is an i.i.d. sequence. In certain cases, as mentioned earlier, a statistical characterization of $W(t)$ is needed, typically in terms of its second-order moments. Since $W(t)$ is a vector that consists of several entries, to fully understand the characteristics of the system we need to obtain the autocorrelation of each entry as well as the cross-correlation between distinct entries.

In general, $W(t)$ may be modeled as a random process, so that we are led to define the (ensemble-averaged) autocorrelation of $W(t)$, viz.,

$$\Phi_W(t, s) = E\{W^*(s)W(t)\}$$

which is an $M \times M$ matrix function of both $t$ and $s$. However, in the overwhelming majority of applications this function has additional structure, which we classify here into three categories:

- **Stationary $W(t)$**: In many practical cases, such as certain short-distance wireless communications (e.g. bluetooth), and any time varying system when the time interval to be considered is short enough to assume that the (2nd order) statistics are constant [9], $W(t)$ can be modeled as a stationary random process. Thus

$$C_W(m) \triangleq \Phi_W(t + m, t)$$

which is only a function of the lag “$m$” and does not depend on $t$. The Fourier transform of $C_W(m)$ with respect to $m$ is the so-called Doppler spectrum of $W(\cdot)$[9]. The diagonal elements of $C_W(m)$ represent the autocorrelation of each individual element of $W(t)$, while the off-diagonal elements of $C_W(m)$ represent the crosscorrelation between two distinct elements of $W(t)$. The method developed in [33, 40] can estimate this time-invariant autocorrelation, using only measurements of the system input and output signals, as we describe in Sec. 1.3 and 1.4.
Asymptotically-mean-stationary $W(t)$: Some other applications, such as moment estimator construction in [31], rely only on knowledge of the average autocorrelation

$$C_W(m) \triangleq \langle \Phi_W(t + m,t) \rangle_t$$

(1.2)

where $\langle \rangle_t$ denotes long-term time averaging, namely,

$$\langle f(t) \rangle_t = \lim_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} f(t)$$

(1.3)

Convergence of the limit in (1.2) is guaranteed when $W(t)$ is asymptotically mean stationary [41] (see also App. B), which includes, in particular, the case when $W(t)$ is stationary. The average autocorrelation $C_W(m)$ can also be estimated by the method discussed in Sec. 1.3 and 1.4.

Slowly-varying non-stationary: In this case, $\Phi_W(t + m,t)$ is approximately constant (as a function of $t$) over short enough windows. A typical example is the underwater acoustic communication channel, which is usually modeled as

$$W(t) = EW(t) + \overline{W}(t)$$

where $\overline{W}(t)$ is a zero-mean stationary random process, and $EW(t)$ is a slowly-varying function of time [37]. Thus, in this case

$$\Phi_W(t + m,t) = [EW(t + m)]^* [EW(t)] + C_{\overline{W}}(m)$$

is a slowly varying function of the time-variable “$t$” for any given lag “$m$”. We propose in Chapter 3 a direct method, based on an extension of the (stationary) technique of [34], for estimation of this time-variant autocorrelation, using only measurements of the system input and output signals.

The most popular assumption about the time-variation of $W(t)$ is the WSSUS (wide-sense sta-
tionary with uncorrelated scatters) property, typically associated with wireless communication channels. In a (discrete-time) WSSUS channel, $W(t)$ is stationary and $w_k(t)$ is orthogonal to $w_l(s)$ for all $t, s$ and distinct $k, l$. The resulting corresponding $C_W(m)$ is diagonal. Hence a WSSUS channel has $M$ distinct Doppler spectra, one for each $0 \leq k \leq M - 1$. In fact, wireless channels often have the same frequency spectrum for each distinct delay path, i.e. system parameter $w_k(t)$.

The Fourier transform of $C_W(m)$ is called the Doppler spectrum of $W(t)$, viz.,

$$S_W(f) \triangleq \sum_{-\infty}^{\infty} \left[ C_W(m) - \mu_W^* \mu_W \right] e^{-j2\pi fm} \quad (1.4)$$

where $\mu_W$ is the non-vanishing average mean of $W(t)$, viz., $\mu_W \triangleq \langle EW(t) \rangle_t$ (see Appendix D for how to estimate it). Notice that this spectrum is the Fourier transform of the centered average autocorrelation

$$C_W(m) - \mu_W^* \mu_W = \langle E \{ [W(t) - \mu_W]^* [W(t + m) - \mu_W] \} \rangle_t$$

When only $\{C_W(m); 0 \leq m \leq L\}$ are available, the corresponding estimated spectrum is

$$\hat{S}_W(f) = \sum_{m=-L}^{L} \left[ \hat{C}_W(m) - \hat{\mu}_W^* \hat{\mu}_W \right] e^{-j2\pi fm} \quad (1.5)$$

In Chapter 4, we are going to present how to conduct this estimation of $S_W(f)$ in a cost efficient way.

The most widely used Doppler spectrum model is the Jakes model [42]: the autocorrelation function of a time-variant system associated with a moving vehicle is given by [35]

$$c_w(m) = \sigma_n^2 J_0(\omega_D m)$$

where $\sigma_n^2$ is a parameter describing the variance of a single channel coefficient, $J_0(\cdot)$ is the zero-order Bessel function of the first kind, and $\omega_D$ is a normalized Doppler frequency, defined
Figure 1.1: Doppler spectrum for the Jakes model

\[
\omega_D = \frac{2\pi}{F_s} \frac{v}{v_c} f_c
\]

where \( F_s \) is the sampling rate, \( v \) is the vehicle speed in meters per second \((m/s)\), \( v_c \) is the speed of light \((3 \times 10^8 m/s)\), and \( f_c \) is the carrier frequency, so that \( \omega_D \) is proportional to the vehicle speed \( v \). The Doppler power spectrum is the Fourier transform of \( r_{kk}(m) \), and for Jakes model, it can be expressed as [35, 42],

\[
s(\omega) = \begin{cases} 
\frac{2\sigma^2}{\omega_D^2} [1 - (\frac{\omega}{\omega_D})^2]^{-\frac{1}{2}} & |\omega| < \omega_D \\
0 & |\omega| > \omega_D 
\end{cases}
\]

Thus, both the Doppler power spectrum and the autocorrelation function are determined solely by \( \omega_D \). The graph of \( s(\omega) \) is shown in Figure 1.1.

1.2 Estimation of \( W(t) \)

Several techniques have been proposed for direct estimation of a time-variant system response
Figure 1.2: System set up

$W(t)$ from its input-output data [21, 32, 40]. We focus here on two techniques that explicitly rely on prior knowledge of the average autocorrelation $C_W(m)$. Both techniques assume the time-varying (FIR) input-output relation (Fig. 1.2)

$$d(t) = W(t)U(t) + v(t)$$ (1.6a)

where

$$W(t) = [w_0(t) w_1(t) w_2(t) \cdots w_{M-1}(t)]$$

and

$$U(t) = \begin{pmatrix} u(t) \\ u(t-1) \\ \vdots \\ u(t-M+1) \end{pmatrix}$$

In addition, both assume that $v(t)$ is independent of $u(t)$.

**Extended RLS algorithm**

The exponentially-weighted RLS algorithm can be interpreted as the Kalman filter associated with the state space model

$$W(t+1) = \lambda^{-1/2}W(t)$$

$$d(t) = W(t)U(t) + v(t)$$

where the (row) weight vector $W(t)$ plays the role of “state” and the output signal $d(t)$ plays the role of “measurement” [21, 32].
Similarly, the extended RLS [21] is obtained by constructing the Kalman filter for the state space model (with a row state vector $W(t)$)

$$
W(t + 1) = W(t)F + \eta(t)
$$

$$
d(t) = W(t)U(t) + v(t)
$$

Notice that now $W(t)$ is assumed to have a dynamic model with a known parameter matrix $F$. An even more general model for the state space equation was proposed in [40], viz.,

$$
W(t) = \sum_{i=1}^{P} W(t - i)F_i + \eta(t)
$$

The matrix $F$ in (1.7) can sometimes be obtained directly from the problem setup (see, e.g. [32]). However, in most applications this parameter matrix is not known a-priori. It can, nevertheless, be determined from an estimate of $C_W(m)$, as shown in [40].

**Optimized moment estimation**

The input-output relation (1.8) implies that the cross-correlation $R_{dU}(t) \triangleq E\{d(t)U^*(t)\}$ is directly related to $W(t)$, namely

$$
R_{dU}(t) = W(t)R_{UU}
$$

Notice that we rely here on two assumptions: (i) $U(t)$ is stationary, so that $R_{UU} = E\{U(t)U^*(t)\}$ is time-invariant, and (ii) $v(t)$ is uncorrelated with $U(t)$. Also, we treat here $W(t)$ as non-random, which is the standard approach in the adaptive filtering literature [21].

The constant covariance matrix $R_{UU}$ can be estimated by standard techniques, such as exponentially weighted time-averaging, viz.,

$$
\hat{R}_{UU}(t) = (1 - \lambda) \sum_{k=0}^{t} \lambda^k U(t - k)U^*(t - k)
$$

so that, for $1 - \lambda \ll 1$, this estimate is asymptotically efficient (i.e., it has a vanishing variance) and unbiased, namely $\lim_{t \to \infty} \hat{R}_{UU}(t) \approx R_{UU}$ in the mean square sense.
Thus the challenge of estimating $W(t)$ reduces to finding an estimate for the time-variant
cross-correlation $R_{dU}(t)$. This can be accomplished, for instance, by employing the moment
estimation technique of [31]. To this end we form the composite signal $d(t)U^*(t)$ and pass it
through a suitably selected \textit{averaging filter} $H(z)$ (see Figure 1.3). The optimal choice for $H(z)$,
as given in [31], depends on certain average autocorrelations of the composite signal $d(t)U^*(t)$.

\[ d(t)U^*(t) \rightarrow H(z) \rightarrow \hat{R}_{dU}(t) \]

Figure 1.3: Optimized moments estimation

A particularly convenient formulation is obtained if we modify the input to the averag-
ing filter to be $d(t)U^*(t)R_{UU}^{-1}$ because then the corresponding moment to be estimated is
$R_{dU}(t)R_{UU}^{-1} \equiv W(t)$, so that the output of the averaging filter is the estimate $\hat{W}(t)$. The
optimized choice for $H(z)$ relies on the two average scalar autocorrelations [34]

\[
C_B(m) = tr\{C_W(m)\} \\
C_V(m) = c_v(m)tr\{R_{UU}^{-2}C_U^*(m)\} + tr\{C_{\Omega}(m)C_W(m)\} 
\]

where $\Omega(t) = U(t)U^*(t)R_{UU}^{-1} - I$.

In summary, $C_W(m)$ is a necessary prerequisite for both extended RLS and the optimized
moment estimation technique. We provide in Sec. 1.3 further details about the estimation of
$C_W(m)$ from the input-output signals of a time-varying channel.

### 1.3 Estimation of $C_W(m)$

The input-output relation (see Figure 1.2)

\[ d(t) = W(t)U(t) + v(t) \quad (1.8) \]
connects the statistics of \( W(t) \) to those of \( U(t) \) and \( d(t) \). Since moments (2nd and 4th order) of the input \( U(t) \) and the output \( d(t) \) can be directly estimated from samples of these signals, it becomes possible to estimate \( C_W(m) \), the average autocorrelation of the system’s impulse response \( W(t) \).

In general, the autocorrelation (or higher order moments) of any stationary multichannel signal \( U(t) \) can be estimated via conventional (uniform) or exponential averaging as discussed in Sec. 1.4. The same holds true for average autocorrelations of asymptotically-mean-stationary (AMS) signals [41]. In particular, we can estimate \( C_U(m) \triangleq \mathbb{E}\{U(t + m)U^*(t)\} \), \( \Gamma_U(m) \triangleq \mathbb{E}\{[\tilde{U}(t + m) \otimes U(t)][\tilde{U}(t + m) \otimes U(t)]^*\} \) and the average autocorrelation \( c_d(m) \triangleq \langle \mathbb{E}\{d(t + m)d^*(t)\} \rangle_t \) of the AMS output signal \( d(t) \). Notice that here (and in the sequel) we assume that \( W(t) \) is an AMS process itself: this includes, as a special case, systems with deterministic \( W(t) \) (as assumed in [31, 34]). Since \( W(t) \) is not directly observed in our identification problem, it may seem that \( C_W(m) \equiv \Gamma_W(m) \), the average autocorrelation of \( W(t) \), cannot be estimated. However, notice that the input-output relation (1.8) induces a relation between \( C_W(m) \) and the average autocorrelations of \( U(\cdot) \) and \( d(\cdot) \). For instance, if \( v(t) = 0 \) and \( M = 1 \), then \( d(t) = w(t)u(t) \) and, assuming that \( w(t) \) is independent of \( u(t) \), we obtain \( c_d(m) = c_w(m)c_u(m) \), which allows direct calculation of \( c_u(m) \). The result for the general case \((M \geq 1, v(t) \neq 0)\), as given in [33, 34], depends on certain average autocorrelations of the composite signal \( \xi(t) = d(t)U^*(t) \), namely

\[
\begin{pmatrix}
1 & \{\text{vec}\{C_U^*(m)\}\}^* \\
\text{vec}\{C_U^*(m)\} & \Gamma_U(m)
\end{pmatrix}
\begin{pmatrix}
c_v(m) \\
\text{vec}\{C_W(m)\}
\end{pmatrix}
= 
\begin{pmatrix}
c_d(m) \\
\text{vec}\{C_\xi(m)\}
\end{pmatrix}
\tag{1.9}
\]

where \( \Gamma_U(m) = \mathbb{E}\{[\tilde{U}(t + m) \otimes U(t)][\tilde{U}(t + m) \otimes U(t)]^*\} \). Here \( \otimes \) denotes a Kronecker product, \( \text{vec}\{\cdot\} \) denotes the vectorization by columns [43], and \( \tilde{U}(t) \) is the element-wise complex conjugate of \( U(t) \).

In Sec. 2.3 we introduce a “Wiener problem” interpretation of (1.9) that allows us to use classical tools and algorithms from the theory of linear least squares estimation. In particular, we apply in Ch.3 standard adaptive filtering techniques to obtain efficient computational
solutions for the fundamental equation (1.9). We also modify (1.9) and its associated “Wiener problem” interpretation to allow for tracking of $E\{W^*(t)W(t+m)\}$ in applications where $W(t)$ is non-stationary, with a slowly-varying time-variant autocorrelation, again using classical results from adaptive filtering theory.

1.4 Prior-information-free estimation of $C_W(m)$

The fundamental equation (1.9) involves probabilistic moments (such as $C_U(m) = E\{U^*(t)U(t+m)\}$) which have to be estimated from the available signals in order to obtain an estimate of $C_W(m)$. Since the system input signal $u(\cdot)$ is assumed to be stationary, $\hat{C}_U(m)$ and $\hat{\Gamma}_U(m)$ can be estimated as

$$\hat{C}_U(m) = \frac{1 - \lambda}{1 - \lambda^{N-m}} \sum_{k=0}^{N-1-m} \lambda^{N-k-1}U(k+m)U^*(k)$$

(1.10a)

$$\hat{\Gamma}_U(m) = \frac{1 - \lambda}{1 - \lambda^{N-m}} \sum_{k=0}^{N-1-m} \lambda^{N-k-1}[\tilde{U}(k+m) \otimes U(k)][\tilde{U}(k+m) \otimes U(k)]^*$$

(1.10b)

Similarly, assuming that $W(t)$ is asymptotically-mean-stationary (which includes stationary as a special case) the average autocorrelations $c_d(m)$ and $C_\xi(m)$ can be estimated by the same method, viz.,

$$\hat{c}_d(m) = \frac{1 - \lambda}{1 - \lambda^{N-m}} \sum_{k=0}^{N-1-m} \lambda^{N-k-1}d(k+m)d^*(k)$$

(1.10c)

$$\hat{C}_\xi(m) = \frac{1 - \lambda}{1 - \lambda^{N-m}} \sum_{k=0}^{N-1-m} \lambda^{N-k-1}\xi^*(k+m)\xi(k)$$

(1.10d)

The expressions (1.10) can be used in two different ways:

- When we need only the autocorrelation of a stationary $W(t)$, or the average autocorrelation of an AMS $W(t)$, we should use conventional (long-term) averaging, which corresponds to setting $\lambda = 1$ in (1.10). We shall call this approach static estimation.

- When we wish to track the slowly varying autocorrelation of a non-stationary $W(t)$, we have to set $\lambda < 1$, selecting a value that matches the rate of change of the time-variant...
\( E\{W^*(t)W(t + m)\} \). We shall call this approach \textit{dynamic estimation}.

In the dynamic case, the estimates obtained via (1.10) will be time-variant (i.e., they will depend on \( N \)). For large \( N \), they become the well-known exponentially-weighted estimates [21]

\[
\widehat{C}_U(m; t) = (1 - \lambda) \sum_{k=0}^{t-m} \lambda^k U(t - k)U^*(t - k - m)
\]

where we replaced \( N - 1 \) by “\( t \)” to emphasize the time dependence of this estimate. Similarly,

\[
\widehat{\Gamma}_U(m; t) = (1 - \lambda) \sum_{k=0}^{t-m} \lambda^k [\widehat{U}(t - k) \otimes U(t - k - m)] [\widehat{U}(t - k) \otimes U(t - k - m)]^*
\]

\[
\widehat{c}_d(m; t) = (1 - \lambda) \sum_{k=0}^{t-m} \lambda^k d(t - k)d^*(t - k - m)
\]

\[
\widehat{C}_\xi(m; t) = (1 - \lambda) \sum_{k=0}^{t-m} \lambda^k \xi^*(t - k)\xi(t - k - m)
\]

Hence the whole process to obtain \( \widehat{C}_W(m) \) consists of two steps: (i) setting up (1.9), i.e., estimating the parameters in (1.9) and (ii) solving (1.9). Since the only information used to set up (1.9) consists of measurements of \( u(t) \) and \( d(t) \), we refer this approach as \textit{prior-information-free} estimation.

The computational cost of estimating \( C_W(m) \) involves a relatively heavy computational load: the cost of setting up the fundamental equation (1.9) is \( \mathcal{O}(M^4NL) \), and the cost of solving this equation is \( \mathcal{O}(M^6L) \) (Table 1.1). For typical values of \( M, L, N \) (say \( 10 \leq M \leq 50, 100 \leq L \leq 500 \) and \( 10,000 \leq N \leq 50,000 \)) these costs can amount to \( 10^{12} \) to \( 10^{14} \) flops.

<table>
<thead>
<tr>
<th></th>
<th>full version</th>
<th>diagonal version</th>
</tr>
</thead>
<tbody>
<tr>
<td>setting up the fundamental equation</td>
<td>( \mathcal{O}(M^4NL) )</td>
<td>( \mathcal{O}(M^2N) )</td>
</tr>
<tr>
<td>solving the fundamental equation</td>
<td>( \mathcal{O}(M^6L) )</td>
<td>( \mathcal{O}(M^3L) )</td>
</tr>
</tbody>
</table>

Table 1.1: Computational cost of fundamental equation

The value of \( M \), the number of elements in \( W(t) \), depends on the delay spread of the channel and on the sampling rate. For most wireless communication channels, \( M \) is normally smaller than 10 [9], while for underwater acoustic communication \( M \) can be as high as 30 [10]. When
over-sampling is used, $M$ can reach values of 100-200. The estimation of $S_W(f)$, the Doppler spectrum of $W(t)$, requires evaluation of $\{C_W(m); 0 \leq m \leq L\}$ for a fairly large $L$ (2000-5000 for a 36kHz sampling rate, with 4-8Hz resolution). Finally, the value of $N$, the number of data points used to estimate those moments in (1.9), determines the accuracy of $C_W(m)$ estimation. The main cause for inaccuracy is the estimation of the fourth-order moment matrix $\Gamma_U(m)$, which may require $N \approx 100,000$ for relative errors in $||\Gamma_U(m)||_F$ of the order of 5% or less.

For the dynamic version of (1.9), the “Wiener problem” interpretation that we introduce in Sec. 2.3 opens up the way for a variety of efficient computational implementations, some of which provide a more efficient alternative to the static solution that was proposed in [34, 40]. This approach reduces the overall cost to $O(M^2NL)$ because it involves time recursive (e.g. LMS type) solutions of a set of $L$ “Wiener problems”, each one of order $M^2$ (see Ch. 3). Thus it is still important to find ways to reduce $M$ (e.g. by using sparsity), and $L$. A detailed discussion is provided in Chs. 4 and 5.
Chapter 2

Estimating the Response Autocorrelation of a Linear Time Varying System

The matrix $C_W(m)$ provides information about the autocorrelation of each individual $w_k(t)$ as well as about the cross-correlation between the elements of $W(t)$. The various elements of $C_W(m)$ and of the associated multichannel spectrum $S_W(f)$ can be used to determine correlation time, correlation frequency, power spread and delay spread of the channel. The importance of $C_W(m)$ also lies in the fact that once acquired, it can be utilized in many ways as discussed in Ch. 1. In particular, our estimate of $C_W(m)$ should provide a complete (second-order) statistical characterization of both the “slow variation” and “fast variation” of $W(t)$, such as observed, for instance, in underwater acoustic channels [37].

Our method of obtaining $C_W(m)$ relies on a linear relation between the unknown $C_W(m)$ and certain average autocorrelations constructed from the input signal $U(t)$ and the output signal $d(t)$ [40]. This relation gives rise to a set of $M^2 + 1$ equations in $M^2 + 1$ unknowns, namely $c_v(m)$ and the $M^2$ elements of $C_W(m)$, for a fixed value of $m$. In addition, we discuss also the case of uncorrelated scatterers, in which $C_W(m)$ is diagonal, so that there are only $M + 1$ unknowns to determine.

The quality of our estimates of $c_v(m)$ and $C_W(m)$ depends on the statistical distribution of the system response $W(t)$ and the signals $U(t)$ and $d(t)$. A “Wiener problem” interpretation
that we introduce in Sec. 2.3 provides explicit results about the accuracy of these estimates in terms of classical results about estimator variance in linear least squares problems.

2.1 Prior-info-free estimation of $C_W(m)$

The starting point of the prior-info-free (PIF) estimation method is the available data: the input signal $u(t)$ and the corresponding output $d(t)$. Under the assumption of an FIR time varying discrete-time channel we can write the input-output relation as (recall (1.6) and Figure 1.2 from Ch. 1)

$$d(t) = W(t)U(t) + v(t)$$

(2.1)

where $u(t)$ is assumed to be stationary and the noise $v(t)$ is assumed to be independent of the input and system parameters. This input-output equation induces relations between $C_W(m)$ and certain autocorrelations and crosscorrelations of $d(t)$, $U(t)$ and the composite signal $\xi(t)$ defined as

$$\xi(t) \triangleq d(t)U^*(t)$$

(2.2)

Such relations make it possible to determine $C_W(m)$, the average autocorrelation of an unobserved (i.e. “hidden”) process $W(t)$ from estimates of certain second order moments of $d(t)$, $U(t)$ and $\xi(t)$.

To gain insight into this approach, we consider first the memoryless case ($M = 1$) and then generalize our conclusions to the dynamic case ($M > 1$). Notice that the definition of $C_W(m)$ has been generalized for non-stationary $W(t)$.

The memoryless channel ($M = 1$):

The simplest illustration of our approach is provided by the noiseless memoryless case (i.e. when $v(t) = 0$). In this case $d(t) = W(t)U(t) = w(t)u(t)$, so that

$$c_d(m) = c_w(m)c_u(m)$$
where
\[ c_u(m) \triangleq E\{u(t + m)u^*(t)\} , \quad c_d(m) \triangleq \langle E\{d(t + m)d^*(t)\} \rangle \]
are the average autocorrelations of the input and output signals, and \( c_w(m) = \langle E\{w^*(t)w(t + m)\} \rangle \equiv C_W(m) \) is the average autocorrelation of the (scalar) channel gain \( w(t) \). Hence,

\[ c_w(m) = \frac{c_d(m)}{c_u(m)} \]

is determined from the average autocorrelations of \( d(t) \) and \( u(t) \).

In the presence of noise this autocorrelation relation becomes

\[ c_d(m) = c_w(m)c_u(m) + c_v(m) \] (2.4)

where
\[ c_v(m) \triangleq E\{v(t + m)v^*(t)\} \]

For a fixed \( m \), this is a single equation in the two unknowns \( c_w(m) \) and \( c_v(m) \). Thus we need one more equation in the same two unknowns in order to be able to uniquely determine \( c_w(m) \) (and \( c_v(m) \)). One possibility, and probably the simplest one, is to consider \( \xi(t) = d(t)u^*(t) \) (still \( M = 1 \)), so that

\[
\begin{align*}
c_\xi(m) &= \langle E\{d(t + m)u^*(t + m)d^*(t)u(t)\}\rangle \\
&= \langle E\{[w(t + m)u(t + m) + v(t + m)][w^*(t)u^*(t) + v^*(t)]u^*(t + m)u(t)\}\rangle
\end{align*}
\]

Since \( v(\cdot) \) is independent of \( u(\cdot) \), this produces the relation

\[ c_\xi(m) = c_v(m)c_u^*(m) + \gamma_u(m)c_w(m) \] (2.6)

where
\[ \gamma_u(m) \triangleq E\{|u(t + m)|^2|u(t)|^2\} \]
and
\[ c_\xi(m) \triangleq \langle E\{\xi(t+m)\xi^*(t)\} \rangle \tag{2.7} \]

Combining equation (2.4) and equation (2.6), we now have two equations in two unknowns, namely
\[
\begin{pmatrix}
1 & c_u(m) \\
c_u^*(m) & \gamma_u(m)
\end{pmatrix}
\begin{pmatrix}
c_v(m) \\
c_w(m)
\end{pmatrix}
=
\begin{pmatrix}
c_d(m) \\
c_\xi(m)
\end{pmatrix}
\tag{2.8}
\]
so that \( c_w(m) \) and \( c_v(m) \) can be determined uniquely, provided that \( \gamma_u(m) \neq |c_u(m)|^2 \). This is indeed so because the \( 2 \times 2 \) system matrix of equation (2.8) can be interpreted as the covariance matrix of the random vector \( \begin{pmatrix} 1 \\ u^*(t+m)u(t) \end{pmatrix} \). Since the two elements of this vector are linearly independent (except when \( u(t) \) is a non-random constant scalar), the corresponding covariance is strictly positive definite. Consequently \( \gamma_u(m) > |c_u(m)|^2 \), and (2.8) has a unique solution.

The dynamic channel (\( M > 1 \)):

When \( M > 1 \), \( W(t), U(t) \) and \( \xi(t) \) are all vectors of length \( M \) and the various average autocorrelations become matrices. The same general approach can be used, but certain matrices need to be vectorized. In particular, the unknown matrix \( C_W(m) \) is represented by the vector \( \text{vec}\{C_W(m)\} \), where \( \text{vec}\{\cdot\} \) denotes the vectorization of a matrix by columns. The end result, obtained in [34, 40] is
\[
\begin{pmatrix}
1 \\ \text{vec}\{C_U^*(m)\}
\end{pmatrix}
\begin{pmatrix}
\{\text{vec}\{C_W^*(m)\}\}^* \\
\Gamma_U(m)
\end{pmatrix}
\begin{pmatrix}
c_v(m) \\
\text{vec}\{C_W(m)\}
\end{pmatrix}
=
\begin{pmatrix}
c_d(m) \\
\text{vec}\{C_\xi(m)\}
\end{pmatrix}
\tag{2.9}
\]
where \( C_W(m), C_U(m) \) and \( C_\xi(m) \) are the average autocorrelations of \( W(t), \xi(t) \) and \( U(t) \) as
defined in Sec. (1.1) (recall equation (1.2)), viz.,

\[ C_U(m) \triangleq E\{U(t + m)U^*(t)\} \]
\[ C_W(m) \triangleq \langle E\{W^*(t)W(t + m)\} \rangle \]
\[ C_\xi(m) \triangleq \langle E\{\xi^*(t)\xi(t + m)\} \rangle \]

and \( c_d(m) \) and \( c_v(m) \) are defined as in equations (2.3), (2.5). Also

\[ \Gamma_U(m) \triangleq E\{[\tilde{U}(t + m) \otimes U(t)][\tilde{U}(t + m) \otimes U(t)]^*\} \]  

(2.10)

Here the tilde \( \sim \) denotes element-wise complex conjugation without transposition and \( \otimes \) denotes the Kronecker product. Equation (2.9) is a matrix equation with \( M^2 + 1 \) unknowns (\( M^2 \) entries of \( C_W(m) \) and the scalar \( c_v(m) \)) and \( M^2 + 1 \) equations. Again, the \( (M^2 + 1) \times (M^2 + 1) \) system matrix can be interpreted as the covariance matrix of the random vector

\[
\begin{pmatrix}
1 \\
\tilde{U}(t + m) \otimes U(t)
\end{pmatrix}
\]

so that in general this matrix is positive definite and equation (2.9) has a unique solution. However, when \( m \leq M - 2 \), the elements of \( \tilde{U}(t + m) \otimes U(t) \) are linearly dependent and the system matrix becomes singular. A method for overcoming this singularity has been introduced in [40].

**Alternative form of (2.9)**

When our concern is \( C_W(m) \) only, we can eliminate \( c_v(m) \) and get the solution for \( C_W(m) \) alone from (2.9), viz.,

\[
\mathcal{R}_U(m) \text{vec}\{C_W(m)\} = \text{vec}\{C_\xi(m)\} - c_d(m) \text{vec}\{C_\xi^*(m)\} \]

(2.11a)

where

\[
\mathcal{R}_U(m) \triangleq \Gamma_U(m) - \left[ \text{vec}\{C_\xi^*(m)\} \right] \left[ \text{vec}\{C_U^*(m)\} \right]^* \]

(2.11b)
and $\Gamma_U(m)$ is as defined in (2.10). Since
\[
\text{vec}\{C^*_U(m)\} = E\{\text{vec}[U(t)U^*(t + m)]\} = E\{\tilde{U}(t + m) \otimes U(t)\} = E\Psi_m(t)
\]
we conclude that
\[
\mathcal{R}_U(m) = E\left\{\left[\Psi_m(t) - E\Psi_m(t)\right]\left[\Psi_m(t) - E\Psi_m(t)\right]^*\right\}
\tag{2.11c}
\]
namely the covariance matrix of the *centered* random vector $\Psi_m(t) - E\Psi_m(t)$. A similar result holds for the right side of (2.11a) by observing that
\[
C_\xi(m) = \langle E\{\xi(t + m)\xi^*(t)\} \rangle = \langle E\{d(t + m)d^*(t)U^*(t + m)U(t)\} \rangle
\]
so that
\[
\text{vec}\{C_\xi(m)\} - c_d(m)\text{vec}\{C^*_U(m)\} = \text{vec}(E\{[d(t + m)d^*(t) - c_d(m)][U^*(t + m)U(t) - C^*_U(m)]\})
\tag{2.11d}
\]
which is a vectorized version of the cross-covariance matrix of a centered random variable $d(t + m)d^*(t) - c_d(m)$ and a centered random matrix $U^*(t + m)U(t) - C^*_U(m)$. Both (2.9) and (2.11) are linked with the same “Wiener problem” — see Sec. 2.3.

### 2.2 Uncorrelated scattering: diagonal $C_W(m)$

As discussed in Ch. 1, *uncorrelated scattering* is a case of special interest in many communications channels. In particular, this is true for atmospheric channels, and approximately true for short underwater acoustic channels. In this case $C_W(m)$ is diagonal, which means there are only $M$ unknowns in $C_W(m)$, so that equation (2.9) has only $M + 1$ unknowns to be considered. We consider two approaches to deal with this diagonal case.
Least squares approach:

Assume $C_W(m)$ is exactly diagonal, viz.,

$$C_W(m) = \begin{pmatrix} c_w^{(m)}(1, 1) & 0 & \cdots & 0 \\ 0 & c_w^{(m)}(2, 2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_w^{(m)}(M, M) \end{pmatrix}$$

In this case, equation (2.9) is over-determined, since there are $M^2 + 1$ equations with only $M + 1$ unknowns. The sparsity of $\text{vec}\{C_W(m)\}$ can be characterized by a “collapsing matrix” $C$, so that [40]

$$\begin{pmatrix} c_w^{(m)}(1, 1) \\ c_w^{(m)}(2, 2) \\ \vdots \\ c_w^{(m)}(M, M) \end{pmatrix} = C \cdot \text{vec}\{C_W(m)\}$$

(2.12)

where

$$C = \begin{pmatrix} e_1 \\ e_{M+1+1} \\ e_{2(M+1)+1} \\ e_{3(M+1)+1} \\ \vdots \\ e_{(M-1)(M+1)+1} \end{pmatrix}$$

and $e_i = [0 \cdots 0 1 0 \cdots 0]$ is the unit vector. The same collapsing matrix can also be used to remove the unused columns of the system matrix in equation (2.9), resulting in a “tall and thin” system matrix $A$, viz.,

$$A = \begin{pmatrix} \{\text{vec}\{C_U^*(m)\}\}^* \\ \text{vec}\{C_U^*(m)\} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \Gamma_U(m) & C^T \end{pmatrix}$$

(2.13)
The collapsed system of equations become

\[
A \begin{pmatrix}
    c_w^{(m)}(1, 1) \\
    c_w^{(m)}(2, 2) \\
    \vdots \\
    c_w^{(m)}(M, M)
\end{pmatrix} = \begin{pmatrix}
    c_d(m) \\
    \text{vec}\{C_\xi(m)\}
\end{pmatrix}
\] (2.14)

Since it is over-determined, we can adopt a least squares solution approach, viz.,

\[
\begin{pmatrix}
    c_w^{(m)}(1, 1) \\
    c_w^{(m)}(2, 2) \\
    \vdots \\
    c_w^{(m)}(M, M)
\end{pmatrix} = (A^*A)^{-1}A^* \begin{pmatrix}
    c_d(m) \\
    \text{vec}\{C_\xi(m)\}
\end{pmatrix}
\]

Since our estimates of \(c_d(m), C_\xi(m)\) and the elements of \(A\) all contain random errors, we expect that the least squares approach will reduce the level of estimation error in \(c_v(m)\) and \(c_w(i, i)\).

**Reduced equation set approach:**

As an alternative to an overdetermined set of equations, we may consider constructing a set of \(M + 1\) equations in the \(M + 1\) unknowns \(\{c_w^{(m)}(k, k); 1 \leq k \leq M\}\) and \(c_v(m)\). One such solution is concealed within the least squares solution itself. If we obtain the SVD of the matrix \(A\) of equation (2.14), then the least squares solution of (2.14) is also the unique solution of the equation

\[
(\Sigma V^*) \begin{pmatrix}
    c_w^{(m)}(1, 1) \\
    c_w^{(m)}(2, 2) \\
    \vdots \\
    c_w^{(m)}(M, M)
\end{pmatrix} = U_s^* \begin{pmatrix}
    c_d(m) \\
    \text{vec}\{C_\xi(m)\}
\end{pmatrix}
\]

where

\[
A = [U_s \ U_o] \begin{pmatrix}
    \Sigma \\
    0
\end{pmatrix} V^*
\]

is the SVD decomposition of \(A\).
2.3 Wiener (MMSE) problem interpretation

In Sec. 2.1 we have observed, following [33, 34], that the matrix

\[
\mathcal{M}_U \triangleq \begin{pmatrix}
1 & \{\text{vec}C_U^*(m)\}\star \\
\text{vec}\{C_U^*(m)\} & \Gamma_U(m)
\end{pmatrix}
\]

occurring on the left-hand-side of (2.9) can be interpreted as

\[
\mathcal{M}_U = E\left\{ \begin{pmatrix} 1 \\ \Psi_m(t) \end{pmatrix} \begin{pmatrix} 1 \\ \Psi_m(t) \end{pmatrix}^\star \right\}
\]

where \(\Psi_m(t) \triangleq \tilde{U}(t + m) \otimes U(t)\) is the Kronecker product of two column vectors. Thus \(\mathcal{M}_U\) is the autocorrelation matrix of \(U(t) \triangleq \begin{pmatrix} 1 \\ \Psi_m(t) \end{pmatrix}\). We now observe that the right-hand-side of (2.9) is, in fact equal to the cross correlation

\[
E\left\{ \begin{pmatrix} 1 \\ \Psi_m(t) \end{pmatrix} \mathcal{D}_m^*(t) \right\}
\]

where \(\mathcal{D}_m(t) \triangleq d^*(t + m)d(t)\). Since

\[
c_d(m) = E\{d(t + m)d^*(t)\} = E\{\mathcal{D}_m^*(t)\}
\]

and

\[
\text{vec}\{C_\xi(m)\} = E\{\text{vec}\{\xi^*(m)\xi(t + m)\}\}
\]

\[
= E\{\text{vec}\{d^*(t)U(t)d(t + m)U^*(t + m)\}\}
\]

\[
= E\{\text{vec}\{d(t + m)d^*(t)U(t)U^*(t + m)\}\}
\]

\[
= E\{d(t + m)d^*(t)\text{vec}\{U(t)U^*(t + m)\}\}
\]

\[
= E\{\Psi_m(t)\mathcal{D}_m^*(t)\}
\]

24
we observe that (2.9) becomes

\[
M_{tt}(m) \begin{pmatrix} c_v(m) \\ \text{vec}\{C_W(m)\} \end{pmatrix} = E \begin{pmatrix} 1 \\ \Psi_m(t) \end{pmatrix} D_m^*(t)
\]

With the following recognition comparing to a standard Wiener problem setting [21],

\[
X \rightarrow \begin{pmatrix} 1 \\ \Psi_m(t) \end{pmatrix} \\
y \rightarrow D_m(t) \\
H \rightarrow \begin{pmatrix} c_v(m) \\ \text{vec}\{C_W(m)\} \end{pmatrix}
\]

we conclude that (2.9) is the Wiener-Hopf equation for the MMSE problem

\[
\min_{H,b} E \left| D_m(t) - \begin{bmatrix} b & H \end{bmatrix} \begin{pmatrix} 1 \\ \Psi_m(t) \end{pmatrix} \right|^2
\]

whose optimal solution is \((H, b)\) is a row vector

\[
b_{opt} = c_v^*(m), \quad H_{opt} = [\text{vec}\{C_W(m)\}]^*
\]

Alternatively, we can write this as

\[
\min_{H,b} E \left| D_m(t) - \left( H\Psi_m(t) + b \right) \right|^2
\]

which is a linear MMSE estimation problem (a.k.a, “Wiener problem”) for non-zero-mean random variables (see App. A for a detailed discussion).

Historically, the Wiener problem \(\min_H E \left| y - HX \right|^2\) was designed to determine a linear estimate \(\hat{y} = HX\) of an unknown variable \(y\) from measurements of the random vector \(X\). However, in the problem setting we use here \(D_m(t)\) and \(\Psi_m(t)\) are both known, and our objective is
simply to determine the unknown coefficients $H = [\text{vec}\{C_W(m)\}]^*$ and $b = c_v^*(m)$. For this reason we will refer to (2.16) as the Wiener Identification Problem (WIP) associated with (2.9). A similar observation applies in the standard formulation of the adaptive filtering paradigm, where one associates a WIP with a system identification problem [21].

Since (2.16) involves non-zero mean random variables, the corresponding Wiener-Hopf equation determines both the additive constant “$b$” and the coefficient vector “$H$” (equivalently both $c_v(m)$ and $\text{vec}\{C_W(m)\}$). We shall refer to this format as the non-centered Wiener-Hopf equation (see App. A). It can be replaced by two separate equations, one for “$H$” alone, and the other one for “$b$” alone. Once $H$ is estimated, $b$ can be estimated as

$$ b = E\mathcal{D}_m(t) - HE\Psi_m(t) $$

The resulting equation for $H$ is a centered Wiener-Hopf equation:

$$ E\{\overline{\Psi}_m(t)\overline{\Psi}_m(t)^*\}H^* = E\{\overline{\Psi}_m(t)\overline{\mathcal{D}}_m(t)^*\} $$

or equivalently,

$$ \mathcal{R}_U(m)H^* = \mathcal{R}_{UD}(m) $$

where $\overline{\Psi}_m(t), \overline{\mathcal{D}}_m(t)$ are centered versions of these random variables, namely,

$$ \overline{\Psi}_m(t) \triangleq \Psi_m(t) - E\Psi_m(t) $$

$$ \overline{\mathcal{D}}_m(t) \triangleq \mathcal{D}_m(t) - E\mathcal{D}_m(t) $$

and

$$ \mathcal{R}_U(m) \triangleq E\{\overline{\Psi}_m(t)\overline{\Psi}_m(t)^*\} $$

$$ \mathcal{R}_{UD}(m) \triangleq E\{\overline{\Psi}_m(t)\overline{\mathcal{D}}_m(t)^*\} $$

This is the standard version of the Wiener-Hopf equation presented in all textbooks (see, e.g.
We now recognize our alternative (reduced) formulation (2.11) as the centered version of the Wiener-Hopf equation for the WIP (2.16), viz.,

$$\min_{H} E \left| D_m(t) - H \Psi_m(t) \right|^2$$  \hspace{1cm} (2.19)

We recall that for lag values in the range $0 \leq m < M$, the random vector $\Psi_m(t)$ has linearly-dependent elements, so that both (2.9) and (2.11) become singular. In fact, for real-valued signals, $\Psi_m(t)$ contains several random variables that occur twice. One solution, proposed in [40], is to eliminate the redundant elements from both $\Psi_m(t)$ and $\text{vec}\{C_W(m)\}$, which leads to the “compressed” WIP

$$\min_{H,b} E \left| D_m(t) - \left( H \tilde{\Psi}_m(t) + b \right) \right|^2$$  \hspace{1cm} (2.20)

where $\tilde{\Psi}_m(t) = T_e \Psi_m(t)$, and $T_e$ is a matrix consisting of only “0” and “1” entries that selects the non-redundant elements of $\Psi_m(t)$. Another approach could be to treat (2.16) as a singular WIP, and use the Moore-Penrose pseudo-inverse to solve the associated (singular) Wiener-Hopf equation.

As for the case of a non-stationary $W(t)$, there are two separate directions that we can proceed: (a) focus only on the average autocorrelation of $W(t)$ (assuming that $W(t)$ is AMS), and (b) focus on the “instantaneous autocorrelation” $C_W(m; t) \triangleq E\{W^*(t)W(t+m)\}$. In the former case $C_W(m)$ is defined via (1.2), viz., $C_W(m) \triangleq \langle E\{W^*(t)W(t+m)\} \rangle_t$: the entire preceding discussion applies in this case, with $c_d(m)$ and $C_\xi(m)$ both defined as average autocorrelations. In this case, the time averaging operation can also be added to the WIP interpretation (2.16) if we assume the time-varying system is AMS, so that (2.16) becomes

$$\min_{H,b} \langle E \left| D_m(t) - \left( H \Psi_m(t) + b \right) \right|^2 \rangle_t$$

and similar for other equations. In the latter case (1.9) still holds with $c_d(t; m) \triangleq E\{d(t + m)d^*(t)\}$ and $C_\xi(t; m) \triangleq E\{\xi^*(t + m)\xi(t)\}$, since both $d(t)$ and $\xi(t)$ will now be non-stationary.
Thus, the WIP interpretation still holds, but in an instantaneous sense, i.e.,

\[
\left[ \begin{array}{c}
\text{vec}^* \{ C_W(m; t) \},
\psi^*_v(m; t)
\end{array} \right] = \underset{(H,b)}{\arg\min} \ E \left| D_m(t) - \left( H \Psi_m(t) + b \right) \right|^2
\]  

(2.21)

The corresponding Wiener-Hopf equations are now time-variant, and so are the solutions for \( C_W(\cdot) \) and \( c_v(\cdot) \).

In order to set up the equations (2.9) or (2.11) we need to estimate the relevant second-order joint statistics of the processes \( D_m(t) \) and \( \Psi_m(t) \). This is straightforward in the static version which uses only average autocorrelations and cross correlations (see Sec. 1.1). However, the dynamic (i.e., instantaneous) version (2.21) involves a distinct set of 2nd order moments for every time instant “\( t \)”. One case when such moments can be estimated by conventional methods is when their time-variation is slow. Using either the LMS or the RLS algorithms with a suitable “forgetting factor” makes it possible to track the slow variation of \( c_d(m; t) \) and \( C_\xi(m; t) \) (see Sec. 1.4). Another advantage of using standard estimation/tracking algorithms is the availability of explicit results about the accuracy (i.e., estimator variance) of \( \hat{C}_W(m) \) in both the static and the dynamic case [21].

### 2.4 The linear regression assumption

The classical (FIR) Wiener problem \( \min_H E \| y - H X \|^2 \) makes no assumptions about the nature of the relation between the random vector \( X \) and the random variable \( y \). The solution \( H_{opt} = E \{ yX^* \} [E \{ XX^* \}]^{-1} \) always holds, provided that the elements of \( X \) are linearly independent. When \( y(t) \) and \( X(t) \) are random signals, the coefficient vector \( H_{opt} \) will be time-varying, except when \( y(\cdot) \) and \( X(\cdot) \) are jointly stationary.

The Wiener problem formulation is often applied to reveal hidden linear relations. For instance, when \( y \) and \( X \) satisfy the so called linear regression assumption [44], namely when

\[
y(t) = AX(t) + v(t)
\]  

(2.22)
the unknown coefficient vector $A$ coincides with the solution of the Wiener problem, viz.,

$$A = \arg\min_{\hat{H}} E |y - HX|^2 \triangleq H_{opt}$$

Here $v(\cdot)$ is a white noise process that is independent of $X(\cdot)$, and $X(\cdot)$ is assumed to be (wide-sense) stationary. The linear regression assumption holds in a wide variety of applications, and is routinely included in the analysis of adaptive filter performance [44].

It should be noticed, however, that the linear regression assumption is not required when we use LMS or RLS (or long-term averaging) to solve the Wiener problem using only measurements of $y(\cdot)$ and $X(\cdot)$. We have discussed in Sec. 2.3 (see also Sec. 1.1) three distinct sets of assumptions that lead to a deterministic solution, (static or dynamic) of the Wiener problem: none of these involved the linear regression assumption. Similarly, the (almost sure) convergence of the deterministic linear least squares estimate of $H_{opt}$, viz.,

$$\hat{H}(N) = \left[\sum_{t=0}^{N} y(t)X^*(t)\right]\left[\sum_{t=0}^{N} X(t)X^*(t)\right]^{-1}$$

(2.23)

to $H_{opt}$ (as $N \to \infty$) requires only\(^1\) that $Ee_{opt} = 0$ and $E\{e_{opt}(t)X^*(t)\} = 0$ where $e_{opt} \triangleq y(t) - H_{opt}X(t)$, but does not rely on the linear regression assumption [44].

A direct examination of the relation between $D_m(t)$ and $\Psi_m(t)$ reveals, in fact, that these random processes do not satisfy the linear regression assumption. Indeed,

$$D_m(t) = d^*(t+m)d(t)$$

$$= [W(t+m)U(t+m) + v(t+m)]^*[W(t)U(t) + v(t)]$$

$$= [W^T(t) \otimes W^*(t+m)]^* \Psi_m(t) + v^*(t+m)v(t)$$

$$+ U^*(t+m)W^*(t+m)v(t) + v^*(t+m)W(t)U(t)$$

(2.24)

so that the presence of the two cross-term expressions in the right-hand side of (2.24) results in a violation of the linear regression assumption. Nevertheless, it is still true that $\hat{C}_W(m; t)$

\(^1\)In addition, $y(\cdot)$ and $X(\cdot)$ must satisfy a mild regularity constraint.
converges almost surely to $C_W(m)$, and similarly for $\hat{c}_v(m,t)$, under mild regularity assumptions on the processes $u(\cdot)$ and $d(\cdot)$. This is so because we can always define

$$\varepsilon_{m,\text{opt}}(t) \triangleq D_m(t) - (H_{\text{opt}}\Psi_m(t) + b_{\text{opt}}) \quad (2.25a)$$

so that

$$D_m(t) = (H_{\text{opt}}\Psi_m(t) + b_{\text{opt}}) + \varepsilon_{m,\text{opt}}(t) \quad (2.25b)$$

where $E\varepsilon_{m,\text{opt}}(t) = 0$, as well as (see App. A)

$$E \{ \varepsilon_{m,\text{opt}}(t)\Psi_m^*(t) \} = 0$$

Since the known results about steady state behavior of the LMS and RLS algorithms rely on the linear regression assumption (2.22) (see, e.g., [21]), one of our research objectives is to examine the applicability of these classical results to our $\hat{C}_W(m; t)$ estimates. In particular, the steady-state variance of elements of $\hat{C}_W(m; t)$ is proportional to the step size $\mu$ for LMS, and to $(1 - \lambda)$ for exponentially weighted RLS (see Sec. 3.2). These results provide explicit guidelines about the choice of the relevant accuracy-control parameter: $\mu$ for LMS, $\lambda$ for exponential RLS, or $N$ for growing window RLS.

Similarly, the variance of each element of the (static) estimate (2.23) is inversely proportional to $N$, the length of our data record (for large $N$ values). In fact, it is known [44] that the probability density function of the scaled estimation error $\sqrt{N}\text{vec}\{\hat{H}(N) - H_{\text{opt}}\}$, where $\hat{H}(N)$ is defined in (2.23), converges as $N \to \infty$ to a Gaussian zero-mean distribution with covariance equal to $J_{\text{min}}R_X^{-1}$, where $R_X$ is the covariance matrix of the stationary zero-mean multichannel signal $X(t)$, and $J_{\text{min}} = \min_{\hat{H}}E|y - HX|^2$. This result, which relies on the linear regression assumption (2.22), provides a direct explicit characterization for the accuracy of the estimate $\hat{H}(N)$, namely for large $N$ we have the explicit error covariance matrix expression $E\left\{ \left[\hat{H}(N) - H_{\text{opt}}\right] \left[\hat{H}(N) - H_{\text{opt}}\right]^* \right\} \approx \frac{1}{N}J_{\text{min}}R_X^{-1}$. Since the relation between $D_m(t)$ and $\Psi_m(t)$ does not satisfy the linear regression assumption, as is evident from (2.24), there is no guarantee that the
scaled estimation error $\sqrt{N}\text{vec}\{\hat{C}_W(m) - C_W(m)\}$ is asymptotically Gaussian. Nevertheless, we conjecture that the covariance of this scaled error is still given, approximately, by $J_{\text{min}}\mathcal{R}_U^{-1}(m)$, where $\mathcal{R}_U(m)$ is the centered covariance of $\Psi_m(t)$ (recall (2.11)), and $J_{\text{min}}$ is the minimal achievable value of the WIP cost function (2.16). In contrast, earlier work [34, 40] could only obtain even approximate results for the accuracy of the intermediate estimates $\hat{C}_U(m)$, $\hat{\Gamma}_U(m)$, $\hat{c}_d(m)$ and $\hat{C}_\xi(m)$.

Our conjecture is motivated, in part, by the observation that the four-term expression (2.24) for $D_m(t)$ implies

$$E\{D_m(t)\big|U(0:\infty)\} = [\text{vec}C_W(m)]^* \Psi_m(t) + c^*_v(m)$$

(2.26)

where $U(0:\infty) \triangleq \{U(t): 0 \leq t \leq \infty\}$, and we relied on the fact that $v(t)$ is a zero-mean signal independent of $U(\cdot)$ and $W(\cdot)$, so that $E\{U^*(t + m)W^*(t)v(t)\} = 0 = E\{v^*(t + m)W(t)U(t)\}$. Combining this result with (2.25a) (2.25b) we conclude that the residual $\varepsilon_{m,\text{opt}}(t)$ can also be expressed as

$$\varepsilon_{m,\text{opt}}(t) = D_m(t) - E\{D_m(t)\big|U(0:\infty)\}$$

so that $\varepsilon_{m,\text{opt}}(t)$ is uncorrelated not only with $\Psi_m(t)$, but with every nonlinear (Borel) function of $U(0:\infty)$, the complete history of the multichannel signal $U(\cdot)$. This “martingale property” is in between full independence (i.e., the linear regression property), which would ensure asymptotic Gaussianity, and lack of correlation between $\varepsilon_{m,\text{opt}}(t)$ and $\Psi_m(t)$, which guarantees almost sure convergence of the estimate $\hat{C}_W(m)$ to its true value $C_W(m)$, as $N \to \infty$.

The conditional mean relation (2.26) is, in fact, the starting point for the construction of a $\hat{C}_W(m)$ estimate in [39]. To be specific, the derivation in [39] involves a number of additional assumptions:

(i) $u(t)$ and $v(t)$ are assumed i.i.d, and

(ii) a state space model is used to characterize the time evolution of $W(t)$.

Since the authors of [39] are not aware of the key relation (2.9) or its WIP interpretation, they resort to an approximation argument, claiming that the difference $D_m(t) - E\{D_m(t)\big|U(0:\infty)\}$
is small. This motivates the introduction of a deterministic “counterpart” of the WIP (2.16), namely
\[ \min_{H,b} \sum_{t=0}^{N} \left| D_m(t) - \left( H \Psi_m(t) + b \right) \right|^2 \]
which leads to non-recursive estimates for \( C_W(m) \) and \( c_v(m) \), similar to the ones proposed in [33].

2.5 Computational cost

The fundamental set of equations (2.9), and its WIP interpretation (2.16), both depend on the structured indices \( M \) and \( L \). We need to determine a block \((M \times M)\) autocorrelation \( C_W(m) \) for a range of lag values, \((0 \leq m \leq L)\). Thus a direct (non-iterative) solution of (2.9) requires \( O(M^6L) \) computations (in addition to the cost of setting up this set of equations). This observation motivates our interest in reducing both \( M \) and \( L \), as well as finding more efficient ways to set up and solve the fundamental equation (2.9).

We demonstrate in Ch. 5 that the effective value of \( M \), the length of the coefficient vector \( W(t) \), can be reduced by exploiting the sparsity of \( W(t) \). This allows a significant reduction in computational cost, in view of the dependence of all cost components on \( M \) (see Table 2.1). Similarly, we propose in Ch. 4 to use the Maximum Entropy Method, combined with a multichannel Levinson algorithm to reduce the effective value of \( L \).

However, the most significant cost reduction is achieved by introducing time-recursive techniques for the solution of the fundamental equation (2.9). A direct (non-iterative) solution as proposed in [33] and [39], requires \( O(M^6L) \) operations, in addition to a setting-up cost of \( O(M^4LN) \), where \( N \) is the length of the signal record used to estimate the parameters of the equation (2.9) (see Table 2.1). Since high value of \( N \) (in excess of \( 10^4 \)) are needed to achieve acceptable accuracy of the estimates \( \hat{C}_W(m) \), we conclude that \( O(M^4LN) \gg O(M^6L) \), so that the overall cost of setting up and solving the fundamental equations (2.9) is proportional to \( M^4 \). In contrast, the time-recursive solution techniques that we develop in Ch. 3 allow us to reduce the overall cost as low as \( O(M^2LN) \).
Table 2.1: Computational cost of static (non-iterative) solution

| \( \hat{c}_d(m) \) | \( N - M + 1 \) | \( N - M - 1 \) |
| \( \hat{C}_U(m) \) | \( M^2 \cdot (N - M) + 1 \) | \( M^2(N - M - 1) \) |
| \( \hat{C}_\xi(m) \) | \( M^2(N - M) + 1 \) | \( M^2(N - M) \) |
| \( \hat{\Gamma}_U(m) \) | \( M^4(N - M) + 1 \) | \( M^4(N - M) \) |
| \( \mathcal{R}_U(m) \) | \( M^4 \) | \( M^4 \) |

\([\mathcal{R}_U(m)]^{-1}\mathcal{R}_{MD}(m)\) \(O(M^b)\) \(O(M^b)\)
Chapter 3

Time Recursive Implementation

The WIP interpretation (2.16) provides us the potential to utilize classical adaptive filter algorithms, such as LMS, NLMS, RLS, QR-RLS and fast RLS, in our problem setting to obtain time-recursive estimates of $C_W(m)$ and $c_v(m)$. This approach has several advantages, as compared with a non-recursive solution, namely:

- WIP allows time recursive implementations (Sec. 3.1). Using LMS and NLMS in the stationary case results in reduced cost compared to an offline (non-recursive) solution of (2.9).

- The time-recursiveness of the implementation also allows for tracking of time-varying $C_W$. Standard adaptive algorithms have a control parameter ($\mu$ or $\lambda$) that allows tradeoff between misadjustment and tracking lag (Sec. 3.2).

- Simulation results suggest that the steady-state accuracy of our time-recursive estimates of $C_W(\cdot)$ and $c_v(\cdot)$ can be predicted from standard metrics (misadjustment, tracking error and/or convergence rate) of adaptive algorithm performance. Even though our WIP interpretation does not satisfy the linear regression assumption (recall Sec. 2.4), the standard metrics provide a useful (albeit conservative) approximate for the misadjustment and convergence rate of our time-recursive solutions.
3.1 Time-recursive implementation

The WIP interpretation (2.19) provides the basic framework for adaptive implementation, as illustrated in Fig. 3.1. We consider three well-known adaptive algorithms: (i) RLS, which is the time-recursive equivalent of the non-recursive solution of Sec. 2.1; (ii) LMS, which offers significant cost reduction as compared with RLS (and with the non-recursive solution) but slower initial convergence; and (iii) NLMS with time-varying step size, which allows faster initial convergence than LMS at a modest increase in cost.

Notice that in order to form the centered signals $\overline{D}_m(t)$ and $\overline{\Psi}_m(t)$ we must first obtain estimates of the means $E\overline{D}_m(t)$ and $E\overline{\Psi}_m(t)$.

**Estimation of the mean**

The composite signal $\Psi_m(t) = \tilde{U}(t + m) \otimes U(t)$ is stationary, by virtue of our assumption that the input signal $u(t)$ is narrow-sense stationary (or, at least, has stationary 4th-order moments). Thus $E\overline{\Psi}_m(\cdot)$ can be estimated by standard time-averaging or, equivalently, by exponential averaging with $\lambda \approx 1$ (Fig. 3.2). On the other hand, the output signal $d(t)$ inherits the statistical properties of the system time-variant response $W(t)$. Indeed, the relation...
\[ d(t) = W(t)U(t) + v(t) \] implies that

\[
ED^*_m(t) = E\{d(t + m)d^*(t)\}
\]

\[ = c_v(m) + [E\Psi_m(t)]^* \text{vec}\{E[W^*(t)W(t + m)]\} \tag{3.1}
\]

Thus, when \( W(t) \) is (wide-sense) stationary so is \( d(t) \), resulting in a constant mean \( ED_m(\cdot) \).

\[
\Psi_m(t) \xrightarrow{\frac{1-\lambda}{1-\lambda z-1}} \text{estimate of } E\Psi_m(\cdot)
\]

Figure 3.2: Time-recursive estimation of \( E\Psi_m(\cdot) \)

Similarly, when \( W(t) \) is AMS, so is \( d(t) \): in this case we use a time-averaged WIP interpretation to determine the time averaged autocorrelation

\[
C_W(m) \triangleq \langle E\{W^*(t)W(t + m)\}\rangle_t
\]

Consequently, the centered version of \( D_m(t) \) is obtained in this case by subtracting the time-averaged mean \( \langle ED_m(t)\rangle_t \). This quantity can also be estimated by exponential averaging with \( \lambda \approx 1 \) (Fig. 3.3).

\[
\mathcal{D}_m(t) \xrightarrow{\frac{1-\lambda}{1-\lambda z-1}} \text{estimate of } \langle ED_m(t)\rangle_t
\]

Figure 3.3: Time-recursive estimation of \( \langle ED_m(t)\rangle_t \)

Finally, when the autocorrelation \( E\{W^*(t)W(t + m)\} \) is slowly time-varying (the “dynamic” case) we observe that (3.1) implies that \( ED_m(t) \) is also slowly time-varying. Thus, it can also be estimated by exponential averaging (Fig. 3.2), but with a somewhat smaller value of \( \lambda \), adjusted to the rate of variation of \( ED_m(t) \).
LMS Recursion

The LMS recursion that adaptively solves the centered WIP (2.19) is

\[ e_m(t) = \overline{D_m(t)} - H_m(t-1)\overline{\Psi_m(t)} \]  

\[ H_m(t) = H_m(t-1) + \mu e_m(t) \left[ \overline{\Psi_m(t)} \right]^* \]

for \( t = 1, 2, 3, \ldots \), where \( \mu \) is the step size parameter, and \( H_m(0) = 0 \). The guideline for choosing \( \mu \) in the time-invariant case is \( 0 < \mu < \mu_{cr} \) where a conservative approximation for \( \mu_{cr} \) is \( 2/3 \overline{R} \) and \( \overline{R} \) is the covariance of the process \( \overline{\Psi_m(t)} \). In our case this is the \( M^2 \times M^2 \) matrix \( \overline{R_{\Psi}}(m) = E\{ \overline{\Psi_m(t)\overline{\Psi_m}^*(t)} \} \) (recall (2.11)). Simulation results are shown in Fig. 3.4 and Fig. 3.5. Fig. 3.4 presents the learning curve of \( c_w(0) \) of the algorithm (3.2) for three different \( \mu \) values: 0.0004, 0.00025, 0.0001. From the figure, we can see that the smaller the \( \mu \), the longer the convergence time. The steady state behavior of LMS is also shown in Fig. 3.4 for the data length from \( 1 \times 10^4 \) to \( 2 \times 10^4 \): the level of steady state (output) error increases with \( \mu \). Thus the value of \( \mu \) controls the tradeoff between steady state error (a.k.a “misadjustment”) and rate of convergence. This is clearly evident in Fig. 3.5: the larger the time constant the smaller the misadjustment, and vice versa. Since the time constant is proportional to the inverse of \( \mu \), the desired choice of \( \mu \) can be obtained intuitively from this figure. The solid line in Fig. 3.5 corresponds to the tradeoff curve of \( D_S \) vs. the time constant defined by the expressions of Table 3.1. However, as explained in Sec. 3.2, these expressions need not apply to our WIP. We found, however, that scaling up the expression for \( D_S \) by a factor of 5 results in a reasonable match with our empirical results (indicated by * in Fig. 3.5).

Since the length of \( \overline{\Psi_m(t)} \) is \( M^2 \), the cost of the LMS algorithm (3.2) is \( O(2M^2) \), as compared with the cost of the non-recursive solutions, which is \( O(M^4) \) per signal sample, for a single value of \( m \). (recall Sec. 2.5).
Figure 3.4: LMS Learning curve for $c_w(0)$ for three different $\mu$ values

Figure 3.5: Tradeoff curve for LMS: $D_S$ vs. the time constant for $c_w(0)$
RLS Recursion

The RLS algorithm for our Wiener problem (2.16) is $(0 < \lambda < 1)$

\[
\begin{align*}
\pi_m(t) &= P_m(t-1)\bar{\Psi}_m(t) \\
k_m(t) &= \frac{\pi_m(t)}{\lambda + \bar{\Psi}_m^*(t)\pi_m(t)} \\
e_m(t) &= D_m(t) - H_m(t-1)\bar{\Psi}_m(t) \\
H_m(t) &= H_m(t-1) + e_m(t)k_m^*(t) \\
P_m(t) &= \lambda^{-1}P_m(t-1) - \lambda^{-1}k_m(t)\bar{\Psi}_m^*(t)P_m(t-1)
\end{align*}
\]

with initial values $H_m(0) = 0$ and $P_m(0) = \delta^{-1}I$ where $\delta \ll 1$. RLS provides lower steady state estimation error and better tracking ability than LMS over a wide range of its control parameter $(\lambda)$ values.

Fig. 3.6 shows the learning curve of RLS for three different values of the exponential forgetting factor $\lambda$. The part of data length from 2000 to 5000 is the steady state performance. This figure confirms the known result that increasing $\lambda$ lengthens the initial convergence time while reducing the steady state error. This tradeoff is illustrated explicitly in Fig. 3.7 (indicated by *). The solid line in Fig. 3.7 describes the tradeoff predicted by the classical expressions for convergence time and misadjustment that we discuss in Sec. 3.2 (Table 3.1). Again we observe that the tradeoff curve defined by the expressions of Table 3.1, and scaled-up by a factor of 9, provides a reasonable match to the empirical results.

The computational cost of RLS for real data is $O(M^4 + 7M^2 + 1)$ multiplications and $O(M^4 + 3M^2 + 1)$ additions. Hence it has a much higher computational cost than LMS, but has the same order of computational cost as the non-iterative (offline) solution of (2.9).
Figure 3.6: RLS Learning curve for $c_w(0)$ for three different $\lambda$ values.

Figure 3.7: Tradeoff curve for RLS: $\mathcal{D}_S$ vs. the time-constant for $c_w(0)$.
The tradeoff between steady state error (= misadjustment) and convergence time (see e.g. Fig. 3.5) motivates the introduction of several variable-step-size LMS and NLMS algorithms [45, 46, 47, 48]. In the stationary case such algorithms gradually decrease the size of $\mu$, which makes it possible to achieve both fast (initial) convergence and low steady state estimation error. Here we will focus on the variable step size NLMS algorithms of [45] and [46].

One convenient NLMS algorithm with time varying step size for our Wiener problem (2.16) is (based on an algorithm presented in [45])

$$e_m(t) = \bar{D}_m(t) - H_m(t-1) \Psi_m(t)$$
$$H_m(t) = H_m(t-1) + \mu(t-1) e_m(t) \Psi_m(t) [\Psi_m(t)^* \Psi_m(t)]^{-1}$$
$$\mu(t) = \mu(t-1) \frac{1 - \mu(t-1)/M^2}{1 - \mu^2(t-1)/M^2}$$

(3.3)

where $M^2$ is the length of the vector $\Psi_m(t)$. The initial value $\mu(0)$ is a control parameter that affects the rate of convergence. The derivation in [45] aims to optimize the choice of $\mu(0)$, resulting in

$$\mu(0) = 1 - \frac{J_{\min}}{\sigma_d^2}$$

(3.4a)

where, in the context of our WIP,

$$J_{\min} = E \left[ |\bar{D}_m(t) - [\text{vec} \{C_W(m)\}]^* \Psi_m(t) |^2 \right]$$

(3.4b)

and

$$\sigma_d^2 = E \left[ |\bar{D}_m(t)|^2 \right] = E \left\{ |d(t+m)|^2 |d(t)|^2 \right\}$$

(3.4c)
Figure 3.8: NLMS simulation learning curve of $c_w(0)$ with time varying step size algorithm of [45]

Figure 3.9: NLMS simulation learning curve of $c_w(0)$ with time varying step size algorithm of [46]
The other NLMS algorithm with time varying step size for our Wiener problem (2.16) is based on [46]

\[
e_m(t) = D_m(t) - H_m(t-1)\Psi_m(t)
\]

\[
H_m(t) = H_m(t-1) + \mu(t-1)e_m(t)\Psi_m^{-1}(t)\left[\Psi_m(t)^*\Psi_m(t)\right]^{-1}
\]

\[
\hat{p}(t) = \alpha\hat{p}(t-1) + (1-\alpha)e_m(t)\left[\Psi_m(t)^*\Psi_m(t)\right]^{-1}\Psi_m(t)
\]

\[
\mu(t) = \mu_{\text{max}} \frac{||\hat{p}(t)||^2}{(||\hat{p}(t)||^2 + C)}
\]

where \(\alpha\) is a smoothing factor \((0 \leq \alpha < 1)\), \(\mu_{\text{max}}\) is the largest value that time varying \(\mu\) can reach (should be chosen less than 2). \(C\) is a positive constant and can be approximated as \(\frac{M^2}{\text{SNR}}\) where \(\text{SNR}\) is the signal-to-noise ratio. In our simulation below, we choose \(M = 1\), \(C = 0.1\), \(\alpha = 0.99\) and \(\mu_{\text{max}} = 1.0\).

Simulation results for the learning curve of \(c_w(0)\) are shown in Fig. 3.8 and Fig. 3.9. The learning curves are almost the same for the two NLMS algorithms we consider here (adopted from [45] and [46]). In particular, the convergence time constant and the steady state error \(D_S(m)\) are practically the same for both NLMS techniques. This performance is indicated in Fig. 3.5 as a circle: it has a little better performance than LMS method from Fig. 3.5 because of its smaller time constant for the same estimation error, or smaller estimation error for the same time constant.

### 3.2 Performance Analysis

Our primary metrics of estimator performance (i.e., of \(\hat{C}_W(m;t)\) quality) are:

- **Estimator variance**, which we define as
  \[
  E\left\|\hat{C}_W(m;t) - E\hat{C}_W(m;t)\right\|^2_F
  \]

- **Convergence time** (for static applications), which we define as the number of data points \((=\text{time-recursion steps})\) needed to reduce the total estimation error \(E\left\|\hat{C}_W(m;t) - E\hat{C}_W(m;t)\right\|^2_F\) to, say, 1% of its initial value.

- **Tracking ability** (for dynamic applications), which we define as the “squared bias”
\( \| E \hat{C}_W(m; t) - C_W(m; t) \|_F^2 \)

Since both the variance and the bias may be time varying (even for large \( t \) values), we will use their time-averaged values as metrics. Thus, we define the average estimation variance as:

\[
\mathfrak{D}_S(m) \triangleq \left\langle E \left\| \hat{C}_W(m; t) - E \hat{C}_W(m; t) \right\|_F^2 \right\rangle_t
\]

the average tracking error (lag) in the tracking case as:

\[
\mathfrak{D}_L(m) \triangleq \left\langle \left\| E \hat{C}_W(m; t) - C_W(m; t) \right\|_F^2 \right\rangle_t
\]

Notice that the mean square deviation (MSD)

\[
\mathfrak{D}(m) \triangleq E \left\| \hat{C}_W(m) - E \hat{C}_W(m) \right\|_F^2 \equiv E \left\| \text{vec} \{ \hat{C}_W(m) \} - E \text{vec} \{ \hat{C}_W(m) \} \right\|_2^2
\]

\[
= \mathfrak{D}_S(m) + \mathfrak{D}_L(m)
\]

is equal to the sum of \( \mathfrak{D}_S(m) \) and \( \mathfrak{D}_L(m) \). In the static case there is no steady-state bias so that \( \mathfrak{D}_L(m) = 0 \) for all \( m \). For this reason we use the convergence time as a measure of agility, instead of \( \mathfrak{D}_L(m) \), in the static case (recall Figs. 3.5, 3.7).

Explicit expressions for both \( \mathfrak{D}_S \) and \( \mathfrak{D}_L \) have been derived in the literature (see, e.g. [21], [49]) by relying on the linear regression property. This means that the (zero-mean) signals \( y(t) \), \( X(t) \) in the generic Wiener problem

\[
\min_H E|y(t) - H X(t)|^2
\]

are assumed to satisfy the following conditions:

(a) There exists a vector \( H_{opt} \) such that \( y(t) = H_{opt} X(t) + v(t) \)

(b) The noise sequence \( \{ v(t) \} \) is i.i.d with variance \( \sigma_v^2 = E|v(t)|^2 \)

(c) The sequence \( v(i) \) is independent of \( X(j) \) for all \( i, j \)
Unfortunately, we don’t have the linear regression property in our WIP problem setting, namely these conditions do not hold in our case where \( y = \overline{D}_m(t) \) and \( X = \overline{\Psi}_m(t) \), so that there is no guarantee that the known expressions for \( D_S \) and \( D_L \) still apply.

Nevertheless, our simulation results in Sec. 3.1 suggest that the theoretical expressions of Tables 3.1 and 3.2 (below) can provide a useful approximation for the actual tradeoff relations, and thus serve as guidelines for selection of the control parameter (\( \mu \) for LMS, \( \lambda \) for RLS) in our time-recursive implementations.

<table>
<thead>
<tr>
<th></th>
<th>( D_S(m) )</th>
<th>time constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS</td>
<td>( \frac{\mu J_{\text{min}} M}{2} )</td>
<td>( \ln \frac{0.1}{\ln (1 - \mu_{\text{min}})} )</td>
</tr>
<tr>
<td>RLS</td>
<td>( \frac{(1 - \lambda) J_{\text{min}} \text{tr} R_{\varphi}^\dagger}{2} )</td>
<td>( \ln \frac{0.1}{\ln \lambda} )</td>
</tr>
</tbody>
</table>

Table 3.1: Steady state performance of LMS and RLS

<table>
<thead>
<tr>
<th></th>
<th>( D_S(m) )</th>
<th>( D_L(m) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS</td>
<td>( \frac{\mu J_{\text{min}} M}{2} )</td>
<td>( \frac{\text{tr} R_{\varphi}^\dagger}{2\mu} )</td>
</tr>
<tr>
<td>RLS</td>
<td>( \frac{(1 - \lambda) J_{\text{min}} \text{tr} R_{\varphi}^\dagger}{2} )</td>
<td>( \frac{\text{tr} R_{w}}{2(1 - \lambda)} )</td>
</tr>
</tbody>
</table>

Table 3.2: Tracking performance of LMS and RLS

### 3.2.1 Stationary (=Static) case

**LMS tradeoff**

The expression for the steady-state error variance of \( H_m(n) \) is given (for small \( \mu \) values) by [49]:

\[
D_S = \frac{\mu J_{\text{min}} M}{2}
\]

where

\[
J_{\text{min}} = \min_{\mu} E \left[ |\overline{D}_m(t) - H\overline{\Psi}_m(t)|^2 \right]
= E|\overline{D}_m(t)|^2 - [\text{vec}\{C_W(m)\}]^* \mathcal{R}_\varphi(m) [\text{vec}\{C_W(m)\}]
\]
and $M$ is the system length. The time constant is (for convergence to 10% of the initial value)

\[
\tau = \frac{\ln 0.1}{\ln(1 - \mu \lambda_{\text{min}})}
\]

and for small $\mu$, $\ln(1 - \mu \lambda_{\text{min}}) \approx -\mu \lambda_{\text{min}}$, so that $\tau$ is (approximately) inversely proportional to $\mu$.

**RLS tradeoff**

The steady state coefficient error variance for RLS is given (for $1 - \lambda \ll 1$) by

\[
\mathcal{D}_S = \frac{1 - \lambda}{2} J_{\text{min}} tr \left( R_{\Psi}^{-1}(m) \right)
\]

and the time constant is

\[
\tau = \frac{\ln 0.1}{\ln \lambda}
\]

for convergence to 10% of the initial value.

### 3.2.2 Dynamic case

When our objective is to track a time-variant autocorrelation $C_W(m; t)$, we should choose $\mu$ (for LMS) or $\lambda$ (for RLS) so as to achieve a desired tradeoff between $\mathcal{D}_L$ and $\mathcal{D}_L$. As discussed before, the known expressions (see Table 3.2) have limited applicability in our case, but can be used, nevertheless, to provide an approximate guideline for the selection of the step-size.

The expressions for $\mathcal{D}_L(m)$ in Table 3.2 assume that $C_W(m; t)$ is a random walk process, where $Q$ denotes the covariance matrix of the increment vec\{\(C_W(m; t) - C_W(m; t - 1)\)}.
Chapter 4

Lag Recursive Implementation

The time-recursive methods we presented in Ch. 3 determine an estimate of $C_W(m)$ for a single value of $m$, and the same statement holds for the non-recursive method discussed in Sec. 2.3. However, many applications require evaluation of $C_W(m)$ over a range of lag values, typically $0 \leq m \leq L$ for some $L$. For the purpose of our discussion here we distinguish between three types of applications:

- **Small $L$**, such as in transmission power control [37] ($L = 1$), or identification of a dynamic model for $W(t)$ [40] ($L = 2$ or $L = M + 1$).

- **Medium $L$**, such as needed for the construction of an optimized averaging filter [31], typically in the range of $10 - 100$.

- **Large $L$**, such as needed for an accurate estimate of $S_W(f)$, the Doppler spectrum of $W(t)$ (typical $L$ values are in the range of hundreds).

For small to medium values of $L$, we present here a technique for lag-recursive computation of $C_W(m)$, which relies on special (shift) structure exhibited by $C_W(m)$ (Sec. 4.1).

For large values of $L$ we propose an approach based on the Maximum Entropy Method (MEM) that makes it possible to obtain an estimate of the Doppler spectrum $S_W(f)$ from a relatively short sequence of autocorrelation lags: we only need to estimate \{\(C_W(m); 0 \leq m \leq L_0\}\} where $L_0$ is in the medium range of values (several tens). In contrast, a direct evaluation
of $S_W(f)$ via a Fourier transform (recall (1.4)) requires hundreds of autocorrelation lags to overcome the Gibbs phenomenon associated with the truncated expression (1.5).

4.1 Lag-recursive implementation

In this section we explore the shift-structure of $\Gamma_U(m) = E\{\psi_m(t)\psi_m^*(t)\}$ or, equivalently, $\mathcal{R}_U(m) = E\{[\psi_m(t) - E\psi_m(t)] [\psi_m(t) - E\psi_m(t)]^*\}$ with respect to the lag variable “m”. Recall that $u(t)$ is assumed stationary (up to 4th order), so that both $\Gamma_U(m)$ and $\mathcal{R}_U(m)$ are time-invariant. Two types of “block-shift-structure” are unveiled: inter-block shift structure and intra-block shift structure, where “block” is defined as an $M \times M$ submatrix, so that $\Gamma_U(m)$ consists of $M$ block rows and $M$ block columns, and similarly for $\mathcal{R}_U(m)$. We show that $\mathcal{R}_U(m)$ and $\mathcal{R}_U(m-1)$ share a large portion of their elements, so that the cost of constructing $\mathcal{R}_U(m)$, once $\mathcal{R}_U(m-1)$ is known, can be significantly reduced. In addition, the cost of solving the equation can also be reduced by using a lag-recursive procedure to efficiently determine $\mathcal{R}_U^{-1}(m)$. [38]

4.1.1 Inter-block shift structure

The inter-block shift structure becomes evident when we write $\psi_m(t)$ in blocks of $M$ elements each, viz.,

$$
\psi_m(t) = \begin{pmatrix}
\psi(t + m)U(t) \\
\vdots \\
\psi(t - M + 1)U(t)
\end{pmatrix}
$$

(4.1a)
Comparing (4.1a) and (4.1b), we conclude that the top \( M - 1 \) blocks (equals to top \( M^2 - M \) elements) of \( \Psi_{m-1}(t) \) coincide with the bottom \( M - 1 \) blocks of \( \Psi_m(t) \). In other words,

\[
\Psi_m(t) = \begin{pmatrix}
  u^*(t + m - 1) U(t) \\
  \vdots \\
  u^*(t + m - M + 1) U(t) \\
  u^*(t + m - M) U(t)
\end{pmatrix}
\]

Comparing (4.1a) and (4.1b), we conclude that the top \( M - 1 \) blocks (equals to top \( M^2 - M \) elements) of \( \Psi_{m-1}(t) \) coincide with the bottom \( M - 1 \) blocks of \( \Psi_m(t) \). In other words,

\[
\Psi_m(t) = \begin{pmatrix}
  u^*(t + m) U(t) \\
  \vdots \\
  Q \Psi_{m-1}(t)
\end{pmatrix}
\]

where

\[
Q = \begin{bmatrix}
  I_{M^2-M} & 0_{(M^2-M) \times M}
\end{bmatrix}
\]

Thus only the top \( M \) elements of \( \Psi_m(t) \) are new, as compared with the elements of \( \Psi_{m-1}(t) \).

As a result

\[
\mu_\Psi(m) \triangleq E\Psi_m(t) \equiv \text{vec}\{C_U^*(m)\}
\]

\[
= \begin{pmatrix}
  \times \\
  \vdots \\
  Q \mu_\Psi(m - 1)
\end{pmatrix}
\]

so that only \( M \) elements (at most) need to be estimated to obtain \( \mu_\Psi(m) \) from \( \mu_\Psi(m - 1) \).

Similarly,

\[
\Gamma_U(m) = \begin{bmatrix}
  \times & \times \\
  \times & Q \Gamma_U(m-1) Q^*
\end{bmatrix}
\]
so that only the first $M$ columns and first $M$ rows of $\Gamma_U(m)$ need to be estimated in addition to the $(M^2 - M) \times (M^2 - M)$ block of elements that is inherited without change from $\Gamma_U(m - 1)$. The same relation exists between $\mathcal{R}_U(m - 1)$ and $\mathcal{R}_U(m)$ (Fig. 4.1).

**Structured matrix inversion**

From the inter-block shift relation (4.2b) we conclude that

$$\mathcal{R}_U(m - 1) = \begin{pmatrix} A & B^* \\ B & D \end{pmatrix}$$

and

$$\mathcal{R}_U(m) = \begin{pmatrix} E & C^* \\ C & A \end{pmatrix}$$

where $[A] = (M^2 - M) \times (M^2 - M)$ is the shared part of $\mathcal{R}_U(m - 1)$ and $\mathcal{R}_U(m)$, and $[D] = M \times M = [E]$. Our objective is to propagate the inverse matrix

$$\mathcal{R}_U^{-1}(m) = \begin{bmatrix} \Theta^{-1} & -\Theta^{-1}C^*A^{-1} \\ -A^{-1}C\Theta^{-1} & A^{-1} + A^{-1}C\Theta^{-1}C^*A^{-1} \end{bmatrix}$$

(4.3)
where $\Theta$ is the $M \times M$ Schur-complement

$$\Theta \triangleq E - C^*A^{-1}C$$

It is evident from these expressions that the highest computational cost is associated with the inversion of the relatively large submatrix $A$. However, since the same submatrix appears also in $R_U(m - 1)$, we can use the block-matrix inversion formula (where $\Omega = D - BA^{-1}B^*$)

$$R_U^{-1}(m - 1) \triangleq \begin{pmatrix} X & Y^* \\ Y & T \end{pmatrix} = \begin{pmatrix} A^{-1} + A^{-1}B\Omega^{-1}BA^{-1} & -A^{-1}B\Omega^{-1} \\ -\Omega^{-1}BA^{-1} & \Omega^{-1} \end{pmatrix}$$

to conclude that

$$A^{-1} = X - Y^*T^{-1}Y \quad (4.4)$$

which requires only matrix products and the inversion of the much smaller ($M \times M$) submatrix $T$.

The cost of evaluating $A^{-1}$ via (4.4) is $O(M^5)$, and it dominates the cost of forming $R_U^{-1}(m)$ via (4.3). Thus the overall cost of solving the equation (2.18) for $C_W(m)$ is $O(M^5L)$, while a direct inversion of $R_U(m)$ would require $O(M^6L)$ (for $0 \leq m \leq L$).

**Generalized Displacement Structure**

An alternative efficient approach to solve the equation (2.18) is suggested by the observation that the generalized displacement matrix $R_U(m) - Z^M R_U(m - 1)(Z^M)^*$ has rank $2M$, where

$$Z = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & \ddots & \cdots & \cdots & \vdots \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$
is the so-called lower shift matrix \([50]\), with ones on the first subdiagonal and zeros elsewhere. The technique of \([50]\) can be used to obtain a lag-recursive (i.e., recursive in \("m\)”) procedure for propagating the Cholesky factor of \(R_U(m)\) at a cost of \(O(K^2r)\) per iteration step, where \(r\) is the displacement rank and \(K\) is the size of the matrix. Since in our case \(K = M^2\) and \(r = 2M\), the cost of propagating the structured Cholesky factor \(R_U^{1/2}(m)\) is \(O(M^3L)\) for \(0 \leq m \leq L\), which is comparable with the cost of the method we proposed earlier. Once the Cholesky factor \(R_U^{1/2}(m)\) is known, the equation (2.18) is solved via two back-substitution steps, at a cost of \(O(M^4L)\).

4.1.2 Intra-block shift structure

Since \(\Psi_m(t) = \text{vec}\{U(t)U^\ast(t + m)\}\), its elements can be expressed in the form

\[
\left[\Psi_m(t)\right]_p = u(t - k)u^\ast(t + m - i)
\]

where \(p = iM + k + 1\). Since \(0 \leq i \leq M - 1\) and \(0 \leq k \leq M - 1\), we conclude that \(1 \leq p \leq M^2\). It now follows by observation that

\[
\left[\Psi_m(t - 1)\right]_p = \left[\Psi_{m-1}(t)\right]_{p+1} \quad \text{for all } p \neq iM \quad (4.5)
\]

In particular, we notice that

\[
E\left[\Psi_m(t)\right]_p = E\left[\Psi_m(t - 1)\right]_p = E\left[\Psi_{m-1}(t)\right]_{p+1}
\]

for all \(1 \leq p \leq M - 1\). Combing this result with (4.2a), we find that only the \(M\)-th element of \(\Psi_m(t)\) needs to be estimated: the remaining \(M^2 - 1\) elements are all inherited, without alteration, from \(\Psi_m(t - 1)\).

Similarly, most of the elements in the first \(M\) columns of \(\Gamma_U(m)\) are inherited from \(\Gamma_U(m-1)\).
Indeed, the identity (4.5) implies that (relying on the stationarity of $\Psi_m(\cdot)$)

$$
\left[ \Gamma_U(m) \right]_{p,q} = E \left\{ \left[ \Psi_m(t - 1) \right]_p \left[ \Psi_m(t - 1) \right]^*_q \right\} \\
= E \left\{ \left[ \Psi_{m-1}(t) \right]_{p+1} \left[ \Psi_{m-1}(t) \right]^*_{q+1} \right\} \\
= \left[ \Gamma_U(m - 1) \right]_{p+1,q+1}
$$

for all $(p,q)$ such that $p \neq iM$, $q \neq jM$. This excludes the $M$-th column ($q = M$) of $\Gamma_U(m)$ as well as the last row of every block within the first $M$ columns (because these have $p = M, 2M, \cdots M^2$). The same relation exists between $R_U(m)$ and $R_U(m-1)$ (Fig 4.2). Combining this result with (4.2b), and the conjugate symmetry of $\Gamma_U(m)$, only $M \times M^2$ elements need to be estimated by considering intra-block alone. The combination of the inter-block and intra-block shift structures is shown in Fig. 4.3. Notice that only the blue area needs to be estimated when constructing $R_U(m)$ from $R_U(m-1)$. Those elements are $\gamma_{i,j}$, where $i$ and $j$ are from the sets $\{[i = M], [j = M], [i = 1, \ldots, M - 1, j = M, 2M, \ldots, M^2], [i = M, 2M, \ldots, M^2, j = 1, \ldots, M - 1]\}$. If we further consider the conjugate symmetry property of $R_U(m)$, then only half of them need to be estimated. Thus, only $M^2 + M \times (M - 1) - (M - 1) = 2M^2 - 2M + 1$ elements need to be estimated, which is $O(M^2)$, a reduction by a factor of $M^2$, compared to the direct (non-recursive) estimation of all the elements of $R_U(m)$.

The intra-block shift structure does not lend itself to block a decomposition such as (4.3), making it rather difficult to obtain a lag-recursive expression for $R_U^{-1}(m)$. It also appears difficult to relate it to the notion of generalized displacement structure. Consequently, using the intra-block structure to further reduce the cost of propagating $R_U^{-1}(m)$ is still an open research problem.

### 4.2 Maximum entropy autocorrelation extension

We propose here to use the maximum entropy autocorrelation extension method [51] to reduce the cost of estimating $\{C_W(m); 0 \leq m \leq L\}$ in applications with medium to large values of
Figure 4.2: Illustration of intra-block shift structure of $R_U(m - 1)$ (left) and $R_U(m)$ (right): yellow area covers the same elements while deep blue covers the different elements.

Figure 4.3: Illustration of combination of inter- and intra-block shift structure of $R_U(m - 1)$ (left) and $R_U(m)$ (right): yellow area are elements shared by two matrix and blue area are elements need to be estimated.
The maximum entropy method allows us to determine the desired autocorrelation estimates from a limited number of $C_W(m)$ values, effectively reducing the cost of estimating the needed lags of $C_W(\cdot)$. This is achieved by: (a) estimating $C_W(m)$ for $0 \leq m \leq L_0$ for some (small to medium) integer $L_0$, (b) solving a set of (multichannel) Yule-Walker equations of order $L_0$, and (c) using the resulting (matrix-valued) AR coefficients to determine $\hat{C}_W(m)$ for all $L_0 + 1 \leq m \leq L$. In addition, we show that the Doppler spectrum $S_W(f)$ can be estimated without explicit evaluation of the missing lags $\{C_W(m); L_0 + 1 \leq m \leq L\}$, so that step (c) above can be completely avoided in this case.

To be specific, the Yule-Walker equation relates the autocorrelation elements $\{C_W(m); 0 \leq m \leq L_0\}$ to the matrix-valued parameters $R_{L_0}^r$ and $\{A_i; 1 \leq i \leq L_0\}$ of the multichannel linear prediction model of order $L_0$, viz.,

$$C_W(m) = \sum_{l=1}^{L_0} A_l C_W(m - l) + R_{L_0}^r \delta(m), \quad m = 0, \ldots, L_0$$

These equations can be solved efficiently via the (multichannel) Levison algorithm [52] (see App. E). Thus the cost of obtaining the AR coefficients is $O(M^2 NL_0) + O(M^2 L_0^3)$: the first term describes the cost of solving the fundamental equation (2.9) for $0 \leq m \leq L_0$ using an LMS-type time-recursive technique, while the second term describes the cost of solving (4.6) via the Levinson algorithm. Since normally $N \gg L_0$, the first term dominates, so that evaluating the AR coefficients does not have a significant impact on the overall computational cost.

Once the AR coefficients have been determined, we use the AR autocorrelation extension recursion, namely, eq. (4.6) with $m \neq 0$, viz.,

$$C_W(m) = \sum_{l=1}^{L_0} A_l C_W(m - l)$$

starting with $m = L_0 + 1$ and progressing to $m = L$. The cost of this step is $O(M^3 L_0 L)$, where we assumed $L \gg L_0$. Clearly, this approach is useful only when its computational cost is below $O(M^2 NL)$, the cost of the approach discussed in Sec. 4.1. This happens when $ML_0 < N$, which defines an upper limit for the choice of $L_0$. 

55
One approach to estimate the Doppler spectrum is by using (1.5). This approach requires $L$ to be very high (such as 1000) to achieve a reasonable spectrum resolution, especially for low frequencies, so that the cost of this standard method (i.e., $O(M^2NL)$) could be high. In contrast, with the ME method, we can obtain a direct AR spectrum estimate using \{\bar{C}(m); 0 \leq m \leq L_0\} only, viz.,

$$S_W(f) = [A_{L_0}^{-1}(e^{2\pi f})] R_{L_0}^r [A_{L_0}^{-1}(e^{2\pi f})]^*$$

(4.7)

where $A_{L_0}(z) = I + A_{L_0,1} z^{-1} + \cdots + A_{L_0,L_0} z^{-L_0}$. Notice that $A_{L_0}(z)$, $R_{L_0}^r$ and $S_W(f)$ are all $M \times M$ matrices. The cost of a single $k$-point DFT is $O(K^2 \log_2 K)$, so that the cost of evaluating $A_{L_0}(e^{2\pi f})$ over a grid of $K$ frequency points is $O(M^2K^2 \log K)$. Thus the resulting total cost of this approach is $O(M^2NL_0) + O(M^2K^2 \log K)$, which is a significant improvement, compared to the standard (truncated Fourier transform) spectrum estimation approach.

Figure 4.4 is the simulation result of power spectrum estimation using periodogram and ME method respectively. The true spectrum is also shown as dotted line. We use $L_0 = 10$ here for maximum entropy method and $L = 1000$ for periodogram. It is clearly shown both method give almost the same result compare to the true power spectrum. Since maximum entropy method only uses 10 lags, it is 100 times less lags used by periodogram.

4.3 Combining time and lag recursiveness

At this point it is natural to ask whether we can combine the order-recursive procedures of this chapter with the adaptive (time-recursive) algorithms of Ch. 3. There appear to be two natural combinations, namely,

1. A time update for $m = 0$ alone, and a lag-recursive evolution at each time instant.
2. A lag-recursive evaluation of $\hat{C}_W(m,t)$ for $t = 0$, followed by a time update, for each individual lag value in the range $1 \leq m \leq L$.

The first approach is useful for estimating $C_W(m)$ using a block of data since it obviously induces considerable delays. The second approach is useful for instantaneous computation of
Figure 4.4: Power spectrum estimation comparison: FFT vs. ME

all the $\hat{C}_W(m; t)$ for $1 \leq m \leq L$.

However, none of these combinations offer a computational cost reduction as compared with a time-recursive evaluation of $\hat{C}_W(m; t)$ separately for each individual $m$. Recall that using LMS-like implementations results in an overall cost of $O(M^2NL)$. In contrast, the cost of the first combination proposed here is

$$O(M^2N) + O(M^3NL) \approx O(M^3NL)$$

while the cost of the second combination is (assuming LMS-like adaptation)

$$O(M^3L) + O(M^2NL) \approx O(M^2NL)$$

which is dominated by the second term, since $M \ll N$. Both combinations involve a higher cost than a time-recursive procedure with no coupling between distinct $m$ values.
Chapter 5

Exploiting Sparsity

Sparsity, in our discrete time setting, means that the impulse response $W(t)$ has significant values in only a few lags while other lags have relatively small values. One example is communication channels with multipath propagation: the (relatively short) responses of each propagation path are typically separated from each other as a result of wide differences in propagation delay. Another example is provided by the underwater acoustic communication channel, whose impulse response consists of a few significant propagation paths involving a relatively small number of (surface and bottom) reflections. While the length of the channel impulse response could be as high as 40, only a few lags (typically 5 or less) have significant values, separated by wide stretches of negligible lag values. Significant computational savings can be achieved by including only the non-negligible system response lags in the coefficient vector $W$, which may also result in improved identification performance.

5.1 Exploiting sparsity

We shall say that the time-varying system response is sparse when the input-output relation $y(t) = W(t)U(t)$ can be (approximately) expressed as $\sum_i w_{ki}(t)u(t - k_i)$ where $\{k_i; 1 \leq i \leq K\}$ is the set of lags with non-negligible $w_k(t)$ values. In matrix-vector notation, this means that (1.1) is replaced by

$$y(t) = \tilde{W}(t)\tilde{U}(t)$$

(5.1a)
where $\tilde{W}(t) \triangleq [w_{k_1}(t) \ w_{k_2}(t) \ \cdots \ w_{k_K}(t)]$ and

$$\tilde{U}(t) \triangleq \begin{pmatrix} u(t - k_1) \\ u(t - k_2) \\ \vdots \\ u(t - k_K) \end{pmatrix} \quad (5.1b)$$

Sparsity can result in significant reduction in computational cost when $K \ll M$, namely, when the reduced coefficient vector $\tilde{W}(t)$ is much shorter than the original coefficient vector $W(t)$.

Sparsity in the input-output relation (5.1) can be mapped directly into the WIP formulation (2.16), and consequently into the autocorrelation $C_W(m)$, which is replaced by the smaller matrix

$$\tilde{C}_W(m) = \langle E\{\tilde{W}(t)\tilde{W}(t+m)\} \rangle_t$$

The reduced WIP formulation is

$$\min_{\{H,b\}} E \left| D_m(t) - \left( H\tilde{\Psi}_m(t) + b \right) \right|^2$$

(5.2)

where $D_m(t) = d^*(t+m)d(t)$ as before, while $\tilde{\Psi}_m(t)$ only has $K^2$ elements, viz.,

$$\tilde{\Psi}_m(t) \triangleq \tilde{U}(t+m) \otimes \tilde{U}(t)$$

resulting in the solution

$$H_{opt} = \left[ \text{vec}\{\tilde{C}_W(m)\} \right]^*$$

The reduction in the length of $W(\cdot)$ from $M$ to $K$ elements results in: (i) a reduction in the size of the composite data vector $\Psi_m$ from $M^2$ to $K^2$, (ii) a similar reduction in the fundamental equation (2.9) from $M^2 \times M^2$ to $K^2 \times K^2$, so that only $K^2$ unknowns need to be estimated compared to $M^2$ unknowns.
5.2 Determining the sparsity profile

In order to use the reduced input-output relation (5.1) we must first determine the indices \( \{k_1, k_2, \ldots, k_K\} \), namely, the lag positions that have significant values in the impulse response. For time-varying systems, this sparsity profile can also be time-varying, viz., the lag positions that have significant values may be changing over time. However, we restrict our discussion here to the case of a static (time-invariant) sparsity profile. The sparsity profile \( \{k_1, k_2, \ldots, k_K\} \) can be determined, for instance, by running one of our time-recursive algorithms (for \( m = 0 \)) and observing the relative magnitude of the elements of the resulting \( M \times M \) autocorrelation \( \hat{C}_W(0) \).

Alternatively, one may consider techniques that enforce/enhance sparsity, such as orthogonal matching pursuit (OMP) or \( l_1 \)-relaxation.

**Orthogonal matching pursuit (OMP)**

Let us suppress the effect of \( m \) in (2.18), so that (2.18) can be written as

\[
\mathcal{R}_{UD} = \mathcal{R}_U H^*_{opt}
\]  

The OMP algorithm, which was proposed in [53], is based on the basic matching pursuit (MP) algorithm. Both algorithms are iterative procedures that identify the dominant entries of \( H \) in (5.3) and estimate the associated values.

The OMP algorithm starts with the initial residual \( x_0 = \mathcal{R}_{UD} \) and iteratively selects one of the column of \( \mathcal{R}_U \) that correlates best with the approximation residual from the previous iteration [54]. Then the residual vector \( x_{k-1} \) is computed as \( \mathcal{R}_{UD} \) minus the contributions of all the columns identified in the previous \( k - 1 \) iteration. At the \( k \)th iteration, the column of \( \mathcal{R}_U \) onto which the residual vector \( x_{k-1} \) has the maximal rank-one projection, denoted as \( r_{a_k} \), is found via

\[
a_k = \arg \max_{1 \leq j \leq M^2} \max_{j \notin P_{k-1}^*} \frac{|r_j^* x_{k-1}|^2}{\|r_j\|^2} \quad 1 \leq k \leq K
\]

where \( P_{k-1}^* \triangleq \{a_1, \ldots, a_{k-1}\} \) is the index set of all previously selected columns in \( \mathcal{R}_U \). Then
the OMP algorithm computes the associated $\hat{c}_k$ as follows:

$$\hat{c}_k = \arg\min_c \| \mathcal{R}_{UD} - \mathcal{R}_{U}^{a,k} c \|^2 = \left[ \left( \mathcal{R}_{U}^{n,k} \right)^* \mathcal{R}_{U}^{n,k} \right]^{-1} \left( \mathcal{R}_{U}^{n,k} \right)^* \mathcal{R}_{UD}$$

where $\mathcal{R}_{U}^{n,k} \triangleq [r_{a_1} \cdots r_{a_k}]$ are the collection of all the selected columns in $\mathcal{R}_{U}$. And the residual vector $x_k$ is computed as

$$x_k = \mathcal{R}_{UD} - \mathcal{R}_{U}^{n,k} \hat{c}_k \quad 1 \leq k \leq K$$

A stopping criterion is used to determine the number of dominant taps $K$. The last $\hat{c}_k$ becomes the solution of (5.3), i.e., $H^*_{opt} = \hat{c}_K$. Generally speaking, the choice of $K$ is a tradeoff between the variance of the estimation and the errors caused by insufficient parametrization [55]. In any case, it should be significantly smaller than the length of $H_{opt}$.

**$l_1$ relaxation**

In the $l_1$-relaxation method one seeks an approximate solution of $\mathcal{R}_{UD} = \mathcal{R}_{U} H^*_{opt}$ (hence the term relaxation), with the goal of reducing the $l_1$-norm of the solution. In other words, one solves the constrained optimization problem

$$\min_{H} \| H_{opt} \|_1 \quad \text{subject to} \quad \| \mathcal{R}_{UD} - \mathcal{R}_{U} H^* \|_2 \leq \beta$$

where the parameter $\beta$ controls the level of relaxation: increasing $\beta$ increases the set of possible $H$ vectors, allowing for a reduction of $\| H \|_1$. An unconstrained form of this objective is

$$\min \| \mathcal{R}_{UD} - \mathcal{R}_{U} H^*_{opt} \|_2^2 + \lambda \| H_{opt} \|_1$$

The $l_2$-term forces the residual $\mathcal{R}_{UD} - \mathcal{R}_{U} H^*_{opt}$ to be small, whereas the $l_1$-term enforces sparsity of the representation. The $\lambda$ controls the tradeoff between the sparsity of the spectrum and the residual norm.
This optimization criterion is again a convex optimization problem and can be readily handled by quadratic programming for real data. Second-order cone (SOC) programming for the complex data case is proposed in [56].
Chapter 6

Bandpass to baseband conversion

The original derivation of (2.9) makes no assumptions about the bandwidth and frequency content of the input signal $u(t)$ and the output signal $d(t)$. However, in many applications, such as communication channels, the input and output signals are narrow-band bandpass signals, obtained by modulating a narrow-band baseband signal by an appropriate carrier frequency. In such cases, it makes sense to replace $u(t)$ and $d(t)$ by their baseband equivalents before setting up the fundamental equation (2.9).

This is not as simple as it sounds because the classical theory of baseband equivalent transfer functions applies only to linear time-invariant channels. Specifically, the construction of such a baseband-equivalent system relies on the fact that linear time-invariant filtering does not involve frequency spreading. This assumption does not hold for LTVS, so we need to reexamine the idea of passband to baseband conversion in a time-varying context.

6.1 Bandwidth constraints

The baseband equivalent of (1.8) can be derived in a straightforward manner when the combined bandwidth of $S_u(f)$ and $S_w(f)$ does not exceed the frequency range $0 < f < \frac{1}{2}$. To illustrate this constraint, consider first the scalar case ($M = 1$), for which the system input-output relation is
\[ d(t) = w(t)u(t) + v(t) \]

and one of the autocorrelation relations is

\[ c_d(m) = c_w(m)c_u(m) + c_v(m) \]

or, equivalently in the frequency domain,

\[ S_d(f) = S_w(f) \otimes S_u(f) + S_v(f) \]

The convolution \( S_w(f) \otimes S_u(f) \) results in frequency spreading, and the bandwidth of the resulting spectrum is the sum of the individual bandwidths of \( S_u(f) \) and \( S_w(f) \). Now the baseband versions of the passband spectra \( S_d(f) \), \( S_v(f) \) and \( S_u(f) \) are obtained simply by shifting each one to be centered around the origin [57](see also App. C), namely

\[ [S_u(f)]_{\text{baseband}} = S_u^{(a)}(f + f_c) \equiv S_u^{(a)}(f) \otimes \delta(f + f_c) \]

where \( f_c \) is the center (=carrier) frequency of channel input signal \( u(t) \) and \( S_u^{(a)}(f) \) is the analytic part of \( S_u(f) \), viz.,

\[
S_u^{(a)}(f) = \begin{cases} 
S_u(f) & f > 0 \\
0 & f < 0
\end{cases}
\]

Now if \( S_w(f) \) is sufficiently narrow-band, then the convolution \( S_w(f) \otimes S_u(f) \) still describes a (somewhat wider) passband spectrum, and

\[
[S_w(f) \otimes S_u(f)]_{\text{baseband}} = S_w(f) \otimes [S_u(f)]_{\text{baseband}} \quad (6.1)
\]

This is so because of the property

\[
[S_w(f) \otimes S_u(f)]^{(a)} = S_w(f) \otimes S_u^{(a)}(f)
\]
so that the resulting baseband version is

\[ [S_w(f) \otimes S_u(f)]_{\text{baseband}} = S_w(f) \otimes S_u^{(a)}(f + f_c) \]

Consequently, we can recover \( S_w(f) \), or equivalently \( c_w(m) \) from the relation

\[ [c_d(m)]_{\text{baseband}} = c_w(m)[c_u(m)]_{\text{baseband}} + [c_v(m)]_{\text{baseband}} \quad (6.2) \]

by using the method discussed in Ch. 2.

In contrast, when \( S_w(f) \) is wideband, the convolution \( S_w(f) \otimes S_u(f) \) may result in an overlap between the positive and negative frequency components of \( S_w(f) \otimes S_u(f) \), so that (6.1) does not hold anymore.

An alternative perspective on the bandwidth constraints imposed on \( w(t) \) and \( u(t) \) is provided by using the relation

\[ u(t) = \text{Re}\{u_b(t)e^{jw_c t}\} \]

where \( u_b(t) \) is the (complex) baseband equivalent of \( u(t) \). Thus the (noise free) input-output relation \( y(t) = w(t)u(t) \) can be also written as

\[ y(t) = \text{Re}\{w(t)u_b(t)e^{jw_c t}\} \]

where we used the fact that \( w(t) \) is real-valued. This relation suggests that we may be able to define the baseband component of \( y(t) \) as \( y_b(t) = w(t)u_b(t) \), which is consistent with (6.2). However, we are constrained by the following fundamental limitation.

**Lemma:**

Given a discrete-time real signal \( x(t) \) and a complex signal \( c(t) \) such that \( x(t) = \text{Re}\{c(t)\} \), then the analytic signal

\[ x^{(a)}(t) = x(t) + j\mathcal{H}\{x(t)\} \]
coincides with \( c(t) \) if, and only if, the DTF of \( c(t) \) is restricted to the frequency range \( 0 < w < \pi \).

Above lemma is proved in [58]. This result imposes a constraint on the frequency content of \( w(t)u_b(t)e^{jw_t} \). However, since the frequency content of \( u_b(t)e^{jw_t} \) is simply the one sided version of the frequency content of the real-valued signal \( u(t) \), the constraint can be expressed in terms of the three frequencies \( f_w, f_{\text{min}}, f_{\text{max}} \) (Fig. 6.1). The frequency content of \( u_b(t)e^{jw_t} \) is restricted to \( f_{\text{min}} < f < f_{\text{max}} \), so that the frequency content of \( w(t)u_b(t)e^{jw_t} \) is restricted to

\[
 f_{\text{min}} - f_w < f < f_{\text{max}} + f_w
\]

and we obtain the specific constraints \( f_{\text{min}} - f_w > 0 \) and \( f_{\text{max}} + f_w < \frac{1}{2} \) or

\[
 f_{\text{min}} > f_w \quad ; \quad f_{\text{max}} < \frac{1}{2} - f_w
\]  

(6.3)

6.2 Baseband-equivalent relations

In the general case \( (M > 1) \) the system input-output relation is \( d(t) = y(t) + v(t) \) where

\[
y(t) = W(t)U(t) = \sum w_i(t)u(t - i)
\]
so that each one of the terms \( w_i(t)u(t-i) \) must satisfy the constraint (6.3). Notice that \( u(t-i) \) has the same frequency content as \( u(t) \), namely \( f_{\text{min}} < f < f_{\text{max}} \), but each \( w_i(t) \) may have its own cutoff frequency \( f_{w_i} \).

Assuming that (6.3) is satisfied by all \( w_i(t) \), we can now express \( y(t) \) as

\[
y(t) = \Re\{y_b(t)e^{j\omega_c t}\} \tag{6.4}
\]

where

\[
y_b(t) = [w_0(t) w_1(t)e^{-j\omega_c} \cdots w_{M-1}(t)e^{-j(M-1)\omega_c}]\begin{pmatrix} u_b(t) \\ u_b(t-1) \\ \vdots \\ u_b(t-M+1) \end{pmatrix} \tag{6.5}
\]

We conclude that \( y_b(t) \) and \( u_b(t) \) are related via a baseband equivalent of \( W(t) \), namely,

\[
y_b(t) = W_b(t)U_b(t) = [w_{b0}(t) w_{b1}(t) \cdots w_{b(M-1)}(t)]\begin{pmatrix} u_b(t) \\ u_b(t-1) \\ \vdots \\ u_b(t-M+1) \end{pmatrix} \tag{6.6}
\]

where

\[
W_b(t) = [w_{b0}(t) w_{b1}(t) \cdots w_{b(M-1)}(t)]
\]

\[
= [w_0(t) w_1(t)e^{-j\omega_c} \cdots w_{M-1}(t)e^{-j(M-1)\omega_c}]
\]

\[
= W(t)T_W \tag{6.7}
\]

and

\[
T_W = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & e^{-j\omega_c} & 0 & \vdots \\ \vdots & 0 & e^{-j2\omega_c} & \vdots \\ 0 & \cdots & \cdots & e^{-j(M-1)\omega_c} \end{pmatrix}
\]
This relation induces a similar one between the autocorrelations of \( W(t) \) and \( W_b(t) \), viz.,

\[
C_{W_b}(m) \triangleq \langle E\{W_b^*(t)W_b(t + m)\}\rangle
= \langle E\{T_W^*W(t)W(t + m)T_W\}\rangle
= T_W^* C_W(m) T_W \quad (6.8)
\]

where \(^*\) is matrix Hermitian conjugate. From (6.8), we can see that the diagonal entries of \( C_{W_b}(m) \) and \( C_W(m) \) are the same, which means the autocorrelation of each tap of the channel impulse response is the same whether it is obtained via a bandpass process or a baseband process. On the other hand, the crosscorrelation between distinct taps is different for baseband processing and bandpass processing.

### 6.3 Baseband WIP

A direct derivation of a baseband version of the fundamental equation (1.8) is hampered by the need to convert the fourth-order moment matrix \( \Gamma_U(m) \) from passband to baseband. In contrast, it is much easier to use the baseband input-output relation \( d_b(t) = w(t)u_b(t) + v_b(t) \) to establish a baseband version of the Wiener Identification Problem (WIP) (2.6). We shall concentrate here only on the memoryless case \((M = 1)\).

We assume in the sequel that the bandwidth constraints (6.3) are satisfied, so that the (noise-free) channel output \( w(t)u(t) \) has a baseband version given by \( w(t)u_b(t) \), as discussed in Sec. 6.2. In addition, we assume that the noisy output signal \( d(t) = w(t)u(t) + v(t) \) is the output of a bandpass filter, so that \( v(t) \) is restricted to the frequency range \( f_{\min} - f_w < f < f_{\max} + f_w \).

We can thus express \( v(t) \) in terms of its baseband equivalent

\[
v(t) = Re \{v_b(t)e^{jw_c t}\}
\]
where \( v_b(t) \) is restricted to the frequency range \([0 \ f_w + \frac{f_{\text{max}} - f_{\text{min}}}{2}]\). We conclude that

\[
d(t) = \text{Re} \left\{ [w(t)u_b(t) + v_b(t)] e^{jw_c t} \right\}
\]

and since \( w(t)u_b(t) + v_b(t) \) is appropriately band-limited, the resulting baseband relation is

\[
d_b(t) = w(t)u_b(t) + v_b(t)
\]

Next, we obtain the baseband equivalent of (2.24), namely (recall that \( M = 1 \) here)

\[
d^*_b(t + m)d_b(t) = w(t)w^*(t + m)u_b^*(t + m)u_b(t) + v_b^*(t + m)v_b(t) + u_b^*(t + m)w^*(t + m)v_b(t) + v_b^*(t + m)w(t)u_b(t)
\]

and use the fact that \( v_b(\cdot) \) is independent of \( u_b(\cdot) \) to derive the baseband version of (2.26), namely

\[
E \{ d^*_b(t + m)d_b(t) | u_b(0 : \infty) \} = c^*_w(m)u_b^*(t + m)u_b(t) + \left[ c_v(m) \right]_{\text{baseband}}
\]

Since, in general, for any random variable \( y \) and any random vector \( X \)

\[
E |y - E(y|X)|^2 \leq \min_{\{H,b\}} E |y - (HX + b)|^2
\]

it follows that \( c_w(m) \) and \( \left[ c_v(m) \right]_{\text{baseband}} \) can be determined via the baseband WIP

\[
\min_{\{H,b\}} E |d^*_b(t + m)d_b(t) - [Hu_b^*(t + m)u_b(t) + b]|^2
\]

because the minimizing solution is, according to (6.10)

\[
H_{\text{opt}} = c_w^*(m), \quad b_{\text{opt}} = \left[ c_v(m) \right]_{\text{baseband}}
\]

In summary, we can determine \( c_w(m) \) by applying our (adaptive) methodology to the complex baseband signals \( d_b(t) \) and \( u_b(t) \). Notice that \( \left[ c_v(m) \right]_{\text{baseband}} \) is the baseband version of \( c_v(m) \),
so that \( c_v(m) = Re \{ [c_v(m)]_{\text{baseband}} e^{j\omega_m} \}\).
Chapter 7

Concluding Remarks

7.1 Summary of contribution

- **Wiener Identification problem (WIP) formulation:**
  We have shown that the unknown autocorrelations $C_W(m)$ and $c_v(m)$ can be determined by solving a (non-zero mean) Wiener problem (2.16) involving the composite signals $D_m(t) \triangleq d^*(t + m)d(t)$ and $\Psi_m(t) \triangleq \tilde{U}(t + m) \otimes U(t)$. We also have shown that the fundamental equation (2.9) is, in fact, the Wiener-Hopf equation associated with this MMSE estimation problem.

- **Efficient time-recursive implementation:**
  The WIP interpretation makes it possible to use standard adaptive algorithm, such as LMS or RLS, to obtain computationally efficient estimates for $C_W(m), c_v(m)$ from the composite signal $D_m(t)$ and $\Psi_m(t)$. In particular, LMS and normalized LMS provide efficient alternatives to the direct (offline, non-recursive) solution used in [33, 39], significantly reducing the length of the data record needed to achieve acceptable estimation accuracy.

- **Linear conditional mean property:**
  We have established the conditional mean relation (2.26), namely

  \[ E\{D_m(t) \mid U(0 : \infty)\} = [\text{vec}C_W(m)]^* \Psi_m(t) + c_v^*(m) \]
The fact that this conditional mean is a *linear function* of $\Psi_m(t)$ provides an independent proof for the WIP formulation (2.16), which can be used to extend this formulation beyond the regime defined by the fundamental equation. In addition, the linear conditional mean property implies that

$$D_m(t) = [\text{vec}C_W(m)]^* \Psi_m(t) + c^*_m + \varepsilon_W(t)$$

where $E\{\varepsilon_W(t)|U(0:\infty)\} = 0$. This martingal-type property could be useful in deriving explicit expression for the steady-state error covariance of the estimate $\hat{C}_W(m)$.

- **Performance metrics and tradeoffs:**

As explained in Sec. 2.4, the explicit expression for the steady-state bias and variance of the estimates provided by LMS and RLS (see, e.g., [21], Sec. 3.2) do not apply to our estimate $\hat{C}_W(m)$. Nevertheless, our empirical study suggests that these expressions can provide (approximate) evaluations of the bias and variance of our $\hat{C}_W(m)$ estimate. In particular, the tradeoff curves obtained from the standard expressions can be made to match the empirical tradeoff curves if we scale $D_S$ (the average estimator variance) up by a factor of $5 - 10$.

- **Lag-recursive relations:**

We have shown that of the $M^4$ elements of the matrix $R_U(m)$, the covariance matrix of the random vector $\Psi_m all(t) - E\Psi_m(t)$, all but $2M^2$ coincide with suitable elements of $R_U(m-1)$. This makes it possible to reduce the cost of setting up $R_U(m)$ from $O(M^4N)$ to $O(M^2N)$ and the cost of a non-recursive solution of the fundamental equation (2.9) from $O(M^6)$ to $O(M^5)$. However, further research is needed to obtain an efficient time- and lag recursive solution: by comparison, our LMS and NLMS implementations provide a complete solution at a cost of $O(M^2N)$ computations per a single lag.

- **Maximum Entropy (ME) spectrum estimation:**

We have demonstrated that the Doppler spectrum $S_W(f)$ can be determined, via the ME method from a relatively small number of autocorrelation lags. Our empirical results confirm that an ME estimate of the Doppler spectrum, using $L_0 = 9$ has comparable quality to a
Fourier-estimate that uses $L = 1000$ lags of the autocorrelation sequence $C_W(m)$. In addition, the cost of the ME method, namely $O(M^2NL_0) + O(M^2K \log K)$, is significantly smaller than the cost of applying an FFT to the sequence $\{C_W(m); \ 0 \leq m \leq L\}$, which is $O(M^2NL) + O(M^2L \log L)$, even if we choose to set $K = L$.

- **Sparse system response:**

We have shown that the $1 \times M$ coefficient vector $W(t)$ can be replaced by a condensed (and much shorter) $1 \times K$ version $\tilde{W}(t)$. Thus the dimension of the fundamental equation (and of the associated WIP) is reduced from $M^2$ to $K^2$, where $K \ll M$, leading to significant reduction in the overall computational cost. The ability to shorter the length of the coefficient vector $W(t)$ offers a controllable tradeoff between cost and accuracy in those cases where the system impulse response is only “approximately sparse”, namely most of the elements of $W(t)$ are negligible as compared with $\|W(t)\|_\infty$.

- **Efficient baseband processing:**

We have shown that the composite baseband signals $d_b(t + m)d_b(t)$ and $\tilde{U}_b(t + m) \otimes U_b(t)$ satisfy the linear conditional mean relation (recall (6.10)), which results in a baseband version of the Wiener identification problem (WIP) formulation. This means we have a choice between passband processing and baseband processing in cases when the signal $U(t)$ and $d(t)$ are sufficiently narrow-band. We have also established explicit constraints on the bandwidth of those two signals that must be satisfied before baseband processing can be applied.

### 7.2 Future research

- **Explicit expressions for steady-state error:**

Since the WIP formulation (in terms of $D_m(t)$ and $\Psi_m(t)$) does not possess the linear regression property, as explained in Sec. 2.4, the standard expressions for self-noise $D_S$ (variance of $\hat{C}_W(m)$) and lag $D_L$ (bias of $\hat{C}_W(m)$) do not apply. Our results indicate that are still useful in predicting the level of self-noise for our time-recursive implementation, up to multiplication
by a constant scaling factor that is independent of the step size ($\mu$ or $\lambda$). There observation suggest two future research directions:

1. Obtaining an explicit expression for the scaling factor needed to match the value of $D_S$ provided by the standard expression with its actual value in our LMS or RLS implementations of the WIP formulation (2.16).

2. Use the linear martingal property (2.26) as a starting point for a new derivation of explicit expressions for the self-noise $D_S$ and the lag $D_L$ associated with our time-recursive implementations of (2.16).

• **Time-recursive estimation of $W(t)$:**

As shown in [31, 40], an improved estimate of $W(t)$ can be obtained by using generalized averaging (in contrast to the narrow-band lowpass averaging used in LMS and RLS). The information needed to construct such a customized averaging filter is provided by $\hat{C}_W(m)$ and $\hat{C}_v(m)$. Since our WIP interpretation allows us to determine these estimates in a time-recursive fashion, the logical next step is to attempt a time-recursive construction and implementation of the resulting generalized averaging filter. This is a formidable challenge, since the determination of the optimal averaging filter involves the solution of a large set of linear equations.

• **Efficient time/lag-recursive solution:**

Our study of inter-block and intra-block relations within the covariance matrix $R_U(m)$ of the composite data vector $\Psi_m(t)$ allowed us to reduce the cost of setting up (2.9) to $O(M^4NL)$ and the cost of solving it to $O(M^5L)$. However, since the cost of an LMS (or NLMS) solution is only $O(M^2NL)$, these time-recursive solutions are still the most efficient, without any reliance on the lag-recursive structure of $R_U(m)$. These observations suggest future research directions:

1. Explore the structure of $R_U(m)$ to reduce the cost of our RLS implementation, which now stands at $O(M^4NL)$. One possibility would be to rely on the displacement structure of $R_U(m)$ to efficiently propagate its Cholesky factor, resulting in a modified “QR-RLS” implementation (see, e.g. [50]).
2. Reformulate the WIP interpretation (2.16) to include multiple lag values, in an attempt to obtain (time-recursive) solutions with a cost that depends sub-linearly on the index $L$. If successful, this should result in an LMS/NLMS implementation whose cost is lower than $O(M^2NL)$.

- **Customized sparse implementation:**

The ability to formulate a sparse version of the WIP interpretation (recall (5.2)) motivates our interest in efficient determination of sparsity profiles. The methods we proposed in Sec. (5.2) need to be examined in terms of the tradeoff they offer between reduction of computational cost and accuracy of the resulting estimates $\{\hat{C}_W(m)\}$. An explicit characterization of this tradeoff will make it possible to select the sparsity profile based on estimator quality, rather than by considering the relative magnitudes of the (unknown) elements of $W(t)$.

- **Baseband processing:**

We have demonstrated the feasibility of estimating $C_W(m)$ from baseband versions of the system’s input and output signals. Future research should aim to:

1. Extend the baseband version of the WIP from the memoryless case ($M = 1$) to the directed one ($M \geq 1$).

2. Examine the utility (e.g. reduction in sampling rate, computational cost) of working with baseband equivalents, as compared to using the original (passband) signals.

3. Examine the performance of our time-recursive algorithms, when applied to baseband equivalents, in terms of accuracy and convergence rate.
Appendix A

Wiener Problem for Non-zero Mean Signals

The standard Wiener problem (see, e.g. [21]), namely \( \min_H E \left| |y - HX|^2 \right| \) assumes that all random variables have zero mean. When dealing with non-zero mean random variables we need to use the affine estimate \( \hat{y} =HX + a \), so that the corresponding Wiener problem becomes

\[
\min_{H, a} E \left| y - (HX + a) \right|^2
\]  

(A.1a)

This can be alternatively written as

\[
\min_{H, a} E \left| y - \begin{bmatrix} a & H \end{bmatrix} \begin{pmatrix} 1 \\ X \end{pmatrix} \right|^2
\]  

(A.1b)

which has the appearance of a standard Wiener problem, but with the augmented measurement vector \( \begin{pmatrix} 1 \\ X \end{pmatrix} \).

The optimal choice for \((H, a)\) can be determined, for instance, by writing (A.1a) as a four
term cost function, viz.,

\[ J = E |y|^2 - [a \ H \ E \left\{ \begin{pmatrix} 1 \\ X \end{pmatrix} y^* \right\} - E \left\{ y \begin{pmatrix} 1 \\ X \end{pmatrix}^* \right\} [a \ H]^* + [a \ H \ E \left\{ \begin{pmatrix} 1 \\ X \end{pmatrix} \begin{pmatrix} 1 \\ X \end{pmatrix}^* \right\} [a \ H]^* \]

and setting the formal derivative \( \frac{\partial J}{\partial [a \ H]^*} = 0 \) (see [21] and App. B). The resulting equation is

\[ E \left\{ [y - (H_{opt}X + a_{opt})] \begin{pmatrix} 1 \\ X \end{pmatrix}^* \right\} = 0 \]  \( \text{(A.2)} \)

which we recognize as the “orthogonality principle” for the case of non-zero mean variables. The corresponding non-centered Wiener-Hopf equation is

\[ \begin{bmatrix} a_{opt} & H_{opt} \end{bmatrix} E \left\{ \begin{pmatrix} 1 \\ X \end{pmatrix} \begin{pmatrix} 1 \\ X \end{pmatrix}^* \right\} = E \left\{ y \begin{pmatrix} 1 \\ X \end{pmatrix}^* \right\} \]  \( \text{(A.3)} \)

Alternatively, we can split the orthogonality principle (A.2) into two separable equations, viz.,

\[ E \left\{ y - (H_{opt}X + a_{opt}) \right\} = 0 \]

and

\[ E \left\{ [y - (H_{opt}X + a_{opt})] X^* \right\} = 0 \]

The former provides an explicit expression for \( a_{opt} \) in terms of the (yet to be determined) \( H_{opt} \), viz.,

\[ a_{opt} = Ey - H_{opt}EX \]  \( \text{(A.4a)} \)

This relation can be used to eliminate \( a_{opt} \) from the second part of the orthogonality principle, viz.,

\[ E \{yX^*\} = H_{opt} E \{XX^*\} + a_{opt} (EX)^* \]

\[ = H_{opt} [E \{XX^*\} - (EX)(EX)^*] + (Ey)(EX)^* \]
which leads to the centered Wiener-Hopf equation

\[ E\left\{ \left( y - Ey \right) \left( X - EX \right)^* \right\} = H_{opt} E \left\{ \left( X - EX \right) \left( X - EX \right)^* \right\} \]  \hspace{1cm} (A.4b)

In summary, the general (non-zero mean) Wiener problem (A.1) can be solved either via the non-centered Wiener-Hopf equation (A.3), or via a two-step procedure in which we first determine \( H_{opt} \) via the centered Wiener-Hopf equation (A.4b), and then evaluate \( a_{opt} \) via the relation (A.4a).
Appendix B

Average autocorrelation for asymptotically-mean-stationary

The concepts of average autocorrelation and average spectrum date back to the 1960s, and are commonly used in the analysis of non-stationary signals [31, 59, 60, 41]. The average autocorrelation of a signal $x(t)$ with bounded second-order moments ($E|x(t)|^2 < \infty$) is defined as

$$c_x(m) = \langle E\{x(t + m)x^*(t)\} \rangle_t$$

When this long term average exists for every (discrete-time) $m$, we say that the random signal $x(t)$ is asymptotically-mean-stationary (AMS) [41]. The average spectrum is the Fourier transform of the average autocorrelation, viz.,

$$S_x(f) = \mathcal{F}\{c_x(m)\}$$

When $x(t)$ is (second order) stationary, then $c_x(m)$ coincides with the conventional (stationary) autocorrelation, and $S_x(f)$ coincides with the conventional power spectrum. If $x(t)$ is non-stationary but harmonizable (see, e.g. [61]), then $c_x(m)$ is always defined; and in fact it can be defined even for signals that are not harmonizable [62, 63]. In all cases (except the stationary) the average spectrum and average autocorrelation capture only part of the informa-
tion contained in the (bivariate) second order moment $E\{x(t)x^*(s)\}$. Nevertheless, this partial information is all that is needed for solving various mean-square estimation problems (see, e.g. [31, 60]).

One of the most important properties of the average autocorrelation and spectrum is their relations with linear systems as described by the (generalized) Einstein-Wiener-Khintchine theorem (see, e.g. [31]):

- The spectrum $S_x(f)$ is real-valued, nonnegative, and integrable over the frequency range $-\frac{1}{2} \leq f \leq \frac{1}{2}$.

- The average autocorrelation can be recovered from the average spectrum via the inverse Fourier relation $c_x(\tau) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_x(f) e^{j2\pi f \tau} df$

- Applying $x(t)$ as an input signal to a linear time-invariant filter $H(z)$ produces an output signal $y(t)$ that also possesses an average spectrum, say, $S_y(f)$, which satisfies the relation $S_y(f) = |H(e^{j2\pi f})|^2 S_x(f)$

This theorem shows that the relation between an average spectrum and linear time-invariant filters is the same as for the (conventional) spectrum of stationary signals. An average spectrum characterizes in a unified manner the power spectrum of both deterministic and random (stationary or non-stationary) signals. When $x(t)$ is non-random, then $c_x(m) = \langle x(t+m)x^*(t) \rangle$ is precisely the deterministic autocorrelation introduced by Wiener (and Einstein) [64].

The autocorrelation of a stationary signal can be estimated from samples of the signal by using (long-term) averaging. The slowly-varying autocorrelation of a non-stationary signal can be estimated by restricting the effective length of the signal record used for averaging, either by using a sliding window, or exponential forgetting. The length of the window used for averaging should agree with the rate of variation we expect to track—the longer the window, the slower our tracking ability. Finally, if we let the window grow to infinite length we have no tracking ability at all. In this case, all we can estimate is the average autocorrelation (see Sec. 1.4 for further detail). Notice that the average autocorrelation can be consistently estimated using the
same techniques commonly used for estimation of stationary autocorrelation [41], namely time averaging. Thus, the estimate

\[ \hat{c}_x(m) = \langle x(t + m)x^*(t) \rangle_t = \lim_{N \to \infty} \frac{1}{N + 1} \sum_{t=0}^{N} x(t + m)x^*(t) \]

has zero bias, and under certain conditions (ergodicity), also has zero variance [65].
Appendix C

Bandpass to baseband conversion

Let $x(t)$ be a continuous-time bandpass signal with frequency correspondence $X(f)$. Let $x_a(t)$ be the analytic signal (which will be defined below) with value only at positive frequency range with frequency correspondence $X_a(f)$ and $x_b(t)$ be the equivalent baseband signal with frequency correspondence $X_b(f)$. Conversion process involves two steps.

**Bandpass signal to analytic signal conversion**

This conversion relies on the notion of the *Hilbert transform*. The Hilbert transform of a signal $x(t)$ is defined as

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau = x(t) * \frac{1}{\pi t}$$

Clearly, $\hat{x}(t)$ can be considered as the output of a linear time-invariant filter with impulse response $h(t) = \frac{1}{\pi t}$ and input $x(t)$. The frequency response of this quadrature filter is

$$H(j\omega) = \begin{cases} 
-j & \omega > 0 \\
0 & \omega = 0 \\
-j & \omega < 0 
\end{cases}$$

Now we can form the complex signal

$$x_a(t) = x(t) + j\hat{x}(t)$$
known as the analytic signal associated with \( x(t) \), so that \( x(t) = Re\{x_a(t)\} \). The corresponding frequency domain transformation is

\[
X_a(f) = X(f) + jX(f)H(f) = \begin{cases} 
2X(f) & f > 0 \\
0 & f < 0 
\end{cases}
\]

where we assumed that \( X(0) = 0 \) because \( x(t) \) is a bandpass signal (see Fig. C.1).

**Analytic signal to baseband signal conversion**

By shifting the central frequency of \( x_a(t) \) to the origin we obtain a baseband representation of \( x(t) \) (see Fig. C.1), viz.,

\[
x_b(t) = x_a(t)e^{-j2\pi f_c t}, \text{ or equivalently, } X_b(f) = X_a(f + f_c)
\]

Notice that the choice of \( f_c \) is not unique: it could be any frequency within the frequency band occupied by the bandpass signal. However, it is customary to select \( f_c \) as the center frequency within this band: this has the effect of minimizing the sampling rate needed for representation of the baseband signal \( x_b(t) \).

Also notice that \( X_b(f) \) need not be (conjugate) symmetric with respect to the origin, so that \( x_b(t) \) is, in general, a complex signal. Similarly, \( x_a(t) \) is complex by construction, since its frequency content is always non-symmetric.

The combination of those two steps, \( x(t) \rightarrow x_a(t) \) and \( x_a(t) \rightarrow x_b(t) \), results in the following relations

\[
\begin{align*}
x(t) &= \text{Re}\{x_b(t)e^{j2\pi f_c t}\} \\
X(f) &= \frac{1}{2}[X_b(f - f_c) + X_b^*(-f - f_c)] \\
X(f) &= \frac{1}{2}[X_a(f) + X_a^*(-f)] \\
x(t) &= \frac{1}{2}[x_a(t) + x_a^*(t)] = \text{Re}\{x_a(t)\}
\end{align*}
\]

These two steps are shown in Fig. C.1.
Figure C.1: Bandpass to baseband conversion
Appendix D

Estimate the mean of a time-varying system impulse response

In Chapter 1, we discussed how to estimate the Doppler spectrum of a time-varying system response, which is one of the application after obtaining $C_W(m)$. One prerequisite for conducting the estimation of the Doppler spectrum is the mean value of the time-varying system response $\mu_W$ (see cite). In this appendix we present how to estimate it in two cases: $W(t)$ is assumed to be stationary or asymptotically-mean-stationary, so that $\mu_W$ is a constant vector for the purpose of estimating Doppler spectrum, we don’t consider time varying $\mu_W$ here.

When $W(t)$ is assumed to be stationary, $\mu_W$ is defined as

$$\mu_W \triangleq EW(t)$$

when $W(t)$ is assumed to be asymptotically-mean-stationary, $\mu_W$ is defined as

$$\mu_W \triangleq \langle EW(t) \rangle_t \quad (D.1)$$

Since (D.1) is a more general case, we consider it here. The estimate of $\mu_W$, the time- and ensemble-averaged mean of $W(t)$, is based on the input-output relation (1.8). This relation
implies that

\[ C_{dU}(m) = \mu_W C_U(m) \]  

(D.2a)

where

\[ C_{dU}(m) \triangleq E(d(t)U^*(t - m)) \]  

(D.2b)

The cross correlation \( C_{dU}(m) \) can be estimated using the exponentially weighted sample average

\[ \hat{C}_{dU}(m) \triangleq \frac{1 - \lambda}{1 - \lambda^{N - m}} \sum_{t=0}^{N - m - 1} \lambda^{N-k-1} d(t + m)U^*(t) \]  

(D.3)

and similarly for the matrix autocorrelation \( C_U(m) \), viz.,

\[ \hat{C}_U(m) \triangleq \frac{1 - \lambda}{1 - \lambda^{N - m}} \sum_{t=0}^{N - m - 1} \lambda^{N-k-1} U(t + m)U^*(t) \]  

(D.4)

where \( N \) indicates the length of our signal record. For our static scenario we can get the exponential weight factor \( \lambda \) almost equal to unity, say \( \lambda = 0.999 \). Since (D.2a) holds for all \( m \) values, we can reduce the effect of estimation errors by solving a least square problem, viz.,

\[
\min_{\mu_W} \sum_{m=0}^{L_1} \left\| \hat{C}_{dU}(m) - \mu_W \hat{C}_U(m) \right\|_2^2
\]  

(D.5)

for some \( L_1 \), where \( \| \cdot \|_2 \) denotes the Euclidean norm.

The quality of a least squares estimate tends, in general, to improve with the length of the summation interval (in our case \( L_1 + 1 \)). However, since the quality of the estimates \( \hat{C}_{dU}(m) \) and \( \hat{C}_U(m) \) deteriorates with increasing \( m \), it is hard to predict which value of \( L_1 \) produces the best estimate for \( \mu_W \).
Appendix E

Multichannel Levinson

The normalized version of the multichannel Levinson-Wiggin-Robinson (LWR) algorithm is described in [66, 52] as:

\[
S_W(f) = \left. \widetilde{A}_L^{-1}(z)\widetilde{A}_L^{-T}(z^{-1}) \right|_{z = e^{j2\pi f}} \\
= \left. \widetilde{B}_L^{-1}(z)\widetilde{B}_L^{-T}(z^{-1}) \right|_{z = e^{j2\pi f}} \tag{E.1}
\]

where the normalized linear prediction polynomials \( \widetilde{A}_L(z) \), \( \widetilde{B}_L(z) \) are determined by the recursion

\[
\overline{\Delta}_{l+1} = \overline{A}_{l,0}C_W(l + 1) \\
+ \overline{A}_{l,1}C_W(l) + \cdots + \overline{A}_{l,l}C_W(1) \\
\rho_{l+1} = \overline{\Delta}_{l+1} (R_l^r)^{-T/2} \\
P_{l+1} = I - \rho_{l+1}\rho_{l+1}^T \\
Q_{l+1} = I - \rho_{l+1}^T\rho_{l+1} \\
\overline{A}_{l+1}(z) = P_{l+1}^{-1/2} [\overline{A}_l(z) - z^{-1}\rho_{l+1}\overline{B}_l(z)] \\
\overline{B}_{l+1}(z) = Q_{l+1}^{-1/2} [z^{-1}\overline{B}_l(z) - \rho_{l+1}^T\overline{A}_l(z)] \\
(R_{l+1}^r)^{-1/2} = Q_{l+1}^{-1/2} (R_l^r)^{-1/2}
\]

87
with initial conditions

\[
R_0^r = R_0^c = P_0 = Q_0 = C_W(0)
\]
\[
\overline{A}_{0,0} = \overline{B}_{0,0} = (R_0^r)^{-1/2}
\]

Here \((R_l^r)^{1/2}\) denotes the Cholesky factor of the positive definite error covariance matrix \(R_l^r\), and \((R_l^r)^{-T/2} = \left[(R_l^r)^{-1/2}\right]^T\). The estimated multichannel spectrum is given by the expression (E.1).
Appendix F

Linear relation of vec\{C_W(m)\} and \(\Psi_m(t)\)

The linear relation of vec\{C_W(m)\} and \(\Psi_m(t)\) derivations are based on two equations:

\[
\begin{align*}
  d(t) &= W(t)U(t) + v(t) \quad \text{(F.1)} \\
  c_d(m) &= [\text{vec} \{C_U^*(m)\}]^* \text{vec} \{C_W(m)\} + c_v(m) \quad \text{(F.2)}
\end{align*}
\]

From the first equation (F.1) we have

\[
\begin{align*}
  d^*(t + m)d(t) &= \left[U^*(t + m)W^*(t + m) + v^*(t + m)\right] \left[W(t)U(t) + v(t)\right] \\
  &= U^*(t + m)W^*(t + m)W(t)U(t) + v^*(t + m)W(t)U(t) + U^*(t + m)W^*(t + m)v(t) + v^*(t + m)v(t) \quad \text{(F.3)}
\end{align*}
\]

From the second equation (F.2) we have

\[
E \{d^*(t + m)d(t)\} = [\text{vec} \{C_W(m)\}]^* E\Psi_m(t) + c^*_v(m)
\]

Now consider \(H \triangleq [\text{vec} \{C_W(m)\}]^*\) as a constant, from the definition of mean value we have

\[
\frac{1}{N} \sum_{t=0}^{N-1} d^*(t + m)d(t) = H \frac{1}{N} \sum_{t=0}^{N-1} \Psi_m(t) + \frac{1}{N} \sum_{t=0}^{N-1} v^*(t + m)v(t)
\]
Hence

\[ d^*(t + m)d(t) = H\Psi_m(t) + v^*(t + m)v(t) + \varepsilon(t) \]  

(F.4)

Compare (F.3) and (F.4), we conclude that

\[ \varepsilon(t) = \text{vec}\left\{ W^*(t + m)W(t + m) \right\}^*\Psi_m(t) + v^*(t + m)W(t)U(t) \]

\[ + U^*(t + m)W^*(t + m)v(t) - H\Psi_m(t) \]

\[ = \left[ \text{vec}\left\{ W^*(t)W(t + m) \right\}^* - H \right] \Psi_m(t) + v^*(t + m)W(t)U(t) \]

\[ + U^*(t + m)W^*(t + m)v(t) \]

where the last two terms of \( \varepsilon \), viz., \( v^*(t + m)W(t)U(t) \) and \( U^*(t + m)W^*(t + m)v(t) \) are uncorrelated with \( H \) and \( \Psi_m(t) \), so that can be included into \( v^*(t + m)v(t) \), while the other term, viz., \( \left[ \text{vec}\left\{ W^*(t)W(t + m) \right\}^* - H \right] \Psi_m(t) \) is the variance part of \( H \), so that is uncorrelated with the mean, hence there is linear relation of \( H \) and \( \Psi_m(t) \) with an uncorrelated noise term \( \eta(t) \), viz.,

\[ D_m(t) = H\Psi_m(t) + \eta(t) \]

where

\[ \eta(t) = \left[ \text{vec}\left\{ W^*(t)W(t + m) \right\}^* - H \right] \Psi_m(t) + v^*(t + m)W(t)U(t) \]

\[ + U^*(t + m)W^*(t + m)v(t) + v^*(t + m)v(t) \]

Since we assume our input signal \( u(\cdot) \) is stationary to the 4th order, the input signal for (2.18) \( \Psi_m(t) \) is stationary.
Bibliography


