

EE210A: Adaptation and Learning

Professor Ali H. Sayed

LECTURE #11

TRACKING PERFORMANCE

Sections in order: 20.1-20.4, 21.1 AND 21.6

TRACKING ABILITY

A fundamental feature of adaptive filters is their ability to track variations in the underlying signal statistics. This is because by relying on instantaneous data, the statistical properties of the weight vector and error signals are able to react to changes in the input signal properties. The purpose of this chapter, and the next one, is to characterize the tracking ability of adaptive filters for nonstationary environments.

20.1 MOTIVATION

In order to motivate our setup for tracking analysis, we start by reviewing the basic linear least-mean-squares estimation problem of Sec. 15.1. Thus let $\boldsymbol{d}$ and $\boldsymbol{u}$ be zero-mean random variables with second-order moments

$$\mathsf{E}\,|\boldsymbol{d}|^2 = \sigma_d^2, \qquad \mathsf{E}\,\boldsymbol{d}\boldsymbol{u}^* = R_{du}, \qquad \mathsf{E}\,\boldsymbol{u}^*\boldsymbol{u} = R_u > 0$$

The coefficient vector w^o that estimates $\boldsymbol{d}$ from $\boldsymbol{u}$ optimally in the linear least-mean-squares sense, i.e., the vector that solves

$$\min_w \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w|^2 \tag{20.1}$$

is given by $w^o = R_u^{-1}R_{du}$. The corresponding minimum cost is

$$J_{\min} = \sigma_d^2 \; - \; R_{ud}R_u^{-1}R_{du} = \mathsf{E}\,|\boldsymbol{d} - \boldsymbol{u}w^o|^2 \tag{20.2}$$

MOTIVATION

In Chapter 10 we developed stochastic-gradient algorithms (i.e., adaptive filters) for approximating w^o. These algorithms rely on data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ with moments $\{\sigma_d^2, R_{du}, R_u\}$ and use update equations of the form, say for the case of an LMS implementation,

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* \boldsymbol{e}(i), \quad \boldsymbol{w}_{-1} = \text{ initial condition}$$

or some other update form. In Chapter 15 we evaluated the performance of such filters by measuring the excess mean-square error that is left in steady-state, namely, by computing the difference

$$\text{EMSE} = \lim_{i\to\infty} \mathsf{E}\,|\boldsymbol{e}(i)|^2 - J_{\min} \tag{20.3}$$

where

$$\boldsymbol{e}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1} \tag{20.4}$$

VARYING MOMENTS

Now if the moments $\{\sigma_d^2, R_{du}, R_u\}$ vary with time, say if

$$\mathsf{E}\,|\boldsymbol{d}(i)|^2 = \sigma_{d,i}^2, \qquad \mathsf{E}\,\boldsymbol{d}(i)\boldsymbol{u}_i^* = R_{du,i}, \qquad \mathsf{E}\,\boldsymbol{u}_i^*\boldsymbol{u}_i = R_{u,i} \tag{20.5}$$

then the optimal weight vector w^o will also vary with time. Specifically, the coefficient vector for estimating $\boldsymbol{d}(i)$ from $\boldsymbol{u}_i$ in the linear least-mean-squares sense will be given by

$$w_i^o = R_{u,i}^{-1} R_{du,i} \tag{20.6}$$

with minimum cost

$$J_{\min}(i) \;=\; \sigma_{d,i}^2 - R_{ud,i}R_{u,i}^{-1}R_{du,i} \;=\; \mathsf{E}\,|\boldsymbol{d}(i) - \boldsymbol{u}_i w_i^o|^2 \tag{20.7}$$

If the $\{R_{du,i}, R_{u,i}\}$ vary slowly with time, then it is justifiable to expect that an adaptive filter will have sufficient time to track the optimal solution w_i^o. If, on the other hand, the moments $\{R_{du,i}, R_{u,i}\}$ vary rapidly with time, then this task becomes challenging (and, at times, impossible).

OBJECTIVE

The purpose of a tracking analysis is to quantify how well an adaptive filter performs under such changing conditions in the signal statistics. In order to make the analysis tractable, it is customary to assume that the statistics of the data vary in a certain manner rather than arbitrarily. For instance, the model that we shall adopt in the next section assumes that R_u and $J_{\min}$ remain fixed, while only $\{\sigma_d^2, R_{du}\}$ may vary with time.

20.2 NONSTATIONARY DATA MODEL

Thus recall from the discussion in Sec. 15.2 that any given data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ can be assumed to be related via a linear model of the form

$$\boxed{\boldsymbol{d}(i) = \boldsymbol{u}_i w_i^o + \boldsymbol{v}(i)} \tag{20.8}$$

where w_i^o is the coefficient vector (20.6) that estimates $\boldsymbol{d}(i)$ from $\boldsymbol{u}_i$ optimally in the linear least-mean squares sense. Moreover, $\boldsymbol{v}(i)$ is uncorrelated with $\boldsymbol{u}_i$ and has variance $\sigma_v^2(i) = J_{\min}(i)$. However, as was also the case in Sec. 15.2, we shall impose the stronger assumption that

$$\boxed{\{\boldsymbol{v}(i)\} \text{ is i.i.d. with constant variance } \sigma_v^2, \text{ and is independent of all } \{\boldsymbol{u}_j\}} \tag{20.9}$$

This condition on $\boldsymbol{v}(i)$ is an assumption because the signal $\boldsymbol{v}(i)$ in the model (20.8) is only uncorrelated with $\boldsymbol{u}_i$; it is not necessarily independent of $\boldsymbol{u}_i$, or of all $\{\boldsymbol{u}_j\}$ for that matter. Moreover, the variance of $\boldsymbol{v}(i)$ is not constant. Still there are situations where conditions (20.8) and (20.9) hold simultaneously, e.g., in the channel estimation application of Sec. 10.5.

RANDOM WALK MODEL

Random-Walk Model

In addition to the model (20.8)–(20.9) for the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$, we shall also adopt a model for the variations in the weight vector w_i^o. It is more convenient to adopt a model for the variation in w_i^o than a model for the variations in the statistics $\{\sigma_d^2, R_{du}\}$. One particular model that is widely used in the adaptive filtering literature is a first-order random-walk model. The model assumes that w_i^o undergoes random variations of the form

$$\boxed{\boldsymbol{w}_i^o = \boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i} \tag{20.10}$$

with $\boldsymbol{q}_i$ denoting some random perturbation that is independent of $\{\boldsymbol{u}_j, \boldsymbol{v}(j)\}$ for all i, j. Observe that we are now using boldface letters for $\{\boldsymbol{w}_i^o, \boldsymbol{w}_{i-1}^o\}$. This is because they become random variables due to the presence of the random quantity $\boldsymbol{q}_i$. The sequence $\{\boldsymbol{q}_i\}$ is assumed to be i.i.d., zero-mean, with covariance matrix

$$\boxed{\mathsf{E}\,\boldsymbol{q}_i\boldsymbol{q}_i^* = Q} \tag{20.11}$$

RANDOM WALK MODEL

It is easy to see from (20.10) that

$$\mathsf{E}\,\boldsymbol{w}_i^o \;=\; \mathsf{E}\,\boldsymbol{w}_{i-1}^o \tag{20.12}$$

so that the $\{\boldsymbol{w}_i^o\}$ have a constant mean, which we shall denote by w^o,

$$\boxed{\mathsf{E}\,\boldsymbol{w}_i^o \;\stackrel{\Delta}{=}\; w^o}$$

The initial condition for model (20.10) is modeled as a random variable $\boldsymbol{w}_{-1}^o$, with mean w^o and independent of all other variables, $\{\boldsymbol{q}_i, \boldsymbol{v}(i), \boldsymbol{u}_i\}$ for all i.

SUITABILITY OF MODEL

How Appropriate is this Model?

Although widely adopted in the literature, model (20.10) is not necessarily meaningful. For one thing, the covariance matrix of $\boldsymbol{w}_i^o$ grows unbounded. To see this, observe from

$$\boldsymbol{w}_i^o - w^o = \boldsymbol{w}_{i-1}^o - w^o + \boldsymbol{q}_i \tag{20.13}$$

that

$$\mathsf{E}\,(\boldsymbol{w}_i^o - w^o)(\boldsymbol{w}_i^o - w^o)^* \;=\; \mathsf{E}\,(\boldsymbol{w}_{i-1}^o - w^o)(\boldsymbol{w}_{i-1}^o - w^o)^* \;+\; Q$$

This means that, at each time instant i, a nonnegative-definite matrix Q is added to the covariance matrix of $\boldsymbol{w}_{i-1}^o$ in order to obtain the covariance matrix of $\boldsymbol{w}_i^o$. As a result, the covariance matrix of $\boldsymbol{w}_i^o$ becomes unbounded as time progresses. A more adequate model for tracking analysis would be to replace (20.10), or equivalently (20.13), by

$$(\boldsymbol{w}_i^o - w^o) = \alpha(\boldsymbol{w}_{i-1}^o - w^o) + \boldsymbol{q}_i \tag{20.14}$$

for some scalar $|\alpha| < 1$.

SUITABILITY OF MODEL

In this case, the covariance matrix of $\boldsymbol{w}_i^o$ would tend to a finite steady-state value given by

$$\lim_{i\to\infty} \mathsf{E}\,(\boldsymbol{w}_i^o - w^o)(\boldsymbol{w}_i^o - w^o)^* \;=\; Q/(1-|\alpha|^2)$$

Still, it is customary in the literature to assume that the value of α is sufficiently close to one, and to use model (20.10). The main reason for assuming $\alpha \approx 1$ is to simplify the derivations during the tracking analysis. It is for this reason that we shall also proceed with the simple (yet contrived) model (20.10) in the body of the chapter in order to illustrate the key concepts. However, in the problems, and especially Probs. IV.29–IV.33, we extend the analysis to models of the form (20.14), which are rewritten in Prob. IV.29 in the equivalent form

$$\left\{\begin{array}{lcl} \boldsymbol{w}_i^o & = & w^o + \boldsymbol{\theta}_i \\ \boldsymbol{\theta}_i & = & \alpha\,\boldsymbol{\theta}_{i-1} + \boldsymbol{q}_i, \quad 0 \le |\alpha| < 1 \end{array}\right. \tag{20.15}$$

Data Model

To summarize, we shall adopt the following model in our study of the tracking performance of adaptive filters in the body of the chapter. Specifically, we shall assume that the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy the following conditions:

(a) There exists a vector $\boldsymbol{w}_i^o$ such that $\boldsymbol{d}(i) = \boldsymbol{u}_i \boldsymbol{w}_i^o + \boldsymbol{v}(i)$.
(b) The weight vector varies according to $\boldsymbol{w}_i^o = \boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i$.
(c) The noise sequence $\{\boldsymbol{v}(i)\}$ is i.i.d. with constant variance $\sigma_v^2 = \mathsf{E}\,|\boldsymbol{v}(i)|^2$.
(d) The noise sequence $\{\boldsymbol{v}(i)\}$ is independent of $\boldsymbol{u}_j$ for all i, j.
(e) The sequence $\boldsymbol{q}_i$ has covariance Q and is independent of $\{\boldsymbol{v}(j), \boldsymbol{u}_j\}$ for all i, j.
(f) The initial conditions $\{\boldsymbol{w}_{-1}, \boldsymbol{w}_{-1}^o\}$ are independent of all $\{\boldsymbol{d}(j), \boldsymbol{u}_j, \boldsymbol{v}(j), \boldsymbol{q}_j\}$.
(g) The regressor covariance matrix is denoted by $R_u = \mathsf{E}\,\boldsymbol{u}_i^* \boldsymbol{u}_i > 0$.
(h) The random variables $\{\boldsymbol{d}(i), \boldsymbol{v}(i), \boldsymbol{u}_i, \boldsymbol{q}_i\}$ are zero mean.
(i) The weight vector $\boldsymbol{w}_i^o$ has constant mean w^o.

(20.16)

DATA MODEL

We refer to these conditions as describing a *nonstationary* environment. Observe that in this model, the covariance matrix of the regression data is assumed to be constant and equal to R_u; likewise for the (co-)variances of the noise components $\{\boldsymbol{v}(i), \boldsymbol{q}_i\}$. The constancy of σ_v^2 means that, although $\boldsymbol{w}_i^o$ is varying with time, the problem of estimating $\boldsymbol{d}(i)$ from $\boldsymbol{u}_i$ optimally in the linear least-mean-squares sense is such that it has a constant minimum cost for all i,

$$J_{\min} = \sigma_v^2 \tag{20.17}$$

This is because, as explained in the beginning of Sec. 15.2, $\boldsymbol{v}(i)$ plays the role of the estimation error that results from estimating $\boldsymbol{d}(i)$ from $\boldsymbol{u}_i$.

INDEPENDENCE RELATIONS

Useful Independence Results

A useful consequence of model (20.16) is that at any particular time instant i, the noise variable $\boldsymbol{v}(i)$ is independent of all previous weight estimators $\{\boldsymbol{w}_j,\ j < i\}$. This fact follows easily from examining the update equation of an adaptive filter. Consider, for instance, the LMS recursion

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* [\boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}], \quad \boldsymbol{w}_{-1} = \text{initial condition}$$

By iterating this recursion we find that, for any time instant j, the weight estimator $\boldsymbol{w}_j$ is a function of $\boldsymbol{w}_{-1}$, the reference signals $\{\boldsymbol{d}(j), \boldsymbol{d}(j-1), \ldots, \boldsymbol{d}(0)\}$, and the regressors $\{\boldsymbol{u}_j, \boldsymbol{u}_{j-1}, \ldots, \boldsymbol{u}_0\}$. We can represent this dependency generically as

$$\boldsymbol{w}_j = \mathcal{F}\,[\,\boldsymbol{w}_{-1};\ \boldsymbol{d}(j), \ldots, \boldsymbol{d}(0);\ \boldsymbol{u}_j, \ldots, \boldsymbol{u}_0\,] \tag{20.18}$$

for some function $\mathcal{F}$. A similar dependency holds for other adaptive schemes.

INDEPENDENCE RELATIONS

Now $\boldsymbol{v}(i)$ is independent of each one of the terms appearing as an argument of $\mathcal{F}$ in (20.18) so that $\boldsymbol{v}(i)$ is independent of $\boldsymbol{w}_j$ for all $j < i$. The independence of $\boldsymbol{v}(i)$ from $\{\boldsymbol{w}_{-1}, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_0\}$ is obvious by assumption, while its independence from $\{\boldsymbol{d}(j), \ldots, \boldsymbol{d}(0)\}$ can be seen as follows. Consider $\boldsymbol{d}(j)$ for example. Then from

$$\boldsymbol{d}(j) = \boldsymbol{u}_j \boldsymbol{w}_j^o + \boldsymbol{v}(j) \;=\; \boldsymbol{u}_j \left(\boldsymbol{w}_{-1}^o + \sum_{k=0}^{j} \boldsymbol{q}_k \right) \;+\; \boldsymbol{v}(j)$$

we see that $\boldsymbol{d}(j)$ is a function of $\{\boldsymbol{u}_j, \boldsymbol{v}(j), \boldsymbol{w}_{-1}^o, \boldsymbol{q}_0, \ldots, \boldsymbol{q}_j\}$, all of which are independent of $\boldsymbol{v}(i)$ for $j < i$. We therefore conclude that $\boldsymbol{v}(i)$ is independent of $\boldsymbol{w}_j$ for all $j < i$.

It is also immediate to verify that $\boldsymbol{v}(i)$ is independent of $\tilde{\boldsymbol{w}}_j$ for all $j < i$, where $\tilde{\boldsymbol{w}}_j$ denotes the weight-error vector that is *now* defined in terms of the time-variant weight vector $\boldsymbol{w}_j^o$,

$$\boxed{\tilde{\boldsymbol{w}}_j \;\stackrel{\Delta}{=}\; \boldsymbol{w}_j^o - \boldsymbol{w}_j}$$

INDEPENDENCE RELATIONS

It is also immediate to verify that $\boldsymbol{v}(i)$ is independent of $\tilde{\boldsymbol{w}}_j$ for all $j < i$, where $\tilde{\boldsymbol{w}}_j$ denotes the weight-error vector that is *now* defined in terms of the time-variant weight vector w_j^o,

$$\tilde{\boldsymbol{w}}_j \triangleq w_j^o - \boldsymbol{w}_j$$

It also follows that $\boldsymbol{v}(i)$ is independent of the *a priori* estimation error $\boldsymbol{e}_a(i)$, which in the nonstationary case is defined as

$$\boldsymbol{e}_a(i) \triangleq \boldsymbol{u}_i w_i^o - \boldsymbol{u}_i \boldsymbol{w}_{i-1}$$

This definition is consistent with our earlier definition in the stationary case in Sec. 15.2, namely, $\boldsymbol{e}_a(i) = \boldsymbol{u}_i w^o - \boldsymbol{u}_i \boldsymbol{w}_{i-1}$. In both cases, $\boldsymbol{e}_a(i)$ is a measure of the error in estimating the *uncorrupted* part of $\boldsymbol{d}(i)$. However, observe that in the nonstationary case, we cannot write $\boldsymbol{e}_a(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1}$ since $\tilde{\boldsymbol{w}}_{i-1} = w_{i-1}^o - \boldsymbol{w}_{i-1}$.

INDEPENDENCE RELATIONS

Lemma 20.1 (Useful properties) From the data model (20.16), it holds that $\boldsymbol{v}(i)$ is independent of each of the following:

$$\{\boldsymbol{w}_j \ \text{ for } j < i\}, \ \ \{\tilde{\boldsymbol{w}}_j = w_j^o - \boldsymbol{w}_j \ \text{ for } j < i\}, \ \ \text{and} \ \ \boldsymbol{e}_a(i) = \boldsymbol{u}_i(w_i^o - \boldsymbol{w}_{i-1})$$

EMSE EXPRESSION

Alternative Expression for the EMSE

Using model (20.16), and the result of Lemma 20.1, we can express the filter EMSE in an alternative form. Indeed, using $\boldsymbol{e}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}$, and part (a) of model (20.16), we have $\boldsymbol{e}(i) = \boldsymbol{v}(i) + \boldsymbol{u}_i(\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1})$, i.e.,

$$\boxed{\boldsymbol{e}(i) = \boldsymbol{v}(i) + \boldsymbol{e}_a(i)} \tag{20.19}$$

Now since $\boldsymbol{v}(i)$ and $\boldsymbol{e}_a(i)$ are independent, and $\boldsymbol{v}(i)$ has zero mean, it follows from (20.19) that

$$\mathsf{E}\,|\boldsymbol{e}(i)|^2 \;=\; \mathsf{E}\,|\boldsymbol{v}(i)|^2 \;+\; \mathsf{E}\,|\boldsymbol{e}_a(i)|^2$$

The first term on the right-hand side is σ_v^2 which, as explained above in (20.17), coincides with $J_{\min}$ and, hence,

$$\mathsf{E}\,|\boldsymbol{e}(i)|^2 - J_{\min} = \mathsf{E}\,|\boldsymbol{e}_a(i)|^2$$

Substituting this equality into definition (20.3) for the EMSE we arrive at the equivalent characterization

$$\boxed{\text{EMSE} = \lim_{i\to\infty} \mathsf{E}\,|\boldsymbol{e}_a(i)|^2} \tag{20.20}$$

Degree of Nonstationarity

A lower bound on the EMSE can be determined as follows. Using $\boldsymbol{w}_i^o = \boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i$ we have

$$\begin{aligned}
\boldsymbol{e}_a(i) &= \boldsymbol{u}_i \boldsymbol{w}_i^o - \boldsymbol{u}_i \boldsymbol{w}_{i-1} \\
&= \boldsymbol{u}_i(\boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1} \\
&= \boldsymbol{u}_i(\boldsymbol{w}_{i-1}^o - \boldsymbol{w}_{i-1}) + \boldsymbol{u}_i \boldsymbol{q}_i
\end{aligned}$$

so that

$$\begin{aligned}
\mathsf{E}\,|\boldsymbol{e}_a(i)|^2 &= \mathsf{E}\left|\boldsymbol{u}_i(\boldsymbol{w}_{i-1}^o - \boldsymbol{w}_{i-1}) + \boldsymbol{u}_i \boldsymbol{q}_i\right|^2 \\
&= \mathsf{E}\left|\boldsymbol{u}_i(\boldsymbol{w}_{i-1}^o - \boldsymbol{w}_{i-1})\right|^2 \;+\; \mathsf{E}\,|\boldsymbol{u}_i \boldsymbol{q}_i|^2 \\
&\geq \mathsf{E}\,|\boldsymbol{u}_i \boldsymbol{q}_i|^2 \\
&= \mathsf{Tr}(R_u Q), \qquad \text{for all } i
\end{aligned}$$

where in the second equality we used the fact that $\boldsymbol{q}_i$ is independent of $\boldsymbol{u}_i(\boldsymbol{w}_{i-1}^o - \boldsymbol{w}_{i-1})$ and, hence, their cross-correlation is zero. In the last equality we used the fact that $\boldsymbol{q}_i$ and $\boldsymbol{u}_i$ are independent.

DEGREE OF NONSTATIONARITY

We therefore find that the misadjustment of an adaptive filter in a nonstationary environment is lower bounded by

$$\boxed{\mathcal{M} \geq \mathsf{Tr}(R_u Q)/\sigma_v^2}$$

The ratio on the right-hand side, involving $\{R_u, Q, \sigma_v^2\}$, is equal to the square of what is called the *degree of nonstationarity* (DN) of the data,

$$\boxed{\mathsf{DN} \stackrel{\Delta}{=} \sqrt{\mathsf{Tr}(R_u Q)/\sigma_v^2}}$$

If the value of DN is larger than unity then this means that the statistical variations in the optimal weight vector w_i^o are too fast for the filter to be able to track them (and the misadjustment will be large). On the other hand, if DN $\ll 1$, then the adaptive filter would generally be able to track the variations in the weight vector. In this chapter, we are interested in evaluating the tracking performance of adaptive filters in this latter situation, i.e., when tracking is possible.

ERROR QUANTITIES

TABLE 20.1 Definitions of several estimation errors.

Error	Definition	Interpretation
$e(i)$	$d(i) - u_i w_{i-1}$	*a priori* output estimation error
$r(i)$	$d(i) - u_i w_i$	*a posteriori* output estimation error
$\tilde{w}_i$	$w_i^o - w_i$	weight estimation error
$e_a(i)$	$u_i w_i^o - u_i w_{i-1}$	*a priori* estimation error
$e_p(i)$	$u_i w_i^o - u_i w_i$	*a posteriori* estimation error

20.3 ENERGY CONSERVATION RELATION

We again consider adaptive filters whose update equations are of the form

$$\boxed{\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* g[\boldsymbol{e}(i)], \quad \boldsymbol{w}_{-1} = \text{initial condition}} \tag{20.21}$$

where $g[\cdot]$ denotes the error function; several examples were listed in Table 15.2.

The update recursion (20.21) can be written in terms of the weight-error vector $\tilde{\boldsymbol{w}}_i = \boldsymbol{w}_i^o - \boldsymbol{w}_i$. Subtracting both sides of (20.21) from $\boldsymbol{w}_i^o$ gives

$$\boxed{\boldsymbol{w}_i^o - \boldsymbol{w}_i = (\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}) - \mu \boldsymbol{u}_i^* g[\boldsymbol{e}(i)]} \tag{20.22}$$

Multiplying both sides of this equation by $\boldsymbol{u}_i$ from the left we find that the *a priori* and *a posteriori* errors $\{\boldsymbol{e}_p(i), \boldsymbol{e}_a(i)\}$ are related via

$$\boxed{\boldsymbol{e}_p(i) = \boldsymbol{e}_a(i) - \mu \|\boldsymbol{u}_i\|^2 g[\boldsymbol{e}(i)]} \tag{20.23}$$

ENERGY CONSERVATION

Theorem 20.1 (Energy-conservation relation) For any adaptive filter of the form (20.21), and for any data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$, it holds that

$$\|\boldsymbol{w}_i^o - \boldsymbol{w}_i\|^2 + \bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2 + \bar{\mu}(i)|\boldsymbol{e}_p(i)|^2 \quad (20.24)$$

where $\boldsymbol{e}_a(i) = \boldsymbol{u}_i(\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1})$, $\boldsymbol{e}_p(i) = \boldsymbol{u}_i(\boldsymbol{w}_i^o - \boldsymbol{w}_i)$, and $\bar{\mu}(i)$ is defined as

$$\bar{\mu}(i) \triangleq (\|\boldsymbol{u}_i\|^2)^\dagger = \begin{cases} 1/\|\boldsymbol{u}_i\|^2 & \text{if } \boldsymbol{u}_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Comparing (20.24) with the energy-conservation relation (15.32) in the stationary case, we see that the only difference pertains to the interpretation of the terms $\boldsymbol{w}_i^o - \boldsymbol{w}_i$ and $\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}$ that appear on both sides of (20.24). While the first difference can be recognized as $\tilde{\boldsymbol{w}}_i$, just like the term on the left-hand side of (15.32), the second difference is *not* $\tilde{\boldsymbol{w}}_{i-1}$ since, in the nonstationary case, $\tilde{\boldsymbol{w}}_{i-1}$ is defined as $\tilde{\boldsymbol{w}}_{i-1} = \boldsymbol{w}_{i-1}^o - \boldsymbol{w}_{i-1}$ (in terms of $\boldsymbol{w}_{i-1}^o$ and not $\boldsymbol{w}_i^o$).

20.4 VARIANCE RELATION

In order to explain the relevance of the energy relation (20.24) to the tracking analysis of adaptive filters, we refer to the data model (20.16) and, in particular, to condition (b). The condition states that $\boldsymbol{w}_i^o$ varies according to the random-walk model

$$\boldsymbol{w}_i^o = \boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i \qquad (20.25)$$

where $\boldsymbol{q}_i$ is an i.i.d. sequence with covariance matrix Q and is independent of the initial conditions $\{\boldsymbol{w}_{-1}^o, \boldsymbol{w}_{-1}\}$, of $\{\boldsymbol{u}_j\}$ for all j, and of $\{\boldsymbol{d}(j)\}$ for all $j < i$. This random-walk model, and the assumptions on $\boldsymbol{q}_i$, are the only conditions that we require from the data model (20.16) for the derivation in this section. [The conditions on $\boldsymbol{v}(i)$, such as its independence of all $\boldsymbol{u}_j$ and $\boldsymbol{q}_j$, are not needed here.]

Taking expectations of both sides of the energy-conservation relation (20.24) we get

$$\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2 \;+\; \mathsf{E}\,\bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \mathsf{E}\,\|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2 \;+\; \mathsf{E}\,\bar{\mu}(i)|\boldsymbol{e}_p(i)|^2 \qquad (20.26)$$

since $\tilde{\boldsymbol{w}}_i = \boldsymbol{w}_i^o - \boldsymbol{w}_i$.

VARIANCE RELATION

Moreover, the model (20.25) allows us to relate $\mathsf{E}\,\|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2$ to $\mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2$ as follows:

$$\begin{aligned}
\mathsf{E}\,\|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2 &= \mathsf{E}\,\|\boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i - \boldsymbol{w}_{i-1}\|^2 \\
&= \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1} + \boldsymbol{q}_i\|^2 \\
&= \mathsf{E}\,(\tilde{\boldsymbol{w}}_{i-1} + \boldsymbol{q}_i)^*(\tilde{\boldsymbol{w}}_{i-1} + \boldsymbol{q}_i) \\
&= \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2 \;+\; \mathsf{E}\,\|\boldsymbol{q}_i\|^2 \;+\; \mathsf{E}\,\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{q}_i \;+\; \mathsf{E}\,\boldsymbol{q}_i^*\tilde{\boldsymbol{w}}_{i-1} \qquad (20.27)
\end{aligned}$$

We can be more explicit about the cross terms $\mathsf{E}\,\tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{q}_i$ and $\mathsf{E}\,\boldsymbol{q}_i^*\tilde{\boldsymbol{w}}_{i-1}$. Indeed, note that

$$\tilde{\boldsymbol{w}}_{i-1} = \boldsymbol{w}_{i-1}^o - \boldsymbol{w}_{i-1} \;=\; \left(\boldsymbol{w}_{-1}^o + \sum_{j=0}^{i-1}\boldsymbol{q}_j\right) \;-\; \boldsymbol{w}_{i-1}$$

VARIANCE RELATION

so that

$$
\begin{aligned}
\mathsf{E}\,\tilde{\boldsymbol{w}}_{i-1}^{*}\boldsymbol{q}_i &= \mathsf{E}\left(w_{-1}^{o}+\sum_{j=0}^{i-1}\boldsymbol{q}_j\right)^{*}\boldsymbol{q}_i \;-\; \mathsf{E}\,\boldsymbol{w}_{i-1}^{*}\boldsymbol{q}_i \\
&= -\mathsf{E}\,\boldsymbol{w}_{i-1}^{*}\boldsymbol{q}_i
\end{aligned}
$$

where we used the fact that $\boldsymbol{q}_i$ is independent of all previous $\boldsymbol{q}_j$ and of w_{-1}^{o}. But since $\boldsymbol{w}_{i-1}$ is a function of the variables

$$\{\, \boldsymbol{u}_{i-1},\ldots,\boldsymbol{u}_0,\; \boldsymbol{d}(i-1),\ldots,\boldsymbol{d}(0)\,\}$$

all of which are independent of $\boldsymbol{q}_i$, we conclude that

$$\mathsf{E}\,\tilde{\boldsymbol{w}}_{i-1}^{*}\boldsymbol{q}_i = 0$$

VARIANCE RELATION

Likewise, $\mathsf{E}\, \boldsymbol{q}_i^* \tilde{\boldsymbol{w}}_{i-1} = 0$. Substituting into (20.27) we find that

$$\mathsf{E}\, \|\boldsymbol{w}_i^o - \boldsymbol{w}_{i-1}\|^2 = \mathsf{E}\, \|\tilde{\boldsymbol{w}}_{i-1}\|^2 \; + \; \mathsf{Tr}(Q) \tag{20.28}$$

and (20.26) becomes

$$\boxed{\mathsf{E}\, \|\tilde{\boldsymbol{w}}_i\|^2 \; + \; \mathsf{E}\, \bar{\mu}(i) |\boldsymbol{e}_a(i)|^2 = \mathsf{E}\, \|\tilde{\boldsymbol{w}}_{i-1}\|^2 \; + \; \mathsf{E}\, \bar{\mu}(i) |\boldsymbol{e}_p(i)|^2 \; + \; \mathsf{Tr}(Q)} \tag{20.29}$$

Comparing with relation (15.36) in the stationary case, we see that the only difference is the appearance of the additional term $\mathsf{Tr}(Q)$. All other terms are identical!

Steady-State Performance

Now assume that an adaptive filter is operating in steady-state (cf. Def. 15.1), i.e.,

$$\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2 = \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2, \quad \text{as} \;\; i \longrightarrow \infty \tag{20.30}$$

It then follows from (20.29) that

$$\mathsf{E}\,\bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \mathsf{Tr}(Q) \;+\; \mathsf{E}\,\bar{\mu}(i)|\boldsymbol{e}_p(i)|^2, \quad \text{as } i \longrightarrow \infty$$

Using (20.23), we can replace $\boldsymbol{e}_p(i)$ in terms of $\boldsymbol{e}_a(i)$ and get

$$\mathsf{E}\,\bar{\mu}(i)|\boldsymbol{e}_a(i)|^2 = \mathsf{Tr}(Q) \;+\; \mathsf{E}\,\bar{\mu}(i)\left|\boldsymbol{e}_a(i) - \mu\|\boldsymbol{u}_i\|^2 g[\boldsymbol{e}(i)]\right|^2, \quad \text{as } i \longrightarrow \infty \tag{20.31}$$

This relation can be simplified by expanding both of its sides, and by following the same steps that we carried out for (15.38), thus leading to the following conclusion.

STEADY-STATE PERFORMANCE

Theorem 20.2 (Variance relation) Consider any adaptive filter of the form (20.21), and assume filter operation in steady-state. Assume further that

$$\boldsymbol{d}(i) = \boldsymbol{u}_i \boldsymbol{w}_i^o + \boldsymbol{v}(i)$$

where $\boldsymbol{w}_i^o$ varies according to the random-walk model $\boldsymbol{w}_i^o = \boldsymbol{w}_{i-1}^o + \boldsymbol{q}_i$ and $\boldsymbol{q}_i$ is a zero-mean i.i.d. sequence with covariance matrix Q. Moreover, $\boldsymbol{q}_i$ is independent of $\{\boldsymbol{d}(j), j < i\}$ and of $\{\boldsymbol{u}_j, \boldsymbol{w}_{-1}^o\}$ for all j. Then it holds that

$$\mu \mathsf{E}\, \|\boldsymbol{u}_i\|^2 |g[\boldsymbol{e}(i)]|^2 + \mu^{-1}\mathsf{Tr}(Q) = 2\mathsf{Re}\big(\mathsf{E}\, \boldsymbol{e}_a^*(i) g[\boldsymbol{e}(i)]\big), \quad \text{as } i \to \infty \qquad (20.32)$$

where $\boldsymbol{e}(i) = \boldsymbol{e}_a(i) + \boldsymbol{v}(i)$. For real-valued data, the above relation becomes

$$\mu \mathsf{E}\, \|\boldsymbol{u}_i\|^2 g^2[\boldsymbol{e}(i)] \;+\; \mu^{-1}\mathsf{Tr}(Q) = 2\mathsf{E}\, \boldsymbol{e}_a(i) g[\boldsymbol{e}(i)], \quad \text{as } i \to \infty \qquad (20.33)$$

TRACKING BY LMS

$$\mu \mathsf{E}\,\|\boldsymbol{u}_i\|^2 |g[\boldsymbol{e}(i)]|^2 + \mu^{-1}\mathsf{Tr}(Q) = 2\mathsf{Re}\big(\mathsf{E}\,\boldsymbol{e}_a^*(i) g[\boldsymbol{e}(i)]\big), \quad \text{as } i \to \infty \tag{20.32}$$

21.1 PERFORMANCE OF LMS

We start with the simplest of algorithms, namely, LMS. Thus assume that $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy model (20.16) and consider the LMS recursion

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* \boldsymbol{e}(i) \tag{21.1}$$

for which

$$g[\boldsymbol{e}(i)] = \boldsymbol{e}(i) = \boldsymbol{e}_a(i) + \boldsymbol{v}(i) \tag{21.2}$$

Relation (20.32) then becomes

$$\mu \mathsf{E}\,\|\boldsymbol{u}_i\|^2 |\boldsymbol{e}_a(i) + \boldsymbol{v}(i)|^2 \; + \; \mu^{-1}\mathsf{Tr}(Q) = 2\mathsf{Re}\big(\mathsf{E}\,\boldsymbol{e}_a^*(i)[\boldsymbol{e}_a(i) + \boldsymbol{v}(i)]\big) \tag{21.3}$$

Except for the term $\mu^{-1}\mathsf{Tr}(Q)$, this identity has the same form as the identity (16.3) that appeared in our study of the mean-square performance of LMS in Chapter 16.

TRACKING BY LMS

Therefore, performing the same expansions that we did in that section following (16.3), we can readily verify that (21.3) leads to

$$\zeta^{\mathsf{LMS}} = \frac{1}{2}\left[\ \mu \mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 \ + \ \mu\sigma_v^2\mathsf{Tr}(R_u) + \mu^{-1}\mathsf{Tr}(Q)\ \right], \quad \text{as} \quad i \to \infty \qquad (21.4)$$

This expression extends the result (16.5) to the nonstationary case. In order to evaluate $\zeta^{\mathsf{LMS}} = \mathsf{E}\,|\boldsymbol{e}_a(\infty)|^2$, we again examine three cases (the similarities with the arguments in Sec. 16 are obvious).

SMALL STEP-SIZES

$$\zeta^{\mathsf{LMS}} = \frac{1}{2}\left[\ \mu\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2\ +\ \mu\sigma_v^2\mathsf{Tr}(R_u) + \mu^{-1}\mathsf{Tr}(Q)\ \right], \quad \text{as}\ \ i \to \infty \tag{21.4}$$

Small Step-Sizes

Assume first that the step-size μ is such that, in *steady-state*, the term $\mu\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ can be neglected compared to the term $\mu\sigma_v^2\mathsf{Tr}(R_u)+\mu^{-1}\mathsf{Tr}(Q)$. This condition occurs for data with a sufficiently small degree of nonstationarity and for step-sizes μ in the vicinity of $\mu_{\text{opt}}^{\mathsf{LMS}}$ in (21.6). Then, expression (21.4) gives

$$\zeta^{\mathsf{LMS}} = \frac{\mu\sigma_v^2\mathsf{Tr}(R_u) + \mu^{-1}\mathsf{Tr}(Q)}{2} \tag{21.5}$$

This result highlights the effect of the step-size on the performance of LMS. The term $\mu\sigma_v^2\mathsf{Tr}(R_u)$ is the same one we encountered earlier in expression (16.6) while studying the EMSE of LMS in stationary environments. The additional term $\mu^{-1}\mathsf{Tr}(Q)$ reflects the effect of the nonstationarity in the weight vector $\boldsymbol{w}_i^o$ on filter performance. Observe in particular that $\mathsf{Tr}(Q)$ appears multiplied by μ^{-1} so that the larger the step-size the smaller the effect of the nonstationarity on the EMSE

OPTIMAL STEP-SIZE

This discussion suggests that there exists a compromise choice for the step-size, which is obtained by minimizing (21.5) with respect to μ. Setting the derivative of ζ^{LMS} equal to zero gives

$$\boxed{\mu_{\text{opt}}^{\mathsf{LMS}} = \sqrt{\mathsf{Tr}(Q)/\sigma_v^2 \mathsf{Tr}(R_u)}} \qquad (21.6)$$

Substituting the above optimal value for μ into (21.5) we find that the resulting minimum EMSE is given by

$$\boxed{\zeta_{\min}^{\mathsf{LMS}} = \sqrt{\sigma_v^2 \mathsf{Tr}(R_u) \mathsf{Tr}(Q)}} \qquad (21.7)$$

Separation Principle

Rather than neglect the effect of the term $\mu \mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ in steady-state, we can call upon the separation assumption (16.7) that we introduced earlier in Sec. 16.3, namely, that

$$\boxed{\text{At steady-state,}\quad \|\boldsymbol{u}_i\|^2 \quad \text{is independent of} \quad \boldsymbol{e}_a(i)} \tag{21.8}$$

Using this assumption we have

$$\mathsf{E}\left(\|\boldsymbol{u}_i\|^2 \cdot |\boldsymbol{e}_a(i)|^2\right) = \left(\mathsf{E}\,\|\boldsymbol{u}_i\|^2\right) \cdot \left(\mathsf{E}\,|\boldsymbol{e}_a(i)|^2\right) = \mathsf{Tr}(R_u)\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$$

so that substituting into (21.4) we obtain

$$\boxed{\zeta^{\mathsf{LMS}} = \frac{\mu\sigma_v^2\mathsf{Tr}(R_u) + \mu^{-1}\mathsf{Tr}(Q)}{2 - \mu\mathsf{Tr}(R_u)}} \tag{21.9}$$

SEPARATION PRINCIPLE

Differentiating (21.9) with respect to μ leads to the following expression for the optimal step-size (see Prob. IV.22):

$$\mu_{\text{opt}}^{\text{LMS}} = \sqrt{\frac{\text{Tr}(Q)}{\sigma_v^2 \text{Tr}(R_u)} + \frac{\left(\text{Tr}(Q)\right)^2}{4\sigma_v^4}} - \frac{\text{Tr}(Q)}{2\sigma_v^2} \tag{21.10}$$

Substituting into (21.9) we can find the minimum value for the EMSE. We forgo this calculation here.

WHITE GAUSSIAN INPUT

As discussed in Sec. 16.4, one particular case for which the term $\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2$ that appears in (21.4) can be evaluated in closed-form occurs when the regressor $\boldsymbol{u}_i$ has a circular Gaussian distribution with a diagonal covariance matrix of the form

$$R_u = \sigma_u^2 \mathrm{I}, \quad \sigma_u^2 > 0 \tag{21.11}$$

The diagonal structure of R_u means that the entries of $\boldsymbol{u}_i$ are uncorrelated among themselves and that each has variance σ_u^2. In addition to (21.11), we also assume that (see the justification in Sec. 16.4):

$$\boxed{\text{At steady state, } \tilde{\boldsymbol{w}}_{i-1} \text{ is independent of } \boldsymbol{u}_i} \tag{21.12}$$

Under conditions (21.11)–(21.12), we showed in Sec. 16.4 that

$$\mathsf{E}\,\|\boldsymbol{u}_i\|^2|\boldsymbol{e}_a(i)|^2 = (M+\gamma)\sigma_u^2\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$$

where $\gamma = 2$ for real-valued data and $\gamma = 1$ for complex-valued data — see (16.17) and (16.19). Substituting into (21.4) we obtain

$$\boxed{\zeta^{\mathsf{LMS}} = \frac{\mu M \sigma_v^2 \sigma_u^2 \;+\; \mu^{-1}\mathsf{Tr}(Q)}{2-\mu(M+\gamma)\sigma_u^2}} \tag{21.13}$$

PERFORMANCE OF LMS

Lemma 21.1 (Tracking EMSE of LMS) Consider the LMS algorithm (21.1) and assume $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy the nonstationary model (20.16) with a sufficiently small degree of nonstationarity. Then its EMSE can be approximated by the following expressions:

1. For small step-sizes, it holds that

$$\zeta^{\text{LMS}} = \frac{\mu\sigma_v^2 \text{Tr}(R_u) + \mu^{-1}\text{Tr}(Q)}{2}$$

$$\mu_{\text{opt}}^{\text{LMS}} = \sqrt{\frac{\text{Tr}(Q)}{\sigma_v^2 \text{Tr}(R_u)}} \quad \text{with} \quad \zeta_{\min}^{\text{LMS}} = \sqrt{\sigma_v^2 \text{Tr}(R_u)\text{Tr}(Q)}$$

2. Under the separation assumption (21.8), it holds that

$$\zeta^{\text{LMS}} = \frac{\mu\sigma_v^2 \text{Tr}(R_u) + \mu^{-1}\text{Tr}(Q)}{2 - \mu\text{Tr}(R_u)}$$

$$\mu_{\text{opt}}^{\text{LMS}} = \sqrt{\frac{\text{Tr}(Q)}{\sigma_v^2 \text{Tr}(R_u)} + \frac{\left(\text{Tr}(Q)\right)^2}{4\sigma_v^4}} - \frac{\text{Tr}(Q)}{2\sigma_v^2}$$

3. If $\boldsymbol{u}_i$ is Gaussian with $R_u = \sigma_u^2 \mathrm{I}$, and under assumption (21.12), it holds that ζ^{LMS} and $\mu_{\text{opt}}^{\text{LMS}}$ are given by (21.13) and (21.14), respectively, where $\gamma = 2$ if the data are real-valued and $\gamma = 1$ if the data is complex-valued with $\boldsymbol{u}_i$ circular. Here M is the dimension of $\boldsymbol{u}_i$.

In all cases, the misadjustment is obtained by dividing the EMSE by σ_v^2. Also, substituting the expressions for μ_{opt} into the expressions for EMSE we find the corresponding optimal EMSE.

PERFORMANCE OF NLMS

Lemma 21.2 (Tracking EMSE of ϵ-NLMS) Consider the ϵ-NLMS recursion (21.15) and assume $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy the nonstationary model (20.16) with a sufficiently small degree of nonstationarity. Then, under the separation assumption (21.8) and for small ϵ, its EMSE can be approximated by

$$\zeta^{\epsilon-\text{NLMS}} = \frac{1}{2-\mu}\left[\mu\sigma_v^2 + \frac{\mu^{-1}\text{Tr}(Q)}{\text{E}\left(1/\|\boldsymbol{u}_i\|^2\right)}\right]$$

or, under the steady-state approximation (17.8),

$$\zeta^{\epsilon-\text{NLMS}} = \frac{\text{Tr}(R_u)}{2-\mu}\left[\mu\sigma_v^2\text{E}\left(\frac{1}{\|\boldsymbol{u}_i\|^2}\right) + \mu^{-1}\text{Tr}(Q)\right]$$

Optimal choices for the step-size parameter, and the resulting minimum EMSE for these approximations, are given in Prob. IV.23. The misadjustment is obtained by dividing the EMSE by σ_v^2.

Lemma 21.3 (Tracking EMSE of sign-error LMS) Consider the sign-error LMS recursion (21.22) and assume the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy model (20.16) with a sufficiently small degree of nonstationarity. Assume also that $\{\boldsymbol{v}(i), \boldsymbol{u}_i\}$ are Gaussian and the step-size is small. Then its EMSE as a function of the step-size can be approximated by

$$\zeta^{\text{sign-error LMS}} = \frac{\alpha}{2}\left(\alpha + \sqrt{\alpha^2 + 4\sigma_v^2}\right)$$

where

$$\alpha = \sqrt{\frac{\pi}{8\gamma}}\left[\mu\gamma\text{Tr}(R_u) + \mu^{-1}\text{Tr}(Q)\right]$$

with $\gamma = 1$ for real-valued data and $\gamma = 2$ for complex-valued data. The optimal step-size, and the resulting minimum EMSE, are given by

$$\mu_{\text{opt}}^{\text{sign-error LMS}} = \sqrt{\frac{\text{Tr}(Q)}{\gamma\text{Tr}(R_u)}}$$

$$\zeta_{\min}^{\text{sign-error LMS}} = \frac{\pi}{4}\text{Tr}(R_u)\text{Tr}(Q)\left(1 + \sqrt{1 + \frac{8\sigma_v^2}{\pi\text{Tr}(R_u)\text{Tr}(Q)}}\right)$$

The misadjustment is obtained by dividing the EMSE by σ_v^2.

PERFORMANCE OF RLS

Lemma 21.4 (Tracking EMSE of RLS) Consider the RLS recursion (21.31)–(21.30) and assume $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy the nonstationary model (20.16) with a sufficiently small degree of nonstationarity. Introduce the approximations (21.38) and (21.45). Then, under the separation condition (21.44), the tracking EMSE of RLS can be approximated by

$$\zeta^{\text{RLS}} = \frac{\sigma_v^2(1-\lambda)M \;+\; \frac{1}{(1-\lambda)}\text{Tr}(QR_u)}{2-(1-\lambda)M}$$

The misadjustment is obtained by dividing the EMSE by σ_v^2. Assuming $(1-\lambda)$ is small so that $2-(1-\lambda)M \approx 2$, the optimal choice of λ that results in minimal EMSE and the minimum EMSE are given by

$$\lambda_{\text{opt}}^{\text{RLS}} = 1 - \frac{1}{\sigma_v}\sqrt{\frac{\text{Tr}(QR_u)}{M}} \quad \text{and} \quad \zeta_{\min}^{\text{RLS}} = \sigma_v\sqrt{M\,\text{Tr}(QR_u)}$$

21.6 COMPARING RLS AND LMS

Although the convergence performance of RLS is significantly superior to that of LMS, it does not necessarily follow that the tracking performance of RLS is similarly superior to that of LMS. Actually, there are situations where one algorithm supersedes the other and vice-versa, so much so that a general statement about how their tracking behaviors relate to each other is difficult to make.

To illustrate this fact, recall from Lemma 21.1 that the excess mean-square error of LMS in nonstationary environments, and for sufficiently small step-sizes, can be approximated by

$$\zeta^{\text{LMS}} = \frac{\mu\sigma_v^2\text{Tr}(R_u) + \mu^{-1}\text{Tr}(Q)}{2}$$

with the corresponding optimal choice for the step-size and minimum achievable EMSE given by

$$\mu_{\text{opt}}^{\text{LMS}} = \sqrt{\text{Tr}(Q)/\sigma_v^2\text{Tr}(R_u)}, \qquad \zeta_{\min}^{\text{LMS}} = \sqrt{\sigma_v^2\text{Tr}(R_u)\text{Tr}(Q)}$$

COMPARING RLS AND LMS

Likewise, from Lemma 21.4 we have that the excess mean-square error of RLS in nonstationary environments, and for forgetting factors that are sufficiently close to one, can be approximated by

$$\zeta^{\mathsf{RLS}} = \frac{\sigma_v^2(1-\lambda)M}{2} + \frac{1}{2(1-\lambda)}\mathsf{Tr}(QR_u)$$

with the corresponding optimal choice for the forgetting factor and minimum achievable EMSE given by

$$\lambda_{\text{opt}}^{\mathsf{RLS}} = 1 - \sqrt{\mathsf{Tr}(QR_u)/\sigma_v^2 M}, \qquad \zeta_{\min}^{\mathsf{RLS}} = \sqrt{\sigma_v^2 M\,\mathsf{Tr}(QR_u)}$$

It follows that

$$\boxed{\frac{\zeta_{\min}^{\mathsf{RLS}}}{\zeta_{\min}^{\mathsf{LMS}}} = \sqrt{\frac{M\,\mathsf{Tr}(QR_u)}{\mathsf{Tr}(R_u)\mathsf{Tr}(Q)}}}$$

from which is it seen that the performance of RLS and LMS will be similar whenever R_u or Q is a multiple of the identity matrix. For other choices of $\{Q, R_u\}$, one algorithm may perform better than the other.

COMPARING RLS AND LMS

Three examples for different choices of Q are listed in Table 21.1: (i) $Q = \sigma_q^2 \mathrm{I}$, i.e., the covariance matrix of the random perturbation vector $\boldsymbol{q}_i$ is a multiple of the identity; (ii) Q is a multiple of R_u and (iii) Q is a multiple of R_u^{-1}. It is seen from the results in the table that the performance of LMS is similar to that of RLS in case (i), while LMS is superior in case (ii), and RLS is superior in case (iii). The conclusions in cases (ii) and (iii) follow from the fact that for any $M \times M$ positive-definite matrix R_u, it always holds that (see Prob. IV.28):

$$\boxed{[\mathsf{Tr}(R_u)]^2 \leq M\,\mathsf{Tr}(R_u^2) \quad \text{and} \quad M^2 \leq \mathsf{Tr}(R_u)\mathsf{Tr}(R_u^{-1})} \qquad \text{(for any } R_u > 0\text{)}$$

Of course, these results on the tracking performance of RLS and LMS assume filter operation in environments with a small degree of nonstationarity.

COMPARING RLS AND LMS

TABLE 21.1 Minimum achievable excess-mean-square error of LMS and RLS (i.e., $\zeta_{\min}^{\text{LMS}}$ and $\zeta_{\min}^{\text{RLS}}$) for three choices of the nonstationary covariance matrix Q in comparison to the regressor covariance matrix R_u. In the table, α^2 is some constant value.

	$Q = \sigma_q^2 \mathrm{I}$	$Q = \alpha^2 R_u$	$Q = \alpha^2 R_u^{-1}$
$\dfrac{\zeta_{\min}^{\text{RLS}}}{\zeta_{\min}^{\text{LMS}}}$	1	$\dfrac{\sqrt{M\,\mathsf{Tr}(R_u^2)}}{\mathsf{Tr}(R_u)} \geq 1$	$\dfrac{M}{\sqrt{\mathsf{Tr}(R_u)\mathsf{Tr}(R_u^{-1})}} \leq 1$
	similar performance	LMS is superior	RLS is superior

OTHER FILTERS

TABLE 21.2 Approximate expressions for the excess mean-square performance of adaptive filters in nonstationary environments for small step-sizes.

Algorithm	EMSE	Reference
LMS	$\frac{\mu\sigma_v^2}{2}\mathsf{Tr}(R_u) + \frac{\mu^{-1}}{2}\mathsf{Tr}(Q)$	Lemma 21.1
ϵ-NLMS	$\frac{\mu\sigma_v^2}{2-\mu}\mathsf{Tr}(R_u)\mathsf{E}\left(\frac{1}{\lVert u_i\rVert^2}\right) + \frac{\mu^{-1}}{2-\mu}\mathsf{Tr}(R_u)\mathsf{Tr}(Q)$	Lemma 21.2
ϵ-NLMS with power normalization	$\frac{\mu(1+\beta)M\sigma_v^2 + \mu^{-1}\gamma\sigma_u^2(1-\beta)\mathsf{Tr}(Q)}{2\gamma(1-\beta)-\mu M(1+\beta)}$	Problem IV.24
sign-error LMS	$\frac{\alpha}{2}\left(\alpha+\sqrt{\alpha^2+4\sigma_v^2}\right)$ $\alpha = \sqrt{\frac{\pi}{8\gamma}}\left[\mu\gamma\mathsf{Tr}(R_u)+\mu^{-1}\mathsf{Tr}(Q)\right]$	Lemma 21.3
LMF	$\frac{\mu\xi_v^6}{2\sigma_v^2}\mathsf{Tr}(R_u) + \frac{\mu^{-1}}{2\sigma_v^2}\mathsf{Tr}(Q)$	Problem IV.25
LMMN	$\frac{\mu a}{2b}\mathsf{Tr}(R_u) + \frac{\mu^{-1}}{2b}\mathsf{Tr}(Q)$	Problem IV.25
leaky-LMS	$\frac{\mu\sigma_v^2}{2}\mathsf{Tr}[R_u^2(R_u+\alpha\mathrm{I})^{-1}] + \frac{\mu^{-1}}{2}\mathsf{Tr}[QR_u(R_u+\alpha\mathrm{I})^{-1}]$	Problem V.34
ϵ-APA	$\frac{\mu\sigma_v^2}{2-\mu}\mathsf{Tr}(R_u)\mathsf{E}\left(\frac{K}{\lVert u_i\rVert^2}\right) + \frac{\mu^{-1}}{2-\mu}\mathsf{Tr}(R_u)\mathsf{Tr}(Q)$	Problem IV.26
RLS	$\frac{\sigma_v^2(1-\lambda)M \ + \ \frac{1}{(1-\lambda)}\mathsf{Tr}(QR_u)}{2-(1-\lambda)M}$	Lemma 21.4
CMA2-2	$\frac{\mu\mathsf{E}\left(\gamma^2\lvert s\rvert^2-2\gamma\lvert s\rvert^4+\lvert s\rvert^6\right)\mathsf{Tr}(R_u)+\mu^{-1}\mathsf{Tr}(Q)}{2\mathsf{E}\left(2\lvert s\rvert^2-\gamma\right)}$	Problem IV.35
CMA1-2	$\frac{\mu}{2}\left(\gamma^2+\mathsf{E}\lvert s\rvert^2-2\gamma\mathsf{E}\lvert s\rvert\right)\mathsf{Tr}(R_u)+\frac{\mu^{-1}}{2}\mathsf{Tr}(Q)$	Problem IV.36

Project IV.1 (Line echo cancellation) In communications over phone lines, a signal travelling from a far-end point to a near-end point is usually reflected at the near-end due to mismatches in circuitry (e.g., hybrid connections). The reflected signal travels back to the far-end point in the form of an echo. As a result, the speaker at the far-end receives, in addition to the desired signal from the near-end speaker, an attenuated replica of his own signal in the form of an echo — see Fig. IV.3.

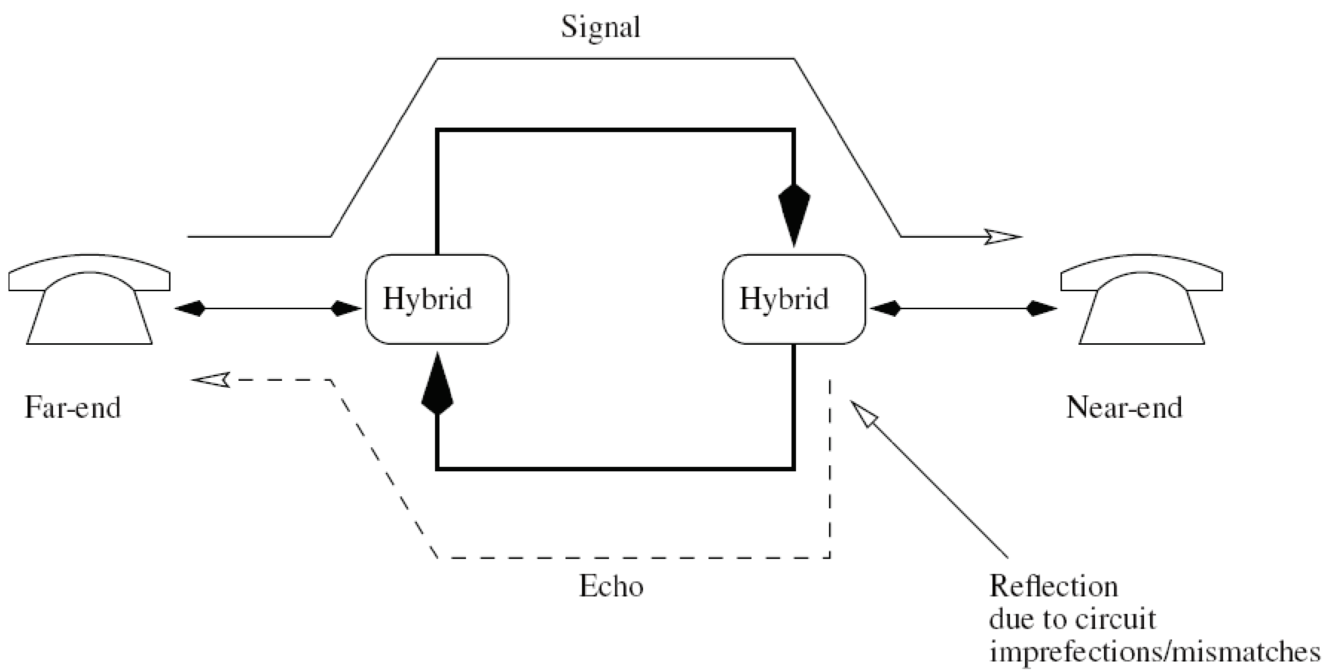

FIGURE IV.3 The signal at the far-end is reflected at the near-end due to circuit mismatches and travels back to the far-end.

LINE ECHO CANCELLATION

The echo interferes with the quality of the received signal. A common way to provide better voice quality at both ends is to employ adaptive line echo cancellers (LEC). At the near-end, for example, the signal feeding the LEC is the far-end signal while the reference signal is its reflected version — see Fig. IV.4. In the figure, the output of the adaptive LEC generates a replica of the echo, and the error signal is therefore a "clean" signal that is transmitted to the far-end. The signals in this project are assumed to be sampled at 8 kHz.

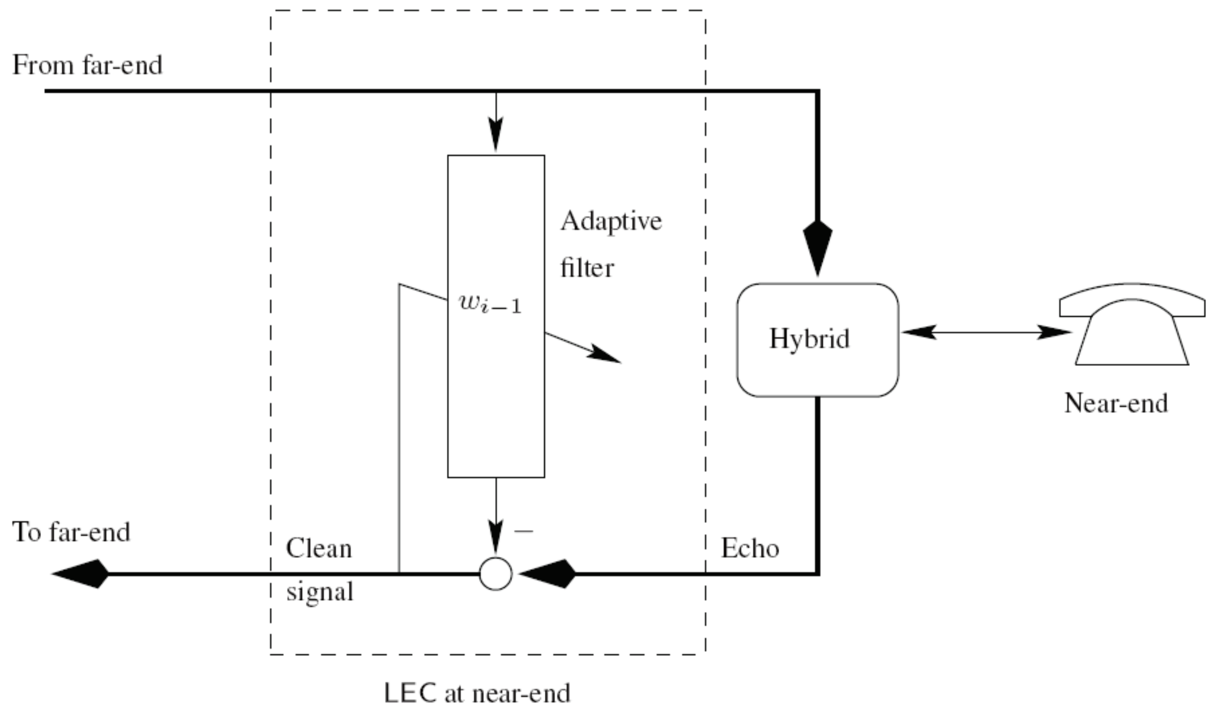

FIGURE IV.4 An adaptive line echo canceller at the near-end.

LINE ECHO CANCELLATION

(a) Load the file path.mat, which contains the impulse response sequence of a typical echo path. Plot the impulse and frequency responses of the echo path.

(b) Load the file css.mat, which contains 5600 samples of a composite source signal; it is a synthetic signal that emulates the properties of speech. Specifically, it contains segments of pause, segments of periodic excitation and segments with white-noise properties. Plot the samples of the CSS data, as well as their spectrum.

(c) Concatenate five such blocks and feed them into the echo path. Plot the resulting echo signal. Estimate the input and output powers in dB using

$$\widehat{P} = 10\log_{10}\left(\frac{1}{N}\sum_{i=1}^{N}|\text{signal}(i)|^2\right)$$

where N denotes the length of the sequence. Evaluate the attenuation in dB that is introduced by the echo path as the signal travels through it; this attenuation is called the echo-return-loss (ERL).

(d) Use 10 blocks of CSS data as far-end signal, and the corresponding output of the echo path as the echo signal. Choose an adaptive line echo canceller with 128 taps. Train the canceller by using as input data the far-end signal, i.e., $u(i) = \text{far\_end}(i)$, and as reference data the echo signal, i.e., $d(i) = \text{echo}(i)$. Use ϵ-NLMS with $\epsilon = 10^{-6}$ and $\mu = 0.25$. Plot the far-end signal, the echo, and the error signal provided by the adaptive filter. Plot also the echo path and its estimate by the adaptive filter at the end of the simulation.

(e) Estimate the steady-state power of the error signal and measure its attenuation in dB relative to the echo signal. Use the last 5600 of the signals to estimate their powers. The difference in power is a measure of the attenuation introduced by the LEC and it is called the echo-return-loss-enhancement (ERLE).

(f) Fix the input power at 0 dB and add white Gaussian noise with variance $\sigma_v^2 = 0.0001$ to the echo signal. Train the LEC using 80 blocks of CSS data and measure the steady-state ERLE. Compare the simulated and theoretical ERLEs.

Project IV.1 (Line echo cancellation) The programs that solve this project are the following.

1. <u>partA.m</u> This program generates two plots showing the impulse response sequence of the channel as well as its frequency response, as shown in Fig. 1. It is seen that the bandwidth of the echo path model is approximately 3.4 kHz.

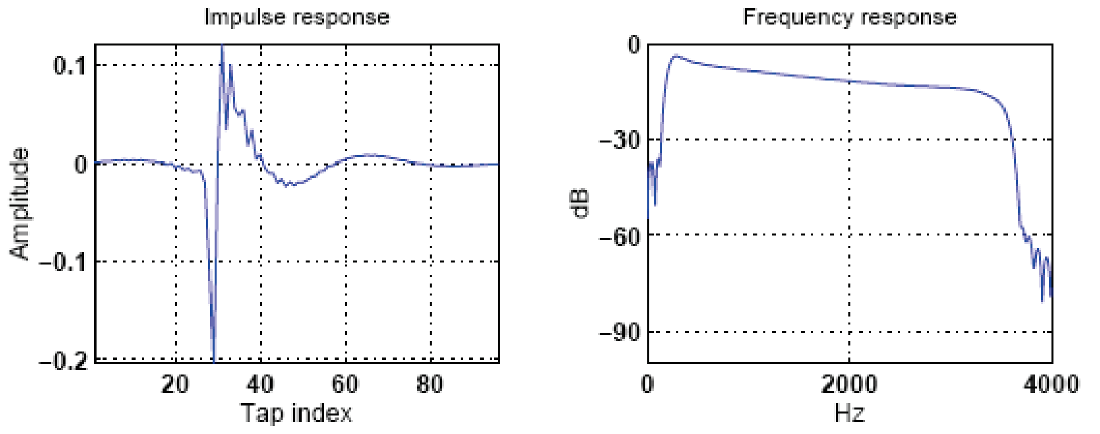

Figure IV.1. Impulse and frequency responses of the echo path model.

2. partB.m This program solves parts B and C — see Fig.2. The powers of the input and output signals are 0 and −6.3 dB, respectively, so that ERL ≈ 6.3 dB and the amplitude of a signal going through the channel is reduced by 1/2.

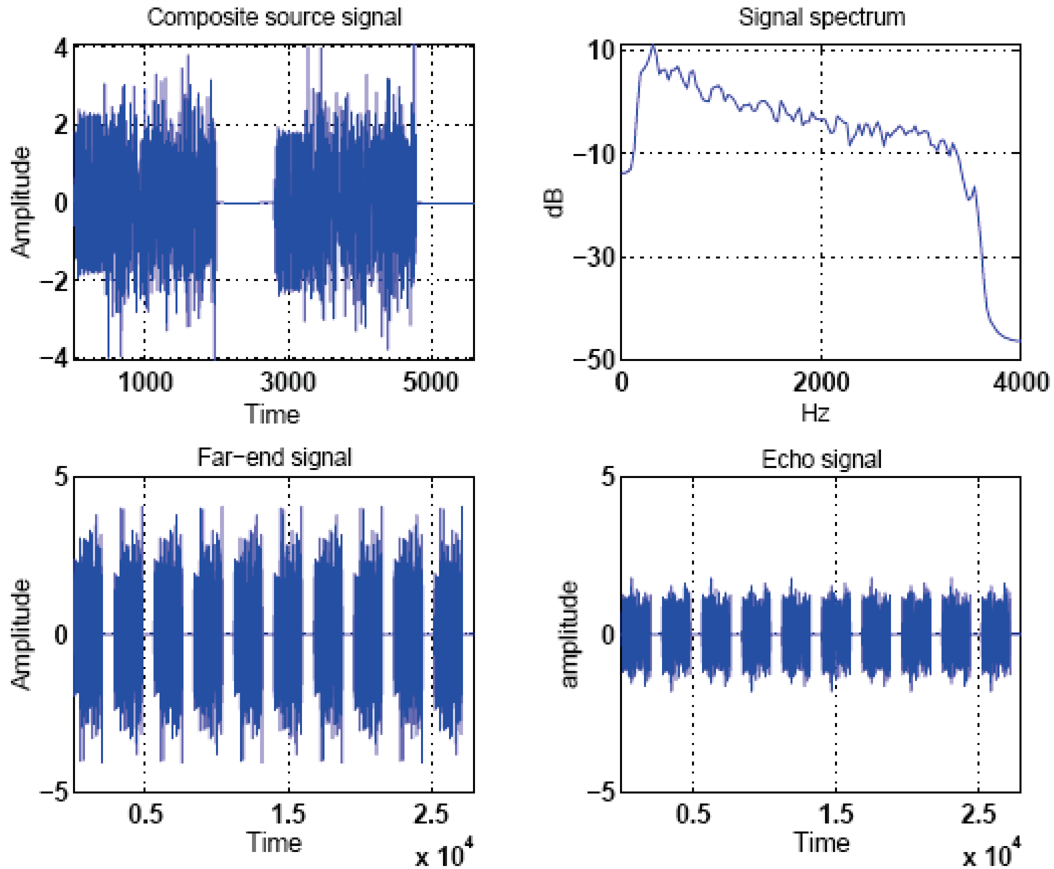

Figure IV.2. The top row shows the CSS data and their spectrum, while the bottom row shows 5 blocks of CSS data and the resulting echo.

3. partD.m This program solves parts D and E and generates two plots. Figure 3 shows the echo path impulse response and its estimate, while Fig. 4 shows the far-end signal, the echo, and the error signal. The resulting ERLE was found to be around 56 dB. Consequently, the total attenuation from the far-end input to the output of the adaptive filter is seen to be $\text{ERL} + \text{ERLE} \approx 62$ dB.

4. partF.m The power of the echo signal is the MSE of the adaptive filter, so that

$$P_{\text{error}}(\text{dB}) \approx 10\log_{10}\left(\sigma_v^2 + \frac{\mu\sigma_v^2}{2-\mu}\right) = -39.42 \text{ dB}$$

Since $P_{\text{echo}} \approx -6.3$ dB, we find that the theoretical ERLE is ≈ 33.1 dB. The simulated value in this experiment is ≈ 29.5 dB. The simulated value will get closer to theory if the simulation is run for a longer period of time.

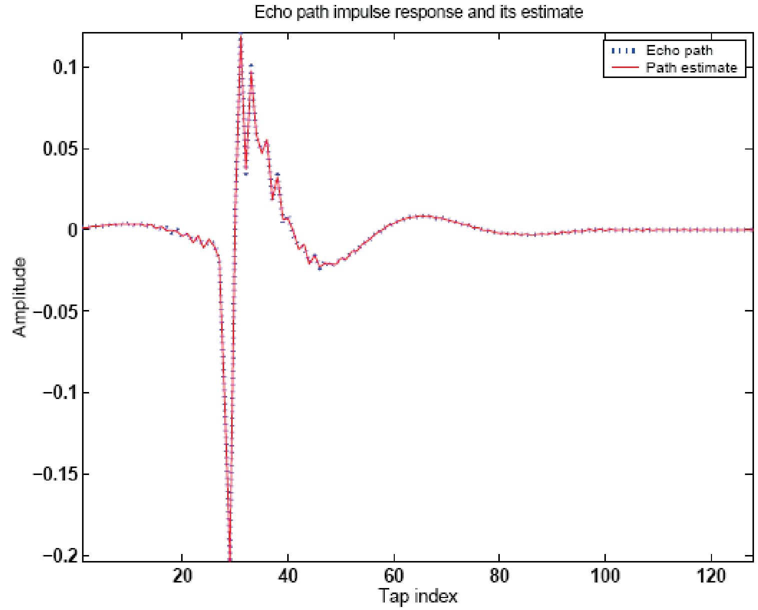

Figure IV.3. The figure shows the impulse response sequences of the echo path and its estimate by the adaptive filter.

LINE ECHO CANCELLATION

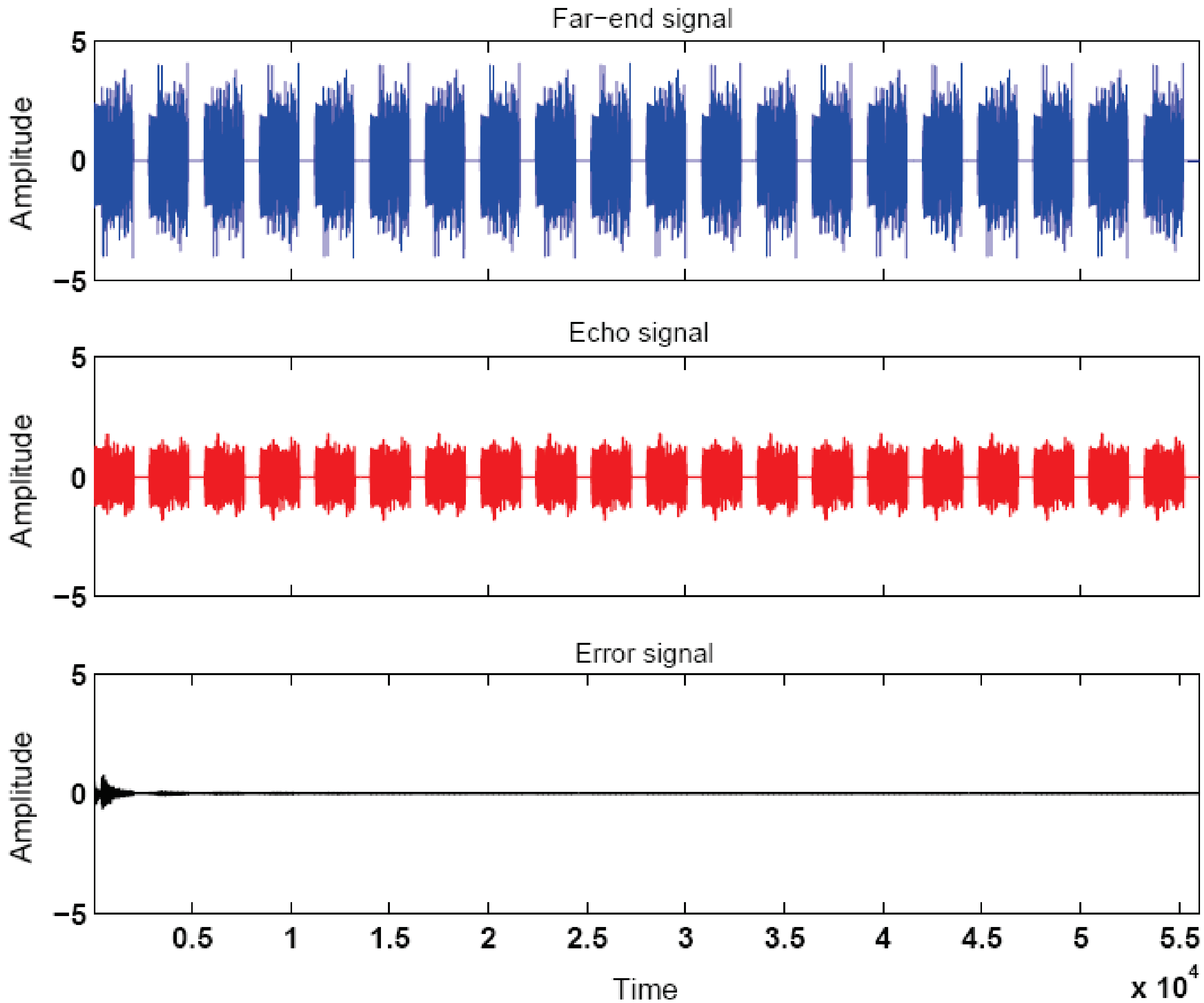

Figure IV.4. The top row shows the far-end signal, while the middle row shows the corresponding echo and the last row shows the resulting error signal.