EE210A: Adaptation and Learning

Professor Ali H. Sayed

TRANSIENT PERFORMANCE

Sections in order: 22.1-22.4, 23.1

MOTIVATION

As is evident by now, adaptive filters are time-variant and nonlinear stochastic systems with inherent learning and tracking abilities. The success of their learning mechanism can be measured in terms of how well they learn the underlying signal statistics given sufficient time (i.e., in terms of their steady-state performance) and in terms of how fast and how stably they adapt to changes in the signal statistics (i.e., in terms of their transient and convergence performance). For this reason, it is customary to study the performance of adaptive filters by examining their transient performance and their steady-state performance. The former is concerned with the stability and convergence rate of an adaptive scheme, while the latter is concerned with the mean-square error that remains in steady-state.

22.1 DATA MODEL

We rely on the same data model that we adopted in Sec. 15.2 for stationary environments. Specifically, we assume that the data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfy the following conditions:

$$\begin{array}{l}
\text{(a) There exists a vector } w^o \text{ such that } \boldsymbol{d}(i) = \boldsymbol{u}_i w^o + \boldsymbol{v}(i). \\
\text{(b) The noise sequence } \{\boldsymbol{v}(i)\} \text{ is i.i.d. with variance } \sigma_v^2 = \mathsf{E}\,|\boldsymbol{v}(i)|^2. \\
\text{(c) The noise sequence } \{\boldsymbol{v}(i)\} \text{ is independent of } \boldsymbol{u}_j \text{ for all } i, j. \\
\text{(d) The initial condition } \boldsymbol{w}_{-1} \text{ is independent of all } \{\boldsymbol{d}(j), \boldsymbol{u}_j, \boldsymbol{v}(j)\}. \\
\text{(e) The regressor covariance matrix is denoted by } R_u = \mathsf{E}\,\boldsymbol{u}_i^* \boldsymbol{u}_i > 0. \\
\text{(f) The random variables } \{\boldsymbol{v}(i), \boldsymbol{u}_i\} \text{ have zero means.}
\end{array} \tag{22.1}$$

22.2 DATA-NORMALIZED ADAPTIVE FILTERS

We shall study filter updates of the form

$$\boxed{\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \frac{\boldsymbol{u}_i^*}{g[\boldsymbol{u}_i]} \boldsymbol{e}(i), \quad i \geq 0} \tag{22.2}$$

where $\boldsymbol{w}_i$ is an estimate for w^o at iteration i, μ is the step-size,

$$\boxed{\boldsymbol{e}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1}} \tag{22.3}$$

is the estimation error, and $g[\cdot]$ is some positive-valued function of $\boldsymbol{u}_i$. For example, the choice $g[\boldsymbol{u}_i] = 1$ results in the LMS algorithm, while $g[\boldsymbol{u}_i] = \epsilon + \|\boldsymbol{u}_i\|^2$ results in the $\epsilon-$ NLMS algorithm. One could also study more general data-normalized updates of the

22.3 WEIGHTED ENERGY CONSERVATION RELATION

Since we shall deal with weighted vector norms on a regular basis in this chapter, we adopt the compact notation $\|x\|^2_\Sigma$ to refer to the weighted squared Euclidean norm of a vector x, i.e.,

$$\boxed{\|x\|^2_\Sigma \;\stackrel{\Delta}{=}\; x^*\Sigma x}$$

for some Hermitian positive-definite weighting matrix Σ. The choice $\Sigma = \mathrm{I}$ results in the standard squared Euclidean norm of x, i.e., $\|x\|^2_{\mathrm{I}} = x^*x = \|x\|^2$. Although in general we shall encounter situations with $\Sigma > 0$, we shall use the same notation $\|x\|^2_\Sigma$ to denote $x^*\Sigma x$ even when Σ is non-negative definite – see Prob. V.1.

WEIGHTED NORMS

The need to consider weighted norms in the context of a transient analysis of adaptive filters can be motivated as follows. It will be seen that the transient performance of an adaptive filter requires that we study the time evolution of expectations of the form $\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2$ and $\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$, where the first expectation relates to the weight-error vector, $\tilde{\boldsymbol{w}}_i = w^o - \boldsymbol{w}_i$, while the second expectation relates to the *a priori* estimation error, $\boldsymbol{e}_a(i) = \boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}$. The evaluation of the expectation $\mathsf{E}\,|\boldsymbol{e}_a(i)|^2$ will in turn require that we evaluate a weighted norm of $\tilde{\boldsymbol{w}}_{i-1}$ of the form

$$\mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{R_u}$$

with the particular weighting matrix $\Sigma = R_u$. Now the energy conservation relation that we encountered in Thm. 15.1 involves the squared Euclidean norm of $\tilde{\boldsymbol{w}}_{i-1}$, $\|\tilde{\boldsymbol{w}}_{i-1}\|^2$, and not any weighted version of it. For this reason, we shall first extend our arguments to allow for weighted vector norms. As we shall see, a weighted version of the energy relation can be obtained rather immediately by following arguments similar to those that led to Thm. 15.1.

WEIGHTED ERRORS

Thus let Σ denote any $M \times M$ Hermitian nonnegative-definite matrix (in general, we shall have $\Sigma > 0$). Later we shall see that different choices for Σ are useful to infer different conclusions about the performance of an adaptive filter. Now define the weighted *a priori* and *a posteriori* error signals

$$\boxed{e_a^{\Sigma}(i) \;\triangleq\; \boldsymbol{u}_i \Sigma \tilde{\boldsymbol{w}}_{i-1}, \qquad e_p^{\Sigma}(i) \;\triangleq\; \boldsymbol{u}_i \Sigma \tilde{\boldsymbol{w}}_i} \tag{22.5}$$

where

$$\boxed{\tilde{\boldsymbol{w}}_i \;\triangleq\; w^o - \boldsymbol{w}_i} \tag{22.6}$$

When $\Sigma = \mathrm{I}$, we recover the standard errors

$$\boxed{\boldsymbol{e}_a(i) \;\triangleq\; \boldsymbol{e}_a^{I}(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1}, \qquad \boldsymbol{e}_p(i) \;\triangleq\; \boldsymbol{e}_p^{I}(i) = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_i} \tag{22.7}$$

WEIGHT ERROR RECURSION

The energy relation that we seek is one that compares the weighted energies of the error quantities:

$$\{\ \tilde{\boldsymbol{w}}_i, \tilde{\boldsymbol{w}}_{i-1}, \boldsymbol{e}_a^{\Sigma}(i), \boldsymbol{e}_p^{\Sigma}(i)\ \} \tag{22.8}$$

To arrive at the desired relation, we follow the same arguments that we employed before in Sec. 15.3.

First, we rewrite the update recursion (22.2)–(22.3) in terms of the weight-error vector $\tilde{\boldsymbol{w}}_i$. Subtracting both sides of (22.2) from w^o we get

$$\boxed{\tilde{\boldsymbol{w}}_i = \tilde{\boldsymbol{w}}_{i-1} - \mu \frac{\boldsymbol{u}_i^*}{g[\boldsymbol{u}_i]} \boldsymbol{e}(i)} \tag{22.9}$$

If we further multiply both sides of (22.9) by $\boldsymbol{u}_i\Sigma$ from the left we find that the *a priori* and *a posteriori* estimation errors $\{\boldsymbol{e}_p^{\Sigma}(i), \boldsymbol{e}_a^{\Sigma}(i)\}$ are related via

$$\boxed{\boldsymbol{e}_p^{\Sigma}(i) = \boldsymbol{e}_a^{\Sigma}(i) - \mu \frac{\|\boldsymbol{u}_i\|_{\Sigma}^2}{g[\boldsymbol{u}_i]} \boldsymbol{e}(i)} \tag{22.10}$$

ENERGY CONSERVATION

We distinguish between two cases:

1. $\|\boldsymbol{u}_i\|^2_\Sigma \neq 0$. In this case, we use (22.10) to solve for $\boldsymbol{e}(i)/g[\boldsymbol{u}_i]$,

$$\frac{\boldsymbol{e}(i)}{g[\boldsymbol{u}_i]} = \frac{1}{\mu\|\boldsymbol{u}_i\|^2_\Sigma}\left[\boldsymbol{e}^\Sigma_a(i) - \boldsymbol{e}^\Sigma_p(i)\right]$$

and substitute into (22.9) to get

$$\tilde{\boldsymbol{w}}_i + \frac{\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_\Sigma}\boldsymbol{e}^\Sigma_a(i) = \tilde{\boldsymbol{w}}_{i-1} + \frac{\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_\Sigma}\boldsymbol{e}^\Sigma_p(i) \qquad (22.11)$$

On each side of this identity we have a combination of *a priori* and *a posteriori* errors. By equating the weighted Euclidean norms of both sides of the equation, i.e., by setting

$$\left\|\tilde{\boldsymbol{w}}_i + \frac{\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_\Sigma}\boldsymbol{e}^\Sigma_a(i)\right\|^2_\Sigma = \left\|\tilde{\boldsymbol{w}}_{i-1} + \frac{\boldsymbol{u}_i^*}{\|\boldsymbol{u}_i\|^2_\Sigma}\boldsymbol{e}^\Sigma_p(i)\right\|^2_\Sigma$$

we find, after a straightforward calculation, that the following energy equality holds:

$$\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma + \frac{1}{\|\boldsymbol{u}_i\|^2_\Sigma}|\boldsymbol{e}^\Sigma_a(i)|^2 = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_\Sigma + \frac{1}{\|\boldsymbol{u}_i\|^2_\Sigma}|\boldsymbol{e}^\Sigma_p(i)|^2$$

ENERGY CONSERVATION

Observe that this equality simply amounts to adding the weighted energies of the individual terms of (22.11); the cross-terms cancel out. Equivalently, we can rewrite the above equality as

$$\boxed{\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma \;+\; \bar{\mu}^\Sigma(i)\cdot|\boldsymbol{e}^\Sigma_a(i)|^2 \;=\; \|\tilde{\boldsymbol{w}}_{i-1}\|^2_\Sigma \;+\; \bar{\mu}^\Sigma(i)\cdot|\boldsymbol{e}^\Sigma_p(i)|^2} \tag{22.12}$$

where

$$\bar{\mu}^\Sigma(i) \;\triangleq\; \begin{cases} 1/\|\boldsymbol{u}_i\|^2_\Sigma & \text{if } \|\boldsymbol{u}_i\|^2_\Sigma \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{22.13}$$

2. $\underline{\|\boldsymbol{u}_i\|^2_\Sigma = 0}$ (since $\Sigma \geq 0$, this implies that $\boldsymbol{u}_i\Sigma = 0$). In this case, it is obvious from (22.10) that $\boldsymbol{e}^\Sigma_a(i) = \boldsymbol{e}^\Sigma_p(i)$ and from (22.9) that $\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_\Sigma$, so that (22.12) is valid again.

ENERGY CONSERVATION

Theorem 22.1 (Weighted energy-conservation relation) For any adaptive filter of the form (22.2), any Hermitian nonnegative-definite matrix Σ, and for any data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$, it holds that

$$\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma \; + \; \bar{\mu}^\Sigma(i) \cdot |\boldsymbol{e}^\Sigma_a(i)|^2 \; = \; \|\tilde{\boldsymbol{w}}_{i-1}\|^2_\Sigma \; + \; \bar{\mu}^\Sigma(i) \cdot |\boldsymbol{e}^\Sigma_p(i)|^2$$

where $\boldsymbol{e}^\Sigma_a(i) = \boldsymbol{u}_i \Sigma \tilde{\boldsymbol{w}}_{i-1}$, $\boldsymbol{e}^\Sigma_p(i) = \boldsymbol{u}_i \Sigma \tilde{\boldsymbol{w}}_i$, $\tilde{\boldsymbol{w}}_i = w^o - \boldsymbol{w}_i$, and $\bar{\mu}^\Sigma(i)$ is defined by (22.13).

The important fact to emphasize here is that *no approximations* have been used to establish the energy relation (22.12); it is an exact relation that shows how the energies of the weight-error vectors at two successive time instants are related to the energies of the *a priori* and *a posteriori* estimation errors. The special choice $\Sigma = \mathrm{I}$ reduces to the energy relation of Thm. 15.1. In addition, the same geometric, physical, and system-theoretic interpretations that were presented in App. 15.A for the case $\Sigma = \mathrm{I}$ can be extended to the weighted case with little effort — see App. 9.G of Sayed (2003).

WEIGHT ERROR RECURSION

We also rewrite below, for later reference, the weight-error recursion (22.9). From the modeling assumption (22.1) we have

$$\boldsymbol{e}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1} = \boldsymbol{u}_i \tilde{\boldsymbol{w}}_{i-1} \ + \ \boldsymbol{v}(i)$$

so that substituting into (22.9) we get the equivalent form

$$\tilde{\boldsymbol{w}}_i = \left(\mathrm{I} - \mu \frac{\boldsymbol{u}_i^* \boldsymbol{u}_i}{g[\boldsymbol{u}_i]} \right) \tilde{\boldsymbol{w}}_{i-1} \ - \ \mu \frac{\boldsymbol{u}_i^*}{g[\boldsymbol{u}_i]} \boldsymbol{v}(i) \tag{22.14}$$

Theorem 22.2 (Weight-error recursion) For any adaptive filter of the form (22.2), and for data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfying (22.1), it holds that

$$\tilde{\boldsymbol{w}}_i = \left(\mathrm{I} - \mu \frac{\boldsymbol{u}_i^* \boldsymbol{u}_i}{g[\boldsymbol{u}_i]} \right) \tilde{\boldsymbol{w}}_{i-1} \ - \ \mu \frac{\boldsymbol{u}_i^*}{g[\boldsymbol{u}_i]} \boldsymbol{v}(i)$$

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

22.4 WEIGHTED VARIANCE RELATION

Relation (22.12) has several useful ramifications in the study of adaptive filters, as was already discussed at some length in Part IV (*Mean-Square Performance*). In this part we focus on its significance to transient analysis. Thus recall that in Sec. 15.2 we used the energy-conservation relation (15.32) to study the steady-state performance of adaptive filters. In that section, we invoked the steady-state condition

$$\mathsf{E}\,\|\tilde{w}_i\|^2 = \mathsf{E}\,\|\tilde{w}_{i-1}\|^2 \qquad \text{as} \quad i \to \infty$$

in order to cancel the effect of the weight-error vector from both sides of the energy relation, and then used the resulting variance relation (15.40) to evaluate the filter EMSE, i.e., the value of $\mathsf{E}\,|e_a(\infty)|^2$.

VARIANCE RELATION

In transient analysis, on the other hand, we will be interested in the time evolution of $\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma$ for some choices of interest for Σ (usually, $\Sigma = \mathrm{I}$ or $\Sigma = R_u$). For this reason, in transient analysis, rather than eliminate the effect of the weight-error vector from (22.12), the contributions of the other error quantities, $\{\boldsymbol{e}_a^\Sigma(i), \boldsymbol{e}_p^\Sigma(i)\}$, will instead be expressed in terms of the weight-error vector itself. In so doing, the energy relation (22.12) will be transformed into a recursion that describes the evolution of $\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma$ — see, e.g., Eq. (22.21) further ahead.

VARIANCE RELATION

Thus returning to (22.12), and replacing $e_p^{\Sigma}(i)$ by its equivalent expression (22.10) in terms of $e_a^{\Sigma}(i)$ and $e(i)$ we get

$$\|\tilde{w}_i\|_{\Sigma}^2 + \bar{\mu}^{\Sigma}(i) \cdot \left|e_a^{\Sigma}(i)\right|^2 = \|\tilde{w}_{i-1}\|_{\Sigma}^2 + \bar{\mu}^{\Sigma}(i) \cdot \left|e_a^{\Sigma}(i) - \frac{\mu \cdot \|u_i\|_{\Sigma}^2}{g[u_i]}\, e(i)\right|^2$$

or, equivalently, after expanding the rightmost term,

$$\begin{aligned} \|\tilde{w}_i\|_{\Sigma}^2 &= \|\tilde{w}_{i-1}\|_{\Sigma}^2 + \frac{\mu^2 \|u_i\|_{\Sigma}^2}{g^2[u_i]} |e(i)|^2 \\ &\quad - \frac{\mu}{g[u_i]} e_a^{\Sigma *}(i) e(i) - \frac{\mu}{g[u_i]} e^*(i) e_a^{\Sigma}(i) \end{aligned} \qquad (22.15)$$

Alternatively, this result can be obtained by starting directly from the weight-error vector recursion (22.9) and equating the weighted Euclidean norms of both sides.

VARIANCE RELATION

Now using the data model (22.1) for $\boldsymbol{d}(i)$, it is clear that $\boldsymbol{e}(i)$ is related to $\boldsymbol{e}_a(i)$ and $\boldsymbol{v}(i)$ as follows:

$$\boldsymbol{e}(i) = \boldsymbol{d}(i) - \boldsymbol{u}_i \boldsymbol{w}_{i-1} = (\boldsymbol{u}_i w^o + \boldsymbol{v}(i)) - \boldsymbol{u}_i \boldsymbol{w}_{i-1} \; = \; \boldsymbol{e}_a(i) + \boldsymbol{v}(i) \qquad (22.16)$$

so that substituting into (22.15) we can eliminate $\boldsymbol{e}(i)$ and get

$$\begin{aligned}
\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2 \;\; = \;\; & \|\tilde{\boldsymbol{w}}_{i-1}\|_\Sigma^2 \\
& + \frac{\mu^2 \|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]} \, |\boldsymbol{e}_a(i)|^2 \; + \; \frac{\mu^2 \|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]} \, |\boldsymbol{v}(i)|^2 \\
& + \frac{\mu^2 \|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]} \, \boldsymbol{v}(i) \boldsymbol{e}_a^*(i) \; + \; \frac{\mu^2 \|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]} \, \boldsymbol{v}^*(i) \boldsymbol{e}_a(i) \\
& - \frac{\mu}{g[\boldsymbol{u}_i]} \boldsymbol{e}_a^{\Sigma *}(i) \boldsymbol{e}_a(i) \; - \; \frac{\mu}{g[\boldsymbol{u}_i]} \boldsymbol{v}(i) \boldsymbol{e}_a^{\Sigma *}(i) \\
& - \frac{\mu}{g[\boldsymbol{u}_i]} \boldsymbol{e}_a^*(i) \boldsymbol{e}_a^{\Sigma}(i) \; - \; \frac{\mu}{g[\boldsymbol{u}_i]} \boldsymbol{v}^*(i) \boldsymbol{e}_a^{\Sigma}(i)
\end{aligned} \qquad (22.17)$$

VARIANCE RELATION

Most of the factors in this equality disappear under expectation, while other factors can be expressed in terms of $\tilde{\boldsymbol{w}}_{i-1}$. To see this, observe first that

$$\boldsymbol{e}_a^*(i)\boldsymbol{e}_a^{\Sigma}(i) = \tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\boldsymbol{u}_i\Sigma\tilde{\boldsymbol{w}}_{i-1}$$

and

$$\boldsymbol{e}_a^{\Sigma*}(i)\boldsymbol{e}_a(i) = \tilde{\boldsymbol{w}}_{i-1}^*\Sigma\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1}$$

so that the first terms on the fourth and fifth lines of (22.17) can be grouped together as a single weighted norm of $\tilde{\boldsymbol{w}}_{i-1}$ as follows:

$$\frac{\mu}{g[\boldsymbol{u}_i]}\boldsymbol{e}_a^{\Sigma*}(i)\boldsymbol{e}_a(i) \;+\; \frac{\mu}{g[\boldsymbol{u}_i]}\boldsymbol{e}_a^*(i)\boldsymbol{e}_a^{\Sigma}(i) \;=\; \frac{\mu}{g[\boldsymbol{u}_i]}\cdot\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\Sigma\boldsymbol{u}_i^*\boldsymbol{u}_i+\boldsymbol{u}_i^*\boldsymbol{u}_i\Sigma}$$

Likewise,

$$|\boldsymbol{e}_a(i)|^2 = \tilde{\boldsymbol{w}}_{i-1}^*\boldsymbol{u}_i^*\boldsymbol{u}_i\tilde{\boldsymbol{w}}_{i-1} = \|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{u}_i^*\boldsymbol{u}_i}$$

VARIANCE RELATION

so that (22.17) becomes

$$\begin{aligned}
\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2 \;=\;& \|\tilde{\boldsymbol{w}}_{i-1}\|_\Sigma^2 \\
&+ \frac{\mu^2\|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]}\;\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{u}_i^*\boldsymbol{u}_i} \\
&+ \frac{\mu^2\|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]}\;|\boldsymbol{v}(i)|^2 \\
&- \frac{\mu}{g[\boldsymbol{u}_i]}\;\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\Sigma\boldsymbol{u}_i^*\boldsymbol{u}_i+\boldsymbol{u}_i^*\boldsymbol{u}_i\Sigma} \\
&+ \frac{\mu^2\|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]}\;\boldsymbol{v}(i)\boldsymbol{e}_a^*(i) \;+\; \frac{\mu^2\|\boldsymbol{u}_i\|_\Sigma^2}{g^2[\boldsymbol{u}_i]}\;\boldsymbol{v}^*(i)\boldsymbol{e}_a(i) \\
&- \frac{\mu}{g[\boldsymbol{u}_i]}\boldsymbol{v}(i)\boldsymbol{e}_a^{\Sigma*}(i) \;-\; \frac{\mu}{g[\boldsymbol{u}_i]}\boldsymbol{v}^*(i)\boldsymbol{e}_a^{\Sigma}(i)
\end{aligned} \tag{22.18}$$

VARIANCE RELATION

Taking expectations of both sides of (22.18) we find

$$\begin{aligned} \mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma \quad = \quad & \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2_\Sigma \;+\; \mathsf{E}\left(\frac{\mu^2\|\boldsymbol{u}_i\|^2_\Sigma}{g^2[\boldsymbol{u}_i]}\;\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{u}_i^*\boldsymbol{u}_i}\right) \\ & -\;\mathsf{E}\left(\frac{\mu}{g[\boldsymbol{u}_i]}\;\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\Sigma\boldsymbol{u}_i^*\boldsymbol{u}_i+\boldsymbol{u}_i^*\boldsymbol{u}_i\Sigma}\right) \;+\; \sigma_v^2\mathsf{E}\left(\frac{\mu^2\|\boldsymbol{u}_i\|^2_\Sigma}{g^2[\boldsymbol{u}_i]}\right) \end{aligned} \tag{22.19}$$

where the expectations of the cross-terms involving $\boldsymbol{v}(i)$ evaluate to zero due to the modeling assumption on $\boldsymbol{v}(i)$ from (22.1), namely, that $\boldsymbol{v}(i)$ is zero-mean, i.i.d., and independent of $\boldsymbol{u}_j$ for all j. Alternatively, we can obtain (22.19) by starting from the weight-error recursion (22.9), equating the weighted norms of both sides, and taking expectations to arrive at (22.19) — see, e.g., the result of Prob. V.27 specialized to $\alpha = 0$.

VARIANCE RELATION

We can further include some of the multiplicative factors in (22.19) into the weighting matrices and write

$$\begin{aligned}\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} \;&=\; \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\Sigma} \;+\; \mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\frac{\mu^2\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\boldsymbol{u}_i^*\boldsymbol{u}_i}\right) \\ &\quad -\; \mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\frac{\mu}{g[\boldsymbol{u}_i]}(\Sigma\boldsymbol{u}_i^*\boldsymbol{u}_i+\boldsymbol{u}_i^*\boldsymbol{u}_i\Sigma)}\right) \;+\; \sigma_v^2\mathsf{E}\left(\frac{\mu^2\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\right)\end{aligned}$$

This equality can be written more compactly as follows. Introduce the random weighting matrix

$$\boldsymbol{\Sigma}' \;\triangleq\; \Sigma \;-\; \frac{\mu}{g[\boldsymbol{u}_i]}\Sigma\boldsymbol{u}_i^*\boldsymbol{u}_i \;-\; \frac{\mu}{g[\boldsymbol{u}_i]}\boldsymbol{u}_i^*\boldsymbol{u}_i\Sigma \;+\; \frac{\mu^2\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\boldsymbol{u}_i^*\boldsymbol{u}_i \qquad (22.20)$$

In accordance with our convention, we are using a boldface symbol $\boldsymbol{\Sigma}'$ since $\boldsymbol{\Sigma}'$ is a random quantity (owing to its dependence on the regressor $\boldsymbol{u}_i$). Then

$$\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} = \mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{\Sigma}'}\right) \;+\; \mu^2\sigma_v^2\mathsf{E}\left(\frac{\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\right) \qquad (22.21)$$

VARIANCE RELATION

In other words, starting from the energy conservation relation (22.12), expanding it, and expressing whatever factors possible in terms of $\tilde{w}_{i-1}$, and then taking expectations to eliminate cross-terms involving the noise variable $v(i)$, we arrive at the variance relation (22.21). This relation shows how the weighted mean-square norm of $\tilde{w}_i$ propagates in time. In particular, observe the important fact that the weighting matrices at the time instants i and $i-1$ are distinct and related via (22.20). Moreover, it is shown in Prob. V.2 that $\Sigma' \geq 0$ when $\Sigma > 0$.

Theorem 22.3 (Weighted variance relation) For adaptive filters of the form (22.2), Hermitian nonnegative-definite matrices Σ, and for data $\{d(i), u_i\}$ satisfying model (22.1), it holds that

$$\begin{aligned}
\mathsf{E}\,\|\tilde{w}_i\|^2_{\Sigma} &= \mathsf{E}\left(\|\tilde{w}_{i-1}\|^2_{\Sigma'}\right) + \mu^2\sigma_v^2\,\mathsf{E}\left(\frac{\|u_i\|^2_{\Sigma}}{g^2[u_i]}\right) \\
\Sigma' &= \Sigma - \mu\Sigma\frac{u_i^* u_i}{g[u_i]} - \mu\frac{u_i^* u_i}{g[u_i]}\Sigma + \mu^2\frac{\|u_i\|^2_{\Sigma}}{g^2[u_i]}u_i^* u_i
\end{aligned}$$

INDEPENDENCE ASSUMPTION

The recursion of Thm. 22.3 provides a compact characterization of the time-evolution (i.e., dynamics) of the expectation $\mathsf{E}\,\|\tilde{w}_i\|^2_{\Sigma}$. However, more is needed in order for this recursion to permit a tractable transient analysis. This is because the recursion for $\mathsf{E}\,\|\tilde{w}_i\|^2_{\Sigma}$ is hard to propagate as it stands due to the presence of the expectation

$$\mathsf{E}\left(\|\tilde{w}_{i-1}\|^2_{\Sigma'}\right) \quad = \quad \mathsf{E}\left(\tilde{w}^*_{i-1}\Sigma'\tilde{w}_{i-1}\right) \tag{22.22}$$

This expectation is difficult to evaluate because of the dependence of Σ' on u_i, and the dependence of $\tilde{w}_{i-1}$ on prior regressors (so that $\tilde{w}_{i-1}$ and Σ' are themselves dependent). These dependencies are among the most challenging hurdles in the transient analysis of adaptive filters. One common way to overcome them is to resort to an independence assumption on the regression sequence $\{u_i\}$, namely to assume that

$$\boxed{\text{The sequence } \{u_i\} \text{ is independent and identically distributed}} \tag{22.23}$$

Actually, it is customary in the literature to start the transient analysis of an adaptive filter with the collection of independence assumptions that were described in Sec. 16.4, which included, in addition to the independence condition (22.23), a Gaussian requirement on u_i as well. Our arguments will not require the regressors to be Gaussian.

INDEPENDENCE ASSUMPTIONS

Now note that condition (22.23) guarantees that

$$\boxed{\tilde{\boldsymbol{w}}_{i-1} \text{ is independent of both } \boldsymbol{\Sigma}' \text{ and } \boldsymbol{u}_i} \tag{22.24}$$

This is because $\tilde{\boldsymbol{w}}_{i-1}$ is a function of past regressors and noise, $\{\boldsymbol{u}_j, \boldsymbol{v}(j), j < i\}$ (cf. the explanation after (16.12)), while $\boldsymbol{\Sigma}'$ is a function of $\boldsymbol{u}_i$ alone. Using (22.24) we can then split the expectation $\mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{\Sigma}'}\right)$ into

$$\mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{\Sigma}'}\right) = \mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\mathsf{E}\,[\boldsymbol{\Sigma}']}\right) \tag{22.25}$$

with the weighting matrix $\boldsymbol{\Sigma}'$ replaced by its mean, and which we denote by Σ', i.e.,

$$\boxed{\Sigma' \triangleq \mathsf{E}\,[\boldsymbol{\Sigma}']}$$

INDEPENDENCE ASSUMPTIONS

Equality (22.25) follows from the identities

$$
\begin{aligned}
\mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\boldsymbol{\Sigma}'}\right) &= \mathsf{E}\,\tilde{\boldsymbol{w}}^*_{i-1}\boldsymbol{\Sigma}'\tilde{\boldsymbol{w}}_{i-1} \\
&= \mathsf{E}\left(\mathsf{E}\left[\tilde{\boldsymbol{w}}^*_{i-1}\boldsymbol{\Sigma}'\tilde{\boldsymbol{w}}_{i-1}|\tilde{\boldsymbol{w}}_{i-1}\right]\right) \\
&= \mathsf{E}\,\tilde{\boldsymbol{w}}^*_{i-1}\left[\mathsf{E}\left(\boldsymbol{\Sigma}'|\tilde{\boldsymbol{w}}_{i-1}\right)\right]\tilde{\boldsymbol{w}}_{i-1} \\
&= \mathsf{E}\left(\tilde{\boldsymbol{w}}^*_{i-1}\left(\mathsf{E}\,\boldsymbol{\Sigma}'\right)\tilde{\boldsymbol{w}}_{i-1}\right) \quad \text{because of (22.24)} \\
&= \mathsf{E}\left(\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\mathsf{E}\,[\boldsymbol{\Sigma}']}\right)
\end{aligned}
$$

Thus observe that the main value of the independence assumption (22.23) lies in guaranteeing that $\tilde{\boldsymbol{w}}_{i-1}$ is independent of $\boldsymbol{\Sigma}'$, in which case it is possible to use (22.25) and thereby simplify the subsequent derivations. We can see from the expression for $\boldsymbol{\Sigma}'$ in (22.20) that this same conclusion will hold if we replace condition (22.23) by the assumption that $\tilde{\boldsymbol{w}}_{i-1}$ is independent of $\boldsymbol{u}^*_i\boldsymbol{u}_i/g[\boldsymbol{u}_i]$.

INDEPENDENCE ASSUMPTIONS

In this way, recursion (22.21) is replaced by

$$\boxed{\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} = \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\Sigma'} \; + \; \mu^2\sigma_v^2\mathsf{E}\left(\frac{\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\right)} \qquad (22.26)$$

with two *deterministic* (not random) weighting matrices $\{\Sigma, \Sigma'\}$ and where, by evaluating the expectation of (22.20),

$$\boxed{\Sigma' \;\triangleq\; \Sigma \;-\; \mu\Sigma\mathsf{E}\left(\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right) \;-\; \mu\mathsf{E}\left(\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right)\Sigma \;+\; \mu^2\mathsf{E}\left(\frac{\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\boldsymbol{u}_i^*\boldsymbol{u}_i\right)} \qquad (22.27)$$

It is further argued in Prob. V.2 that $\Sigma' \geq 0$ when $\Sigma > 0$.

VARIANCE RELATION WITH INDEP.

Theorem 22.4 (Weighted variance relation with independence) For adaptive filters of the form (22.2), Hermitian nonnegative-definite matrices Σ, and for data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfying model (22.1) and the independence assumption (22.23), it holds that:

$$\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} = \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\Sigma'} \;+\; \mu^2\sigma_v^2\mathsf{E}\left(\frac{\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\right)$$

$$\Sigma' = \Sigma \;-\; \mu\Sigma\mathsf{E}\left(\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right) \;-\; \mu\mathsf{E}\left(\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right)\Sigma \;+\; \mu^2\mathsf{E}\left(\frac{\|\boldsymbol{u}_i\|^2_{\Sigma}}{g^2[\boldsymbol{u}_i]}\boldsymbol{u}_i^*\boldsymbol{u}_i\right)$$

Theorem 22.5 (Mean weight error recursion) For any adaptive filter of the form (22.2), and for any data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfying (22.1) and (22.23), it holds that

$$\mathsf{E}\,\tilde{\boldsymbol{w}}_i = \left[\mathrm{I} - \mu\mathsf{E}\left(\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right)\right]\mathsf{E}\,\tilde{\boldsymbol{w}}_{i-1} \qquad (22.29)$$

DATA MOMENTS

Observe that the expression for Σ' is data dependent only; i.e., it depends on $\boldsymbol{u}_i$ alone and does not depend on the weight-error vectors. In this way, the recursion for $\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma$ is decoupled from the computation of Σ' in the statement of theorem. Moreover, the expressions for $\mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_\Sigma$ and Σ' show that studying the transient behavior of an adaptive filter requires evaluating the three multivariate moments:

$$\boxed{\mathsf{E}\left(\frac{\|\boldsymbol{u}_i\|^2_\Sigma}{g^2[\boldsymbol{u}_i]}\right), \qquad \mathsf{E}\left(\frac{\boldsymbol{u}_i^*\boldsymbol{u}_i}{g[\boldsymbol{u}_i]}\right) \qquad \text{and} \qquad \mathsf{E}\left(\frac{\|\boldsymbol{u}_i\|^2_\Sigma}{g^2[\boldsymbol{u}_i]}\boldsymbol{u}_i^*\boldsymbol{u}_i\right)} \tag{22.28}$$

which are solely dependent on $\boldsymbol{u}_i$. Note further that the last moment in the above list appears multiplied by μ^2 in the expression for Σ'. What this means is that sometimes, when the step-size is sufficiently small, this last moment could be ignored in lieu of simplification; see the remark following Thm. 24.1 and also Probs. V.14 and V.38, where this observation is pursued in greater detail.

CHANGE OF COORDINATES

Evaluation of the moments (22.28), and the subsequent analysis, can at times be simplified if we introduce a convenient change of coordinates by appealing to the eigen-decomposition of $R_u = \mathsf{E}\,\boldsymbol{u}_i^*\boldsymbol{u}_i$. So let

$$\boxed{R_u = U\Lambda U^*} \tag{22.30}$$

where Λ is diagonal with the eigenvalues of R_u, $\Lambda = \text{diag}\{\lambda_k\}$, and U is unitary (i.e., it satisfies $UU^* = U^*U = \mathrm{I}$). Then define the transformed quantities:

$$\boxed{\overline{\boldsymbol{w}}_i \;\stackrel{\Delta}{=}\; U^*\tilde{\boldsymbol{w}}_i, \qquad \overline{\boldsymbol{u}}_i \;\stackrel{\Delta}{=}\; \boldsymbol{u}_i U, \qquad \overline{\Sigma} \;\stackrel{\Delta}{=}\; U^*\Sigma U} \tag{22.31}$$

Since U is unitary, it is easy to see that

$$\|\tilde{\boldsymbol{w}}_i\|_\Sigma^2 = \|\overline{\boldsymbol{w}}_i\|_{\overline{\Sigma}}^2 \quad \text{and} \quad \|\boldsymbol{u}_i\|_\Sigma^2 = \|\overline{\boldsymbol{u}}_i\|_{\overline{\Sigma}}^2 \tag{22.32}$$

CHANGE OF COORDINATES

For example,

$$\|\overline{w}_i\|^2_{\overline{\Sigma}} \;=\; \overline{w}_i^*\overline{\Sigma}\,\overline{w}_i \;=\; (\tilde{w}_i^*U)\cdot(U^*\Sigma U)\cdot(U^*\tilde{w}_i) \;=\; \tilde{w}_i^*\Sigma\tilde{w}_i \;=\; \|\tilde{w}_i\|^2_{\Sigma}$$

Likewise, for $\|\overline{u}_i\|^2_{\overline{\Sigma}}$. In the special case $\Sigma = \mathrm{I}$, we have $\overline{\Sigma} = \mathrm{I}$, $\|\overline{w}_i\|^2 = \|\tilde{w}_i\|^2$, and $\|\overline{u}_i\|^2 = \|u_i\|^2$.

Now under the change of variables (22.31), the variance relation of Thm. 22.4 retains the same form, namely

$$\boxed{\mathsf{E}\,\|\overline{w}_i\|^2_{\overline{\Sigma}} = \mathsf{E}\,\|\overline{w}_{i-1}\|^2_{\overline{\Sigma}'} \;+\; \mu^2\sigma_v^2\mathsf{E}\left[\frac{\|\overline{u}_i\|^2_{\overline{\Sigma}}}{g^2[\overline{u}_i]}\right]} \qquad (22.33)$$

where

$$\overline{\Sigma}' = U^*\Sigma' U \qquad (22.34)$$

The data nonlinearity $g[\cdot]$ is usually invariant under unitary transformations, i.e., $g[u_i] = g[\overline{u}_i]$. However, the invariance property of $g[\cdot]$ is not necessary

TRANSFORMED VARIANCE RELATION

Continuing, from the equation for Σ' in Thm. 22.4 we find that

$$\boxed{\overline{\Sigma}' = \overline{\Sigma} - \mu\overline{\Sigma}\mathsf{E}\left[\frac{\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i}{g[\overline{\boldsymbol{u}}_i]}\right] - \mu\mathsf{E}\left[\frac{\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i}{g[\overline{\boldsymbol{u}}_i]}\right]\overline{\Sigma} + \mu^2\mathsf{E}\left[\frac{\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}}}{g^2[\overline{\boldsymbol{u}}_i]}\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i\right]} \tag{22.35}$$

Theorem 22.6 (Transformed weighted-variance relation) For any adaptive filter of the form (22.2), any Hermitian nonnegative-definite matrix Σ, and for data $\{\boldsymbol{d}(i), \boldsymbol{u}_i\}$ satisfying model (22.1) and the independence assumption (22.23), it holds that:

$$\begin{aligned}
\mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\Sigma}} &= \mathsf{E}\,\|\overline{\boldsymbol{w}}_{i-1}\|^2_{\overline{\Sigma}'} + \mu^2\sigma_v^2\mathsf{E}\left[\frac{\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}}}{g^2[\overline{\boldsymbol{u}}_i]}\right] \\
\overline{\Sigma}' &= \overline{\Sigma} - \mu\overline{\Sigma}\mathsf{E}\left[\frac{\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i}{g[\overline{\boldsymbol{u}}_i]}\right] - \mu\mathsf{E}\left[\frac{\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i}{g[\overline{\boldsymbol{u}}_i]}\right]\overline{\Sigma} + \mu^2\mathsf{E}\left[\frac{\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}}}{g^2[\overline{\boldsymbol{u}}_i]}\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i\right]
\end{aligned}$$

where the transformed variables $\{\overline{\boldsymbol{w}}_i, \overline{\boldsymbol{u}}_i, \overline{\Sigma}, \overline{\Sigma}'\}$ are related to the original variables $\{\tilde{\boldsymbol{w}}_i, \boldsymbol{u}_i, \Sigma, \Sigma'\}$ via (22.31) and (22.34), so that $\mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\Sigma}} = \mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma}$.

TRANSFORMED MEAN RECURSION

Likewise, the transformed version of the mean-weight error recursion (22.29) is

$$\mathsf{E}\,\overline{\boldsymbol{w}}_i = \left[\mathrm{I} - \mu \mathsf{E}\left(\frac{\overline{\boldsymbol{u}}_i^* \overline{\boldsymbol{u}}_i}{g[\overline{\boldsymbol{u}}_i]}\right)\right] \mathsf{E}\,\overline{\boldsymbol{w}}_{i-1} \qquad (22.36)$$

LMS WITH GAUSSIAN REGRESSORS

We use the mean and variance relations of the last chapter to study the transient performance of the LMS algorithm,

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} + \mu \boldsymbol{u}_i^* \boldsymbol{e}(i) \qquad (23.1)$$

for which the data normalization in (22.2) is given by

$$g[\boldsymbol{u}_i] = 1 \qquad (23.2)$$

In this case, relations (22.26)–(22.27) and (22.29) become

$$\begin{cases} \mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2_{\Sigma} = \mathsf{E}\,\|\tilde{\boldsymbol{w}}_{i-1}\|^2_{\Sigma'} \;+\; \mu^2\sigma_v^2 \mathsf{E}\,\|\boldsymbol{u}_i\|^2_{\Sigma} \\ \Sigma' = \Sigma - \mu\Sigma\mathsf{E}\,[\boldsymbol{u}_i^*\boldsymbol{u}_i] - \mu\mathsf{E}\,[\boldsymbol{u}_i^*\boldsymbol{u}_i]\,\Sigma \;+\; \mu^2\mathsf{E}\,\left[\|\boldsymbol{u}_i\|^2_{\Sigma}\boldsymbol{u}_i^*\boldsymbol{u}_i\right] \\ \mathsf{E}\,\tilde{\boldsymbol{w}}_i = [\mathrm{I} - \mu\mathsf{E}\,(\boldsymbol{u}_i^*\boldsymbol{u}_i)]\,\mathsf{E}\,\tilde{\boldsymbol{w}}_{i-1} \end{cases} \qquad (23.3)$$

We therefore need to evaluate the three moments:

$$\mathsf{E}\,\boldsymbol{u}_i^*\boldsymbol{u}_i, \quad \mathsf{E}\,\|\boldsymbol{u}_i\|_\Sigma^2 \quad \text{and} \quad \mathsf{E}\,\|\boldsymbol{u}_i\|_\Sigma^2\boldsymbol{u}_i^*\boldsymbol{u}_i \tag{23.4}$$

The first two moments are obvious, and can be evaluated regardless of any assumed distribution for the regression data since

$$\mathsf{E}\,\boldsymbol{u}_i^*\boldsymbol{u}_i = R_u \quad \text{(by definition)} \tag{23.5}$$

$$\mathsf{E}\,\|\boldsymbol{u}_i\|_\Sigma^2 = \mathsf{E}\,\boldsymbol{u}_i\Sigma\boldsymbol{u}_i^* = \mathsf{E}\,\mathsf{Tr}(\boldsymbol{u}_i^*\boldsymbol{u}_i\Sigma) = \mathsf{Tr}(R_u\Sigma) \tag{23.6}$$

The difficulty lies in evaluating the last moment in (23.4). To do so, we shall treat two cases. First we treat the case of Gaussian regressors for which the last moment can be evaluated explicitly. Afterwards, we treat the general case of non-Gaussian regressors.

23.1 MEAN AND VARIANCE RELATIONS

We assume in this chapter that the regressors $\{\boldsymbol{u}_i\}$ arise from a circular Gaussian distribution with covariance matrix R_u (cf. App. A.5).We say *circular* because we are treating the general case of complex-valued regressors; otherwise, the circularity assumption is not needed. In the Gaussian case, as we shall explain ahead following (23.11), it is more convenient to work with the transformed versions (22.33)–(22.35) and (22.36) of the variance and mean relations, which for LMS are given by

$$\begin{cases} \mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\Sigma}} = \mathsf{E}\,\|\overline{\boldsymbol{w}}_{i-1}\|^2_{\overline{\Sigma}'} \;+\; \mu^2\sigma_v^2\mathsf{E}\,\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}} \\ \\ \overline{\Sigma}' = \overline{\Sigma} - \mu\overline{\Sigma}\mathsf{E}\left[\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i\right] - \mu\mathsf{E}\left[\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i\right]\overline{\Sigma} \;+\; \mu^2\mathsf{E}\left[\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}}\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i\right] \\ \\ \mathsf{E}\,\overline{\boldsymbol{w}}_i = \left[\mathrm{I} - \mu\mathsf{E}\left(\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i\right)\right]\mathsf{E}\,\overline{\boldsymbol{w}}_{i-1} \end{cases} \tag{23.7}$$

TRANSIENT PERFORMANCE OF LMS

The moments that we need to evaluate in the transformed domain are

$$\mathsf{E}\,\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i, \qquad \mathsf{E}\,\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}} \qquad \text{and} \qquad \mathsf{E}\,\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}}\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i \tag{23.8}$$

where the first two are again immediate to compute since

$$\mathsf{E}\;\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i \;=\; \Lambda \tag{23.9}$$

and

$$\mathsf{E}\,\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}} = \mathsf{E}\,\overline{\boldsymbol{u}}_i\overline{\Sigma}\overline{\boldsymbol{u}}_i^* \;=\; \mathsf{E}\,\mathsf{Tr}(\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i\overline{\Sigma}) \;=\; \mathsf{Tr}(\Lambda\overline{\Sigma}) \tag{23.10}$$

With regards to the last moment in (23.8), we use the fact that $\overline{\boldsymbol{u}}_i$ is circular Gaussian with a diagonal covariance matrix, and invoke the result of Lemma A.3, to write

$$\begin{aligned}
\mathsf{E}\,\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}}\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i &= \mathsf{E}\;\left(\overline{\boldsymbol{u}}_i\overline{\Sigma}\overline{\boldsymbol{u}}_i^*\right)\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i \\
&= \mathsf{E}\;\overline{\boldsymbol{u}}_i^*\left(\overline{\boldsymbol{u}}_i\overline{\Sigma}\overline{\boldsymbol{u}}_i^*\right)\overline{\boldsymbol{u}}_i \\
&= \Lambda\mathsf{Tr}(\overline{\Sigma}\Lambda) \;+\; \Lambda\overline{\Sigma}\Lambda
\end{aligned} \tag{23.11}$$

TRANSIENT PERFORMANCE OF LMS

Recall that the statement of Lemma A.3 requires the variable $\boldsymbol{z}$ to have a diagonal covariance matrix, which explains why we introduced the transformed vector $\overline{\boldsymbol{u}}_i$ and the transformed relations (23.7) in the Gaussian regressor case. Moreover, if the regressors were real-valued rather than complex-valued, then we would invoke Lemma A.2 and use instead

$$\mathsf{E}\,\|\overline{\boldsymbol{u}}_i\|^2_{\overline{\Sigma}}\overline{\boldsymbol{u}}_i^*\overline{\boldsymbol{u}}_i \;=\; \Lambda\mathsf{Tr}(\overline{\Sigma}\Lambda) \;+\; 2\Lambda\overline{\Sigma}\Lambda$$

with an additional factor of 2 compared to (23.11).

Using (23.9)–(23.11), recursions (23.7) become

$$\begin{aligned}
\mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\Sigma}} &= \mathsf{E}\,\|\overline{\boldsymbol{w}}_{i-1}\|^2_{\overline{\Sigma}'} \;+\; \mu^2\sigma_v^2\mathsf{Tr}(\Lambda\overline{\Sigma}) && (23.12)\\
\overline{\Sigma}' &= \overline{\Sigma} - \mu\overline{\Sigma}\Lambda - \mu\Lambda\overline{\Sigma} \;+\; \mu^2\left[\Lambda\mathsf{Tr}(\overline{\Sigma}\Lambda) \;+\; \Lambda\overline{\Sigma}\Lambda\right] && (23.13)\\
\mathsf{E}\,\overline{\boldsymbol{w}}_i &= (\mathrm{I} - \mu\Lambda)\,\mathsf{E}\,\overline{\boldsymbol{w}}_{i-1} && (23.14)
\end{aligned}$$

Observe the interesting fact that $\overline{\Sigma}'$ will be diagonal if $\overline{\Sigma}$ is. Now since we are free to choose Σ and, therefore, $\overline{\Sigma}$, we can assume that $\overline{\Sigma}$ is diagonal. Under these conditions, it is possible to rewrite (23.13) in a more compact form in terms of the diagonal entries of $\{\overline{\Sigma}, \overline{\Sigma}'\}$.

DIAGONAL NOTATION

To do so, we define the vectors

$$\boxed{\overline{\sigma} \stackrel{\Delta}{=} \text{diag}\{\overline{\Sigma}\} \qquad \text{and} \qquad \lambda \stackrel{\Delta}{=} \text{diag}\{\Lambda\}} \tag{23.15}$$

That is, $\{\overline{\sigma}, \lambda\}$ are $M \times 1$ vectors with the diagonal entries of the corresponding matrices; $\overline{\sigma}$ contains the diagonal entries of $\overline{\Sigma}$, while λ contains the diagonal entries of Λ. Actually, in this book, we shall use the notation diag$\{\cdot\}$ in two directions, both of which will be obvious from the context. Writing diag$\{A\}$, for an arbitrary matrix A, extracts the diagonal entries of A into a vector. This is the convention we used in (23.15) to define $\{\overline{\sigma}, \lambda\}$. On the other hand, writing diag$\{a\}$ for a column vector a, results in a diagonal matrix whose entries are obtained from a. Therefore, we shall also write, whenever necessary,

$$\boxed{\overline{\Sigma} = \text{diag}\{\overline{\sigma}\} \qquad \text{and} \qquad \Lambda = \text{diag}\{\lambda\}} \tag{23.16}$$

in order to recover $\{\overline{\Sigma}, \Lambda\}$ from $\{\overline{\sigma}, \lambda\}$.

LINEAR VECTOR RELATION

Now in terms of the vectors $\{\overline{\sigma}, \lambda\}$, it is easy to see that the matrix relation (23.13) is equivalent to the vector relation

$$\overline{\sigma}' = \left(\mathrm{I} - 2\mu\Lambda + \mu^2\Lambda^2\right)\overline{\sigma} \; + \; \mu^2\left(\lambda^{\mathsf{T}}\overline{\sigma}\right)\lambda$$

which can in turn be written more compactly as

$$\boxed{\overline{\sigma}' = \overline{F}\,\overline{\sigma}} \tag{23.17}$$

with an $M \times M$ coefficient matrix $\overline{F}$ defined by

$$\boxed{\overline{F} \;\stackrel{\Delta}{=}\; \left(\mathrm{I} - 2\mu\Lambda + \mu^2\Lambda^2\right) + \mu^2\lambda\lambda^{\mathsf{T}}} \tag{23.18}$$

Expression (23.17) shows that the relation between the diagonal elements of $\overline{\Sigma}$ and $\overline{\Sigma}'$ is actually *linear*. Moreover, since $\overline{\Sigma}' = \mathrm{diag}\{\overline{\sigma}'\}$, the linear relation (23.17) translates into the matrix relation $\overline{\Sigma}' = \mathrm{diag}\left\{\overline{F}\,\overline{\sigma}\right\}$.

VARIANCE RELATION

We can rewrite recursion (23.12) by using the vectors $\{\overline{\sigma}, \overline{\sigma}'\}$ instead of the matrices $\{\overline{\Sigma}, \overline{\Sigma}'\}$. Using (23.17) and the notation (23.16), recursion (23.12) is equivalent to

$$\mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2_{\text{diag}\{\overline{\sigma}\}} \;=\; \mathsf{E}\,\|\overline{\boldsymbol{w}}_{i-1}\|^2_{\text{diag}\{\overline{F\sigma}\}} \;+\; \mu^2\sigma_v^2(\lambda^{\mathsf{T}}\overline{\sigma})$$

where, for the last term, we used the fact that $\mathsf{Tr}(\Lambda\overline{\Sigma}) = \lambda^{\mathsf{T}}\overline{\sigma}$. For compactness of notation, we shall drop the diag$\{\cdot\}$ notation from the subscripts and keep the vectors only, so that the above will be rewritten more compactly as

$$\boxed{\mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\sigma}} = \mathsf{E}\,\|\overline{\boldsymbol{w}}_{i-1}\|^2_{\overline{F\sigma}} \;+\; \mu^2\sigma_v^2(\lambda^{\mathsf{T}}\overline{\sigma})} \qquad (23.19)$$

The vector weighting factors $\{\overline{\sigma}, \overline{F\sigma}\}$ in this expression should be understood as compact representations for the actual weighting matrices $\{\text{diag}\{\overline{\sigma}\}, \text{diag}\{\overline{F\sigma}\}\}$. In other words, if σ is any column vector, the notation $\|x\|^2_\sigma$ is used to mean

$$\|x\|^2_\sigma \;\triangleq\; \|x\|^2_{\text{diag}\{\sigma\}} \;=\; x^*\Sigma x, \quad \text{where} \quad \Sigma = \text{diag}\{\sigma\}$$

TRANSIENT PERFORMANCE OF LMS

In summary, starting from (23.12)–(23.14), we argued that for Gaussian regressors the weighting matrix $\overline{\Sigma}'$ is diagonal if $\overline{\Sigma}$ is chosen as diagonal, so that (23.12)–(23.13) can be equivalently expressed more compactly as in (23.17)–(23.18) and (23.19), namely,

$$\mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2_{\overline{\sigma}} = \mathsf{E}\,\|\overline{\boldsymbol{w}}_{i-1}\|^2_{\overline{F}\overline{\sigma}} + \mu^2\sigma_v^2(\lambda^{\mathsf{T}}\overline{\sigma}) \tag{23.20}$$

$$\overline{\sigma}' = \overline{F}\overline{\sigma} \tag{23.21}$$

$$\overline{F} = \left(\mathrm{I} - 2\mu\Lambda + \mu^2\Lambda^2\right) + \mu^2\lambda\lambda^{\mathsf{T}} \tag{23.22}$$

$$\mathsf{E}\,\overline{\boldsymbol{w}}_i = [\mathrm{I} - \mu\Lambda]\,\mathsf{E}\,\overline{\boldsymbol{w}}_{i-1} \tag{23.23}$$

Stability and performance analyses are now possible to pursue by using these relations. Recall that in transient analysis we are interested in the time evolution of the expectations $\{\mathsf{E}\,\tilde{\boldsymbol{w}}_i, \mathsf{E}\,\|\tilde{\boldsymbol{w}}_i\|^2\}$ or, equivalently, $\{\mathsf{E}\,\overline{\boldsymbol{w}}_i, \mathsf{E}\,\|\overline{\boldsymbol{w}}_i\|^2\}$ since $\overline{\boldsymbol{w}}_i$ and $\tilde{\boldsymbol{w}}_i$ are related via a unitary matrix as in (22.31). We start with the mean behavior.