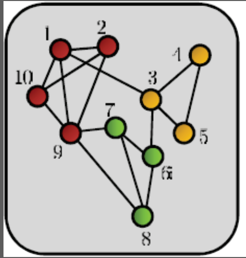
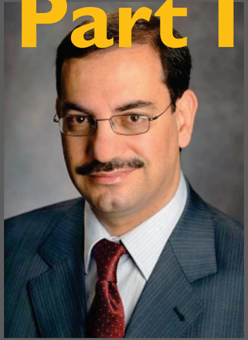


INFERENCE OVER NETWORKS

LECTURE #20: Performance of Multi-Agent Networks, Part I



Professor Ali H. Sayed
UCLA Electrical Engineering

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Reference

<u>Chapter 11 (Performance of Multi-Agent Networks, pp. 574-621):</u>

A. H. Sayed, ``Adaptation, learning, and optimization over networks," ***Foundations and Trends in Machine Learning***, vol. 7, issue 4-5, pp. 311-801, NOW Publishers, 2014.

Setting

We established in Theorem 9.1 that a multi-agent network running the distributed strategy (8.46) is mean-square stable for sufficiently small step-size parameters. More specifically, we showed that, for each agent k, the error variance relative to the limit point, $w^\star$, defined by (8.55), enters a bounded region whose size is in the order of $O(\mu_{\max})$:

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \;=\; O(\mu_{\max}), \qquad k = 1, 2, \ldots, N \tag{11.1}$$

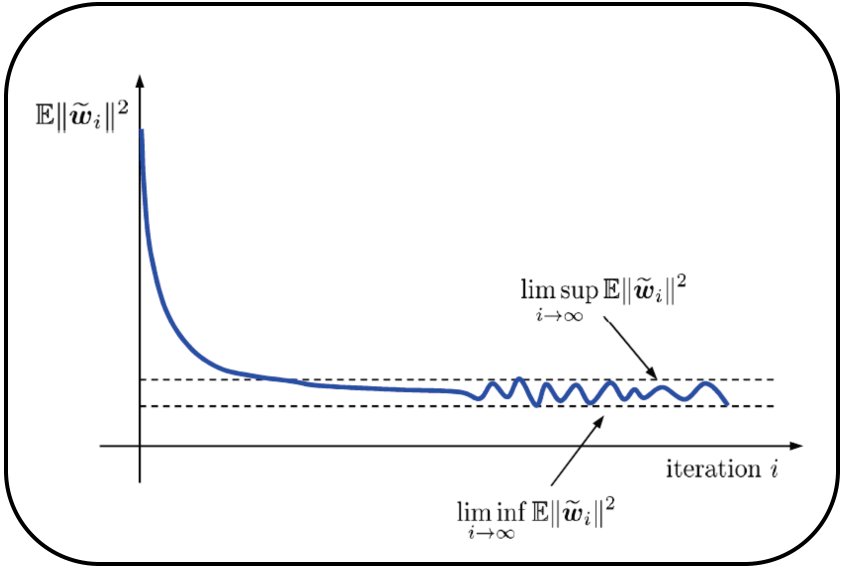
In this chapter, we will assess the size of these mean-square errors for both cases of real and complex data.

Recall#1: Limits Superior & Inferior

For single-agent adaptation:

$$\begin{cases} \limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{w}_i\|^2 &= \mu\cdot\overline{\text{MSD}} + o(\mu) \\ \liminf_{i\to\infty} \mathbb{E}\,\|\widetilde{w}_i\|^2 &= \mu\cdot\overline{\text{MSD}} - o(\mu) \end{cases}$$

for some positive constant $\overline{\text{MSD}}$.



Recall#1: Mean-Square-Deviation

$$\lim_{\mu\to 0}\left(\limsup_{i\to\infty}\frac{1}{\mu}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2\right)=\lim_{\mu\to 0}\left(\liminf_{i\to\infty}\frac{1}{\mu}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2\right)=\overline{\text{MSD}}$$

- Therefore, as $\mu\to 0$, the quantity $\frac{1}{\mu}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$ approaches a limit.

- Once multiplied by μ, this limit assesses $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$ to first-order in μ:

$$\text{MSD}\;\stackrel{\Delta}{=}\;\mu\cdot\left(\lim_{\mu\to 0}\limsup_{i\to\infty}\frac{1}{\mu}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2\right)$$

(simplification)

$$\text{MSD}\;\stackrel{\Delta}{=}\;\lim_{i\to\infty}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$$

Definition: MSD over Networks

$$\text{MSD}_{\text{dist},k} \triangleq \mu_{\max} \cdot \left(\lim_{\mu_{\max} \to 0} \limsup_{i \to \infty} \frac{1}{\mu_{\max}} \mathbb{E} \, \|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \right) \quad (11.2)$$

$$\text{MSD}_{\text{dist,av}} \triangleq \frac{1}{N} \left(\sum_{k=1}^{N} \text{MSD}_{\text{dist},k} \right) \quad (11.3)$$

Conditions on Costs and Noise

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Conditions

The presentation will assume the same conditions we used in the last two chapters to examine the stability of multi-agent networks. In particular, we assume the aggregate cost (9.10) and the individual costs, $J_k(w)$, satisfy the conditions in Assumptions 6.1 and 10.1. We also assume that the first and fourth-order moments of the gradient noise process satisfy the conditions of Assumption 8.1 with the second-order moment condition (8.115) replaced by the fourth-order moment condition (8.121), in addition to a smoothness condition on the noise covariance matrices defined as follows.

Gradient Noise

We refer to the definition of the individual gradient noise processes in (8.109), namely, for any $\phi \in \boldsymbol{\mathcal{F}}_{i-1}$:

$$\boldsymbol{s}_{k,i}(\phi) \;\triangleq\; \widehat{\nabla_{w^*} J}_k(\phi) \;-\; \nabla_{w^*} J_k(\phi) \qquad (11.4)$$

where $\boldsymbol{\mathcal{F}}_{i-1}$ denotes the filtration corresponding to all past iterates across all agents:

$$\boldsymbol{\mathcal{F}}_{i-1} \;=\; \text{filtration defined by } \{\boldsymbol{w}_{k,j},\; j \leq i-1,\; k = 1,2,\ldots,N\} \qquad (11.5)$$

Noise Covariance

We define the extended gradient noise vector of size $2M \times 1$:

$$\boldsymbol{s}^{e}_{k,i}(\phi) \overset{\Delta}{=} \begin{bmatrix} \boldsymbol{s}_{k,i}(\phi) \\ \left(\boldsymbol{s}^{*}_{k,i}(\phi)\right)^{\mathsf{T}} \end{bmatrix} \tag{11.6}$$

We denote its conditional covariance matrix by

$$R^{e}_{s,k,i}(\phi) \overset{\Delta}{=} \mathbb{E}\left[\boldsymbol{s}^{e}_{k,i}(\phi)\boldsymbol{s}^{e*}_{k,i}(\phi)\,|\,\boldsymbol{\mathcal{F}}_{i-1}\right] \tag{11.7}$$

Noise Covariance

We further assume that, in the limit, the following moment matrices tend to constant values when evaluated at the limit point $w^\star$:

$$R_{s,k} \triangleq \lim_{i\to\infty} \mathbb{E}\left[\boldsymbol{s}_{k,i}(w^\star)\boldsymbol{s}^*_{k,i}(w^\star) \,|\, \boldsymbol{\mathcal{F}}_{i-1} \right] \tag{11.8}$$

$$R_{q,k} \triangleq \lim_{i\to\infty} \mathbb{E}\left[\boldsymbol{s}_{k,i}(w^\star)\boldsymbol{s}^\mathsf{T}_{k,i}(w^\star) \,|\, \boldsymbol{\mathcal{F}}_{i-1} \right] \tag{11.9}$$

Definition

Definition 11.1 (Hessian and moment matrices). We associate with each agent k a pair of matrices $\{H_k, G_k\}$, both of which are evaluated at the location of the limit point $w = w^\star$. The matrices are defined as follows:

$$H_k \triangleq \nabla_w^2 \, J_k(w^\star), \quad G_k \triangleq \begin{cases} R_{s,k} & \text{(real case)} \\ \begin{bmatrix} R_{s,k} & R_{q,k} \\ R_{q,k}^* & R_{s,k}^\mathsf{T} \end{bmatrix} & \text{(complex case)} \end{cases} \tag{11.12}$$

Both matrices are dependent on the data type (whether real or complex); in particular, each H_k is $2M \times 2M$ for complex data and $M \times M$ for real data. Note that $H_k \geq 0$ and $G_k \geq 0$.

Aggregate Hessian Matrix

In view of the lower bound condition in (6.13), it follows that

$$\sum_{k=1}^{N} q_k H_k \; > \; 0 \qquad (11.13)$$

so that the weighted sum of the $\{H_k\}$ matrices is invertible. This matrix sum will appear in the performance expressions.

Condition

Assumption 11.1 (Smoothness condition on noise covariance). It is assumed that the conditional second-order moments of the individual noise processes satisfy smoothness conditions similar to (5.37), namely,

$$\left\|R^e_{s,k,i}(w^\star + \Delta w) - R^e_{s,k,i}(w^\star)\right\| \;\le\; \kappa_d \,\|\Delta w\|^\gamma \qquad (11.10)$$

in terms of the extended covariance matrix, for small perturbations $\|\Delta w\| \le \epsilon$, and for some constants $\kappa_d \ge 0$ and exponent $0 < \gamma \le 4$.

Condition

Following the argument that led to (4.24) in the single-agent case, we can similarly show that the conditional noise covariance matrix satisfies more globally a condition of the following form for all $\boldsymbol{\phi} \in \boldsymbol{\mathcal{F}}_{i-1}$:

$$\left\| R^e_{s,k,i}(\boldsymbol{\phi}) - R^e_{s,k,i}(w^\star) \right\| \quad \leq \quad \kappa_d \, \|\widetilde{\boldsymbol{\phi}}\|^\gamma + \kappa'_d \, \|\widetilde{\boldsymbol{\phi}}\|^2 \qquad (11.11)$$

where $\widetilde{\boldsymbol{\phi}} = w^\star - \boldsymbol{\phi}$ and for some constant $\kappa'_d \geq 0$.

The performance expressions that will be derived in this chapter will be expressed in terms of the following quantities, defined for both cases of real or complex data.

Second-Order Moment

Lemma 11.1 (Limiting second-order moment of gradient noise). Under the smoothness condition (11.10), and for sufficiently small step-sizes, it holds that the covariance matrix of the extended gradient noise process, $\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})$, at each agent k satisfies for $i \gg 1$:

$$\mathbb{E}\,\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* \;=\; G_k \;+\; O\left(\mu^{\min\{1,\frac{\gamma}{2}\}}\right) \qquad (11.14)$$

where $0 < \gamma \leq 4$ and G_k is given by (11.12). Consequently, it holds for $i \gg 1$ that the trace of the covariance matrix satisfies:

$$\mathrm{Tr}(G_k) - b_o \;\leq\; \mathbb{E}\,\|\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\|^2 \;\leq\; \mathrm{Tr}(G_k) + b_o \qquad (11.15)$$

for some nonnegative value $b_o = O\left(\mu^{\min\{1,\frac{\gamma}{2}\}}\right)$.

Proof

Proof. By adding and subtracting the same term we have [71, 278]:

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right] \\
&\quad = \quad \mathbb{E}\left[\boldsymbol{s}^e_{k,i}(w^\star)\left(\boldsymbol{s}^e_{k,i}(w^\star)\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right] + \\
&\qquad\quad\; \mathbb{E}\left[\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right] - \\
&\qquad\quad\; \mathbb{E}\left[\boldsymbol{s}^e_{k,i}(w^\star)\left(\boldsymbol{s}^e_{k,i}(w^\star)\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right]
\end{aligned} \tag{11.16}$$

which, upon using definition (11.7), can be rewritten as:

Proof

$$\mathbb{E}\left[\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right]$$
$$= \mathbb{E}\left[\boldsymbol{s}^e_{k,i}(w^\star)\left(\boldsymbol{s}^e_{k,i}(w^\star)\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right] + R^e_{s,k,i}(\boldsymbol{\phi}_{k,i-1}) - R^e_{s,k,i}(w^\star) \quad (11.17)$$

Subtracting the covariance matrix G_k defined by (11.12) from both sides, and computing expectations, we get:

$$\mathbb{E}\,\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* - G_k$$
$$= \mathbb{E}\left(\mathbb{E}\left[\boldsymbol{s}^e_{k,i}(w^\star)\left(\boldsymbol{s}^e_{k,i}(w^\star)\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right] - G_k\right) + \mathbb{E}\left(R^e_{s,k,i}(\boldsymbol{\phi}_{k,i-1}) - R^e_{s,k,i}(w^\star)\right) \quad (11.18)$$

Proof

It then follows from the triangle inequality of norms, and from Jensen's inequality (F.29) in the appendix, that

$$
\begin{aligned}
&\left\| \mathbb{E}\, \boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1}) \left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* - G_k \right\| \\
&\qquad \le \left\| \mathbb{E}\left(\mathbb{E}\left[\boldsymbol{s}^e_{k,i}(w^\star) \left(\boldsymbol{s}^e_{k,i}(w^\star)\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1} \right] - G_k \right) \right\| + \\
&\qquad\qquad \left\| \mathbb{E}\left(R^e_{s,k,i}(\boldsymbol{\phi}_{k,i-1}) - R^e_{s,k,i}(w^\star) \right) \right\| \\
&\qquad \overset{(\text{F.29})}{\le} \mathbb{E} \left\| \mathbb{E}\left[\boldsymbol{s}^e_{k,i}(w^\star) \left(\boldsymbol{s}^e_{k,i}(w^\star)\right)^* \,|\, \boldsymbol{\mathcal{F}}_{i-1} \right] - G_k \right\| + \\
&\qquad\qquad \mathbb{E} \left\| R^e_{s,k,i}(\boldsymbol{\phi}_{k,i-1}) - R^e_{s,k,i}(w^\star) \right\| \qquad (11.19)
\end{aligned}
$$

Proof

Computing the limit superior of both sides, and using (11.8)–(11.9) to annihilate the limit of the first term on the right-hand side, we conclude that

$$\limsup_{i\to\infty} \left\| \mathbb{E}\, \boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1}) \left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* - G_k \right\| \leq$$

$$\limsup_{i\to\infty} \mathbb{E} \left\| R^e_{s,k,i}(\boldsymbol{\phi}_{k,i-1}) - R^e_{s,k,i}(w^\star) \right\| \qquad (11.20)$$

We next use the smoothness condition (11.11) to bound the right-most term as follows:

Proof

$$\|R^e_{s,k,i}(\phi_{k,i-1}) - R^e_{s,i}(w^\star)\| \leq \kappa_d \|\widetilde{\phi}_{k,i-1}\|^\gamma + \kappa'_d \|\widetilde{\phi}_{k,i-1}\|^2 \tag{11.21}$$

where

$$\widetilde{\phi}_{k,i-1} \triangleq w^\star - \phi_{k,i-1} \tag{11.22}$$

Recall from the distributed algorithm (8.46) that

$$\widetilde{\phi}_{k,i-1} = \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \widetilde{\boldsymbol{w}}_{\ell,i-1} \tag{11.23}$$

so that exploiting the convexity of the functions $f(x) = x^2$ and $f(x) = x^4$, and applying Jensen's inequality (F.26), we get:

Proof

$$
\begin{aligned}
\|\widetilde{\phi}_{k,i-1}\|^2 &= \left\| \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \widetilde{\boldsymbol{w}}_{\ell,i-1} \right\|^2 \\
&\overset{(F.26)}{\leq} \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2 \\
&\leq \sum_{\ell \in \mathcal{N}_k} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2 \\
&\leq \sum_{\ell=1}^{N} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2 \qquad (11.24)
\end{aligned}
$$

Proof

Likewise, we have

$$\|\widetilde{\phi}_{k,i-1}\|^4 \;\leq\; \sum_{\ell=1}^{N} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^4 \tag{11.25}$$

and since the function $f(x) = x^{\gamma/4}$ is increasing over $x \geq 0$:

$$\|\widetilde{\phi}_{k,i-1}\|^\gamma \;=\; \left(\|\widetilde{\phi}_{k,i-1}\|^4\right)^{\gamma/4} \;\leq\; \left(\sum_{\ell=1}^{N} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^4\right)^{\gamma/4} \tag{11.26}$$

Substituting (11.24) and (11.26) into (11.21), we obtain

Proof

$$\|R^e_{s,k,i}(\phi_{k,i-1}) - R^e_{s,i}(w^\star)\| \leq \kappa_d \left(\sum_{\ell=1}^{N} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^4\right)^{\gamma/4} + \kappa'_d \left(\sum_{\ell=1}^{N} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2\right) \tag{11.27}$$

Using arguments similar to the steps that led to (4.31) in the single-agent case, we find under expectation and in the limit that:

Proof

$$\begin{aligned}
&\limsup_{i\to\infty} \mathbb{E}\,\|R^e_{s,k,i}(\phi_{i-1}) - R^e_{s,k,i}(w^\star)\| \\
&\leq \quad \limsup_{i\to\infty} \left\{ \kappa_d\, \mathbb{E}\left(\sum_{\ell=1}^{N} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^4 \right)^{\gamma/4} + \kappa_d'\, \mathbb{E}\left(\sum_{\ell=1}^{N} \|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2 \right) \right\} \\
&\overset{(a)}{\leq} \quad \limsup_{i\to\infty} \left\{ \kappa_d \left(\sum_{\ell=1}^{N} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^4 \right)^{\gamma/4} + \kappa_d' \left(\sum_{\ell=1}^{N} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{\ell,i-1}\|^2 \right) \right\} \\
&\overset{(9.11)}{=} \quad O(\mu_{\max}^{\gamma'/2}) \qquad (11.28)
\end{aligned}$$

Proof

where in step (a) we applied Jensen's inequality (F.30) to the function $f(x) = x^{\gamma/4}$; this function is concave over $x \geq 0$ for $\gamma \in (0, 4]$. Moreover, in the last step we called upon results (9.11) and (9.107), namely, that the second and fourth-order moments of $\widetilde{\boldsymbol{w}}_{\ell,i-1}$ are asymptotically bounded by $O(\mu_{\max})$ and $O(\mu_{\max}^2)$, respectively. Accordingly, the exponent γ' in the last step is given by

$$\gamma' \stackrel{\Delta}{=} \min\{\gamma,\, 2\} \tag{11.29}$$

since $O(\mu_{\max}^{\gamma/2})$ dominates $O(\mu_{\max})$ for values of $\gamma \in (0, 2]$ and $O(\mu_{\max})$ dominates $O(\mu_{\max}^{\gamma/2})$ for values of $\gamma \in [2, 4]$. Substituting (11.28) into (11.20) gives

Proof

$$\limsup_{i\to\infty} \left\| \mathbb{E}\, \boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1}) \left(\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\right)^* \;-\; G_k \right\| \;=\; O(\mu_{\max}^{\gamma'/2}) \qquad (11.30)$$

which leads to (11.14). Moreover, since for any square matrix X, it holds that $|\mathrm{Tr}(X)| \leq c\,\|X\|$, for some constant c that is independent of γ', we conclude that

$$\limsup_{i\to\infty} \left| \mathbb{E}\, \|\boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1})\|^2 - \mathrm{Tr}\left(G_k\right) \right| \;=\; O(\mu_{\max}^{\gamma'/2}) \;=\; b_1 \qquad (11.31)$$

Proof

in terms of the absolute value of the difference. We are denoting the value of the limit superior by the nonnegative number b_1; we know from (11.31) that $b_1 = O(\mu^{\gamma'/2})$. The above relation then implies that, given $\epsilon > 0$, there exists an I_o large enough such that for all $i > I_o$ it holds that

$$\Big| \mathbb{E} \| \boldsymbol{s}^e_{k,i}(\boldsymbol{\phi}_{k,i-1}) \|^2 - \text{Tr}(G_k) \Big| \;\leq\; b_1 + \epsilon \tag{11.32}$$

If we select $\epsilon = O(\mu^{\gamma'/2})$ and introduce the sum $b_o = b_1 + \epsilon$, then we arrive at the desired result (11.15).

$\square$

Performance Metrics

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

Definition: Mean-Square-Deviation

$$\text{MSD}_{\text{dist},k} \triangleq \mu_{\max} \cdot \left(\lim_{\mu_{\max} \to 0} \limsup_{i \to \infty} \frac{1}{\mu_{\max}} \mathbb{E} \, \|\widetilde{\boldsymbol{w}}_{k,i}\|^2 \right) \qquad (11.2)$$

$$\text{MSD}_{\text{dist,av}} \triangleq \frac{1}{N} \left(\sum_{k=1}^{N} \text{MSD}_{\text{dist},k} \right) \qquad (11.3)$$

Recall#2: Weighted Aggregate Cost

$$J^{\text{glob}}(w) \triangleq \sum_{k=1}^{N} J_k(w) \longrightarrow \boxed{\nabla_w J^{\text{glob}}(w^o) = 0 \iff \sum_{k=1}^{N} \nabla_w J_k(w^o) = 0}$$

$$J^{\text{glob},\star}(w) \triangleq \sum_{k=1}^{N} q_k J_k(w) \longrightarrow \boxed{\nabla_w J^{\text{glob},\star}(w^\star) = 0 \iff \sum_{k=1}^{N} q_k \nabla_w J_k(w^\star) = 0}$$

(is also **strongly convex**)

$$q \triangleq \text{diag}\{\mu_1, \mu_2, \ldots, \mu_N\} A_2 p$$

Recall (Example #8.5)

Example 8.5 (Normalization of weights in aggregate cost). If desired, we may normalize the positive weighting coefficients $\{q_k\}$ defined by (8.50) to have their sum add up to one, say, by introducing instead the coefficients:

$$\bar{q}_k \;\triangleq\; q_k \Big/ \sum_{k=1}^{N} q_k \tag{8.58}$$

and replacing (8.53) by the convex combination:

$$\bar{J}^{\text{glob},\star}(w) \;\triangleq\; \sum_{k=1}^{N} \bar{q}_k J_k(w) \tag{8.59}$$

Recall (Example #8.5)

Clearly, both aggregate functions, $J^{\text{glob},\star}(w)$ defined by (8.53) and $\bar{J}^{\text{glob},\star}(w)$, are scaled multiples of each other and, hence, their unique minimizers occur at the same location $w^\star$. One advantage of working with the normalized aggregate cost (8.59) is that when all individual costs happen to coincide, say, $J_k(w) \equiv J(w)$, then expression (8.59) reduces to

$$\bar{J}^{\text{glob},\star}(w) = J(w) \tag{8.60}$$

whereas $J^{\text{glob},\star}(w)$ will be a scaled multiple of $J(w)$.

Definition: Excess-Risk

$$\text{ER}_{\text{dist},k} \stackrel{\Delta}{=} \tag{11.33}$$

$$\mu_{\max} \cdot \left(\lim_{\mu_{\max}\to 0} \limsup_{i\to\infty} \frac{1}{\mu_{\max}} \mathbb{E}\left\{ \bar{J}^{\text{glob},\star}(\boldsymbol{w}_{k,i-1}) - \bar{J}^{\text{glob},\star}(w^{\star}) \right\} \right)$$

The main difference in relation to (4.95) is that we are now scaling by $\mu_{\max}$ and using the normalized aggregate cost (8.59). The reason why we are using this normalized cost in (11.33), rather than the regular

Excess-Risk

aggregate cost $J^{\text{glob},\star}(w)$ from (9.6), is to ensure that the above definition of the excess-risk is compatible with the definition used earlier for non-cooperative agents in (4.95) and for centralized processing in (5.53). For example, when the individual costs happen to coincide, say, $J_k(w) \equiv J(w)$, then the expectation on the right-hand side of (11.33) reduces to $\mathbb{E}\{J(\boldsymbol{w}_{k,i-1}) - J(w^o)\}$, which is consistent with the earlier expression (4.95).

Excess-Risk

$$\text{ER}_{\text{dist},k} \;=\; \mu_{\max} \cdot \left(\lim_{\mu_{\max}\to 0} \limsup_{i\to\infty} \frac{1}{\mu_{\max}} \mathbb{E}\, \|\widetilde{\boldsymbol{w}}^{e}_{k,i-1}\|^2_{\frac{1}{2}\bar{H}} \right) \tag{11.35}$$

where the matrix $\bar{H}$ denotes the value of the Hessian matrix of the normalized cost, $\bar{J}^{\text{glob},\star}(w)$, evaluated at $w = w^\star$. It follows from (8.59) that this matrix is given by

$$\bar{H} \;\triangleq\; \sum_{k=1}^{N} \bar{q}_k H_k \tag{11.36}$$

Excess-Risk

We further define the network ER measure as the average ER values across all agents:

$$\text{ER}_{\text{dist,av}} \triangleq \frac{1}{N}\left(\sum_{k=1}^{N} \text{ER}_{\text{dist},k}\right) \tag{11.34}$$

It is straightforward to verify that the MSD and ER performance measures defined so far can be equivalently expressed as follows in terms of the extended error vectors $\{\widetilde{\boldsymbol{w}}^e_{k,i}, \widetilde{\boldsymbol{w}}^e_i\}$ defined by (8.133) and (8.143):

Performance Metrics

$$\text{MSD}_{\text{dist},k} \triangleq \mu_{\max} \cdot \left(\lim_{\mu_{\max}\to 0} \limsup_{i\to\infty} \frac{1}{\mu_{\max}} \frac{1}{2} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{k,i}\|^2 \right) \tag{11.37}$$

$$\text{MSD}_{\text{dist,av}} \triangleq \mu_{\max} \cdot \left(\lim_{\mu_{\max}\to 0} \limsup_{i\to\infty} \frac{1}{\mu_{\max}} \frac{1}{2N} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i}\|^2 \right) \tag{11.38}$$

$$\text{ER}_{\text{dist,av}} = \mu_{\max} \cdot \left(\lim_{\mu_{\max}\to 0} \limsup_{i\to\infty} \frac{1}{\mu_{\max}} \frac{1}{2N} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|^2_{(I_N \otimes \bar{H})} \right) \tag{11.39}$$

Compact Representations

$$\text{MSD}_{\text{dist},k} \triangleq \lim_{i\to\infty} \frac{1}{2}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^{e}_{k,i}\|^2 \tag{11.40}$$

$$\text{MSD}_{\text{dist,av}} \triangleq \lim_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^{e}_{i}\|^2 \tag{11.41}$$

$$\text{ER}_{\text{dist},k} = \lim_{i\to\infty} \frac{1}{2}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^{e}_{k,i-1}\|^2_{\bar{H}} \tag{11.42}$$

$$\text{ER}_{\text{dist,av}} = \lim_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^{e}_{i-1}\|^2_{(I_N\otimes\bar{H})} \tag{11.43}$$

MSE Performance

Course EE210B
Spring Quarter 2015

Proc. IEEE, vol. 102, no. 4, pp. 460-497, April 2014.
Foundations and Trends in Machine Learning, vol. 7, no. 4-5, pp. 311-801, July 2014.

MSE Performance

We examine first the mean-square-error performance of the multi-agent network and derive closed-form expressions for the MSD measures of the individual agents and the entire network. The expressions given below involve the bvec and block Kronecker operations defined in Sec. F.1 in the appendix.

MSE Performance

Theorem 11.2 (Network limiting performance). Consider a network of N interacting agents running the distributed strategy (8.46) with a primitive matrix $P = A_1 A_o A_2$. Assume the aggregate cost (9.10) and the individual costs, $J_k(w)$, satisfy the conditions in Assumptions 6.1 and 10.1. Assume further that the first and fourth-order moments of the gradient noise process satisfy the conditions of Assumption 8.1 with the second-order moment condition (8.115) replaced by the fourth-order moment condition (8.121). Assume also (11.10). Let

$$\gamma_m \triangleq \frac{1}{2} \min\{1, \gamma\} > 0 \tag{11.44}$$

MSE Performance

with $\gamma \in (0, 4]$ from (11.10). Then, it holds that

$$\limsup_{i\to\infty} \frac{1}{2}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{k,i}\|^2 = \frac{1}{h}\text{Tr}(\mathcal{J}_k\mathcal{X}) + O\left(\mu_{\max}^{1+\gamma_m}\right) \tag{11.45}$$

$$\limsup_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i}\|^2 = \frac{1}{hN}\text{Tr}(\mathcal{X}) + O\left(\mu_{\max}^{1+\gamma_m}\right) \tag{11.46}$$

and, for large enough i, the convergence rate of the error variances, $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2$, towards the steady-state region (11.45) is given by

MSE Performance

$$\alpha \;=\; 1 - 2\lambda_{\min}\left(\sum_{k=1}^{N} q_k H_k\right) + O\left(\mu_{\max}^{(N+1)/N}\right) \tag{11.47}$$

where q_k is defined by (9.7) and $\alpha \in (0,1)$; the smaller the value of α is, the faster the convergence of $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2$ towards (11.45). Moreover, the matrix $\mathcal{X}$ that appears in (11.45)–(11.46) is Hermitian non-negative definite and corresponds to the unique solution of the (discrete-time) Lypaunov equation:

$$\mathcal{X} - \mathcal{B}\mathcal{X}\mathcal{B}^* = \mathcal{Y} \tag{11.48}$$

where the quantities $\{\mathcal{Y}, \mathcal{B}, \mathcal{J}_k\}$ are defined by:

MSE Performance

$$\mathcal{A}_o = A_o \otimes I_{hM}, \quad \mathcal{A}_1 = A_1 \otimes I_{hM}, \quad \mathcal{A}_2 = A_2 \otimes I_{hM} \tag{11.49}$$

$$\mathcal{M} = \text{diag}\{\, \mu_1 I_{hM}, \mu_2 I_{hM}, \ldots, \mu_N I_{hM} \} \tag{11.50}$$

$$\mathcal{H} = \text{diag}\{\, H_1, H_2, \ldots, H_N \,\} \tag{11.51}$$

$$H_k = \nabla_w^2 J_k(w^\star) \tag{11.52}$$

$$\mathcal{S} = \text{diag}\{\, G_1,\ G_2, \ldots, G_N \,\} \tag{11.53}$$

$$\mathcal{Y} = \mathcal{A}_2^\mathsf{T} \mathcal{M} \mathcal{S} \mathcal{M} \mathcal{A}_2 \tag{11.54}$$

$$\mathcal{B} = \mathcal{A}_2^\mathsf{T} \left(\mathcal{A}_o^\mathsf{T} - \mathcal{M}\mathcal{H} \right) \mathcal{A}_1^\mathsf{T} \tag{11.55}$$

$$\mathcal{F} = \mathcal{B}^\mathsf{T} \otimes_b \mathcal{B}^* \tag{11.56}$$

$$\mathcal{J}_k = \text{diag}\{\, 0_{hM}, \ldots, 0_{hM}, I_{hM}, 0_{hM}, \ldots, 0_{hM} \,\} \tag{11.57}$$

MSE Performance

with $\mathcal{J}_k$ having an identity matrix at the $k-$th diagonal block, and $h = 1$ for real data and $h = 2$ for complex data. Furthermore, the following are equivalent characterizations for the matrix $\mathcal{X}$ or its trace:

$$\mathcal{X} = \sum_{n=0}^{\infty} \mathcal{B}^n \mathcal{Y} (\mathcal{B}^*)^n \quad (11.58)$$

$$\text{bvec}(\mathcal{X}) = (I - \mathcal{F}^*)^{-1} \text{bvec}(\mathcal{Y}) \quad (11.59)$$

$$\text{Tr}(\mathcal{X}) = \left(\text{bvec}\left(\mathcal{Y}^{\mathsf{T}}\right)\right)^{\mathsf{T}} (I - \mathcal{F})^{-1} \text{bvec}\left(I_{hMN}\right) \quad (11.60)$$

$$\text{Tr}(\mathcal{J}_k \mathcal{X}) = \left(\text{bvec}\left(\mathcal{Y}^{\mathsf{T}}\right)\right)^{\mathsf{T}} (I - \mathcal{F})^{-1} \text{bvec}\left(\mathcal{J}_k\right) \quad (11.61)$$

Proof

Proof. We start from the long-term model (10.19), namely,

$$\widetilde{\boldsymbol{w}}_i^{e'} = \mathcal{B}\,\widetilde{\boldsymbol{w}}_{i-1}^{e'} + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \tag{11.62}$$

We drop the argument $\boldsymbol{w}_{i-1}^e$ from $\boldsymbol{s}_i^e$ for compactness of presentation. Conditioning on the past history and taking expectations gives

$$\mathbb{E}\left(\widetilde{\boldsymbol{w}}_i^{e'}\,|\,\boldsymbol{\mathcal{F}}_{i-1}\right) = \mathcal{B}\,\widetilde{\boldsymbol{w}}_{i-1}^{e'} - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \tag{11.63}$$

so that taking expectations again we obtain the mean recursion:

$$\mathbb{E}\,\widetilde{\boldsymbol{w}}_i^{e'} = \mathcal{B}\left(\mathbb{E}\,\widetilde{\boldsymbol{w}}_{i-1}^{e'}\right) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \tag{11.64}$$

Proof

Now observe that recursion (11.62) includes a constant driving term on the right-hand side represented by the factor $\mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e$. To facilitate the variance analysis, we introduce the centered variable:

$$\boldsymbol{z}_i \;\stackrel{\Delta}{=}\; \widetilde{\boldsymbol{w}}_i^{e'} - \mathbb{E}\,\widetilde{\boldsymbol{w}}_i^{e'} \tag{11.65}$$

Subtracting (11.64) from (11.62) we find that $\boldsymbol{z}_i$ satisfies the following recursion:

$$\boldsymbol{z}_i \;=\; \mathcal{B}\boldsymbol{z}_{i-1} \;+\; \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) \tag{11.66}$$

Proof

where the deterministic driving terms are also removed. Although we are interested in evaluating the asymptotic size of $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'}\|^2$, we can still rely on the centered variable $\boldsymbol{z}_i$ for this purpose. This is because it holds for $i \gg 1$:

$$\begin{aligned} \mathbb{E}\,\|\boldsymbol{z}_i\|^2 &= \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'}\|^2 - \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_i^{e'}\|^2 \\ &\overset{(10.108)}{=} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'}\|^2 + O(\mu_{\max}^2) \end{aligned} \qquad (11.67)$$

Moreover, we established earlier in (10.30) that under the fourth-order moment condition (8.121) on the gradient noise processes, the error variances $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'}\|^2$ and $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e}\|^2$ are within $O(\mu_{\max}^{3/2})$ from each other. Therefore, we

Proof

may evaluate the network error variance (or MSD) in terms of the mean-square value of the variable $\boldsymbol{z}_i$ (similarly for any weighted square measure of $\widetilde{\boldsymbol{w}}_i^e$ such as the ER) by employing the correction:

$$\limsup_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^2 \;=\; \limsup_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\boldsymbol{z}_i\|^2 \;+\; O(\mu_{\max}^{3/2}) \qquad (11.68)$$

We therefore continue with recursion (11.66) and proceed to examine how the mean-square value of $\boldsymbol{z}_i$ evolves over time by relying on energy conservation arguments [6, 205, 206, 269, 278].

Proof

Let Σ denote an arbitrary Hermitian positive semi-definite matrix that we are free to choose. Equating the squared weighted values of both sides of (11.66) and taking expectations conditioned on the past history gives:

$$\mathbb{E}\left(\|\boldsymbol{z}_i\|^2_{\Sigma}\,|\,\boldsymbol{\mathcal{F}}_{i-1}\right) \;=\; \|\boldsymbol{z}_{i-1}\|^2_{\mathcal{B}^*\Sigma\mathcal{B}} + \mathbb{E}\left(\|\boldsymbol{s}_i^e\|^2_{\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}}\,|\,\boldsymbol{\mathcal{F}}_{i-1}\right) \qquad (11.69)$$

Taking expectations again removes the conditioning on $\boldsymbol{\mathcal{F}}_{i-1}$ and we get

$$\mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma} \;=\; \mathbb{E}\left(\|\boldsymbol{z}_{i-1}\|^2_{\mathcal{B}^*\Sigma\mathcal{B}}\right) + \mathbb{E}\left(\|\boldsymbol{s}_i^e\|^2_{\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}}\right) \qquad (11.70)$$

We now evaluate the right-most term. For that purpose, we shall call upon the results of Lemma 11.1. To begin with, note that

Proof

$$\mathbb{E}\left(\|\boldsymbol{s}_i^e\|^2_{\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}}\right) = \mathrm{Tr}\left[\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\,\mathbb{E}\left(\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\left(\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\right)^*\right)\right] \tag{11.71}$$

where the entries of the covariance matrix $\mathbb{E}\,\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\left(\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\right)^*$ that appears in the above expression were already evaluated earlier in (11.14). Using that result, and the fact that the gradient noises across the agents are uncorrelated with each other and second-order circular, we obtain

$$\limsup_{i\to\infty}\left\|\mathbb{E}\,\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\left(\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\right)^* - \mathcal{S}\right\| = O(\mu_{\max}^{\gamma'/2}) \tag{11.72}$$

$$\gamma' = \min\{\gamma, 2\}$$

Proof

where γ' was defined in (11.29) as $\gamma' = \min\{\gamma, 2\}$. Using the sub-multiplicative property of norms, namely, $\|AB\| \leq \|A\|\,\|B\|$, we conclude from (11.72) that

$$\limsup_{i\to\infty} \left\| \mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\left(\mathbb{E}\,\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\left(\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)\right)^* - \mathcal{S}\right)\right\| = \mathrm{Tr}(\Sigma)\cdot O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \qquad (11.73)$$

where an additional factor $\mu_{\max}^2$ has been added to the big-O term; it arises from the fact that $\|\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\| = \mathrm{Tr}(\Sigma)\cdot O(\mu_{\max}^2)$. Note that we are keeping the factor $\mathrm{Tr}(\Sigma)$ explicit on the right-hand side of (11.73); this is convenient

Proof

for later use in (11.92) — the reason we have $\text{Tr}(\Sigma)$ in (11.73) is because $\|\Sigma\| \leq \text{Tr}(\Sigma)$ for any Hermitian positive semi-definite Σ. Using again the fact that $|\text{Tr}(X)| \leq c\|X\|$ for any square matrix X, we conclude that

$$\limsup_{i\to\infty} \left| \mathbb{E}\,\|\boldsymbol{s}_i^e\|^2_{\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}} - \text{Tr}\left(\Sigma\mathcal{Y}\right)\right| \;=\; \text{Tr}(\Sigma)\cdot O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \;=\; b_1 \tag{11.74}$$

in terms of the absolute value of the difference and where we are denoting the value of the limit superior by the nonnegative number b_1; we know from (11.74) that $b_1 = \text{Tr}(\Sigma)\cdot O(\mu_{\max}^{2+(\gamma'/2)})$. The same argument that led to (11.15) then gives for $i \gg 1$:

Proof

$$\text{Tr}(\Sigma\mathcal{Y}) \;-\; b_o \;\leq\; \mathbb{E}\left(\|\boldsymbol{s}_i^e\|^2_{\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}}\right) \;\leq\; \text{Tr}(\Sigma\mathcal{Y}) \;+\; b_o \qquad (11.75)$$

for some nonnegative constant $b_o = \text{Tr}(\Sigma)\cdot O(\mu_{\max}^{2+(\gamma'/2)})$. It follows from (11.75) that we can also write for $i \gg 1$:

$$\mathbb{E}\left(\|\boldsymbol{s}_i^e\|^2_{\mathcal{M}\mathcal{A}_2\Sigma\mathcal{A}_2^{\mathsf{T}}\mathcal{M}}\right) \;=\; \text{Tr}(\Sigma\mathcal{Y}) \;+\; \text{Tr}(\Sigma)\cdot O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \qquad (11.76)$$

Substituting (11.75) into (11.70) we obtain for $i \gg 1$:

$$\mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma} \;\leq\; \mathbb{E}\left(\|\boldsymbol{z}_{i-1}\|^2_{\mathcal{B}^*\Sigma\mathcal{B}}\right) \;+\; \text{Tr}(\Sigma\mathcal{Y}) \;+\; b_o \qquad (11.77)$$

$$\mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma} \;\geq\; \mathbb{E}\left(\|\boldsymbol{z}_{i-1}\|^2_{\mathcal{B}^*\Sigma\mathcal{B}}\right) \;+\; \text{Tr}(\Sigma\mathcal{Y}) \;-\; b_o \qquad (11.78)$$

Recall#3: Properties

$$\limsup_{i\to\infty} \,(a(i)+b(i)) \;\leq\; \limsup_{i\to\infty}\, a(i) \;+\; \limsup_{i\to\infty}\, b(i) \qquad (4.117)$$

$$\liminf_{i\to\infty} \,(a(i)+b(i)) \;\geq\; \liminf_{i\to\infty}\, a(i) \;+\; \liminf_{i\to\infty}\, b(i) \qquad (4.118)$$

Proof

Using the sub-additivity and super-additivity properties (4.117)–(4.118) of the limit superior and limit inferior operations, we conclude from the above relations that:

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma} \;\; \leq \;\; \limsup_{i\to\infty} \mathbb{E}\left(\|\boldsymbol{z}_{i-1}\|^2_{\mathcal{B}^*\Sigma\mathcal{B}}\right) \; + \; \mathrm{Tr}(\Sigma\mathcal{Y}) \; + \; b_o \qquad (11.79)$$

$$\liminf_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma} \;\; \geq \;\; \liminf_{i\to\infty} \mathbb{E}\left(\|\boldsymbol{z}_{i-1}\|^2_{\mathcal{B}^*\Sigma\mathcal{B}}\right) \; + \; \mathrm{Tr}(\Sigma\mathcal{Y}) \; - \; b_o \qquad (11.80)$$

Grouping terms we get:

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma-\mathcal{B}^*\Sigma\mathcal{B}} \;\; \leq \;\; \mathrm{Tr}(\Sigma\mathcal{Y}) \; + \; b_o \qquad (11.81)$$

$$\liminf_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma-\mathcal{B}^*\Sigma\mathcal{B}} \;\; \geq \;\; \mathrm{Tr}(\Sigma\mathcal{Y}) \; - \; b_o \qquad (11.82)$$

Proof

and, consequently, by using the fact that the limit inferior of a sequence is upper bounded by its limit superior, we obtain the following inequality relation:

$$\mathrm{Tr}(\Sigma\mathcal{Y}) \; - \; b_o \; \leq \; \liminf_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma-\mathcal{B}^*\Sigma\mathcal{B}}$$

$$\leq \; \limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma-\mathcal{B}^*\Sigma\mathcal{B}} \; \leq \; \mathrm{Tr}(\Sigma\mathcal{Y}) + b_o \quad (11.83)$$

Recalling that $b_o = \mathrm{Tr}(\Sigma) \cdot O(\mu_{\max}^{2+(\gamma'/2)})$, we conclude that the limit superior and limit inferior of the error variance satisfy:

Proof

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma-\mathcal{B}^*\Sigma\mathcal{B}} \;=\; \text{Tr}(\Sigma\mathcal{Y}) \;+\; \text{Tr}(\Sigma)\cdot O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \qquad (11.84)$$

$$\liminf_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma-\mathcal{B}^*\Sigma\mathcal{B}} \;=\; \text{Tr}(\Sigma\mathcal{Y}) \;-\; \text{Tr}(\Sigma)\cdot O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \qquad (11.85)$$

We can now use (11.84) to justify (11.46). To do so, it is useful to review first two properties of block Kronecker products, which will be used in the derivation.

Recall #4: Properties

Table F.2: Properties of the block Kronecker product definition (F.2).

1.	$(\mathcal{A}+\mathcal{B}) \otimes_b \mathcal{C} = (\mathcal{A} \otimes_b \mathcal{C}) \; + \; (\mathcal{B} \otimes_b \mathcal{C})$
2.	$(\mathcal{A} \otimes_b \mathcal{B})(\mathcal{C} \otimes_b \mathcal{D}) = (\mathcal{A}\mathcal{C} \otimes_b \mathcal{B}\mathcal{D})$
3.	$(A \otimes B) \otimes_b (C \otimes D) = (A \otimes C) \otimes (B \otimes D)$
4.	$(\mathcal{A} \otimes_b \mathcal{B})^{\mathsf{T}} = \mathcal{A}^{\mathsf{T}} \otimes_b \mathcal{B}^{\mathsf{T}}$
5.	$(\mathcal{A} \otimes_b \mathcal{B})^{*} = \mathcal{A}^{*} \otimes_b \mathcal{B}^{*}$
6.	$\{\lambda(\mathcal{A} \otimes_b \mathcal{B})\} = \{\lambda_i(\mathcal{A})\lambda_j(\mathcal{B})\}_{i=1,j=1}^{np,mp}$
7.	$\mathrm{Tr}(\mathcal{A}\mathcal{B}) = \left[\mathrm{bvec}(\mathcal{B}^{\mathsf{T}})\right]^{\mathsf{T}} \mathrm{bvec}(\mathcal{A}) = \left[\mathrm{bvec}(\mathcal{B}^{*})\right]^{*} \mathrm{bvec}(\mathcal{A})$
8.	$\mathrm{bvec}(\mathcal{A}\mathcal{C}\mathcal{B}) = (\mathcal{B}^{\mathsf{T}} \otimes_b \mathcal{A})\mathrm{bvec}(\mathcal{C})$
9.	$\mathrm{bvec}(xy^{\mathsf{T}}) = y \otimes_b x$

Recall#5: Lyapunov Equation

Given $N \times N$ matrices X, A, and Q, where Q is Hermitian and non-negative definite, we consider first discrete-time Lyapunov equations, also called Stein equations, of the following form:

$$X - A^* X A = Q \tag{F.38}$$

Let $\lambda_k(A)$ denote any of the eigenvalues of A. In the discrete-time case, a stable matrix A is one whose eigenvalues lie inside the unit disc (i.e., their magnitudes are strictly less than one).

Recall#5: Lyapunov Equation

Lemma F.2 (Discrete-time Lyapunov equation). Consider the Lyapunov equation (F.38). The following facts hold:

(a) The solution X is unique if, and only if, $\lambda_k(A)\lambda_\ell^*(A) \neq 1$ for all $k, \ell = 1, 2, \ldots, N$. In this case, the unique solution X is Hermitian.

(b) When A is stable (i.e., all its eigenvalues are inside the unit disc), the solution X is unique, Hermitian, and nonnegative-definite. Moreover, it admits the series representation:

$$X = \sum_{n=0}^{\infty} (A^*)^n Q A^n \tag{F.39}$$

Proof

Returning to (11.84), we recall that we are free to choose the weighting matrix Σ. Assume we select Σ as the solution to the following (discrete-time) Lyapunov equation:

$$\Sigma - \mathcal{B}^* \Sigma \mathcal{B} \;=\; I_{hMN} \tag{11.88}$$

We know from (9.173) that the matrix $\mathcal{B}$ is stable for sufficiently small step-sizes. Accordingly, we are guaranteed from the statement of Lemma F.2 that the above Lyapunov equation has a unique solution Σ and, moreover, this solution is Hermitian and non-negative definite, as desired. The advantage of this choice for Σ is that it reduces the weighting matrix on the mean-square value of $\boldsymbol{z}_i$ in (11.84) to the identity matrix. We can then focus on evaluating the value of the right-hand side of expression (11.84).

Proof

For this purpose, we start by applying the block vectorization operation to both sides of (11.88) and use (11.86) to find that

$$\text{bvec}(\Sigma) - \left(\mathcal{B}^{\mathsf{T}} \otimes_b \mathcal{B}^*\right) \text{bvec}(\Sigma) \;=\; \text{bvec}(I_{hMN}) \qquad (11.89)$$

so that in terms of the matrix $\mathcal{F}$ defined by (11.56), which is also stable, we can write

$$\text{bvec}(\Sigma) \;=\; (I - \mathcal{F})^{-1}\text{bvec}(I_{hMN}) \qquad (11.90)$$

Now, substituting this Σ into (11.84), we obtain $\mathbb{E}\,\|\boldsymbol{z}_i\|^2$ on the left-hand side while the term $\text{Tr}(\Sigma\mathcal{Y})$ on the right-hand side becomes:

Proof

$$
\begin{aligned}
\text{Tr}(\Sigma\mathcal{Y}) &\overset{(11.87)}{=} \left(\text{bvec}(\mathcal{Y}^{\mathsf{T}})\right)^{\mathsf{T}} \text{bvec}(\Sigma) \\
&= \left(\text{bvec}(\mathcal{Y}^{\mathsf{T}})\right)^{\mathsf{T}} (I - \mathcal{F})^{-1}\text{bvec}(I_{hMN}) \qquad (11.91)
\end{aligned}
$$

Likewise, the second term on the right-hand side of (11.84) becomes:

$$
\begin{aligned}
O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \cdot \text{Tr}(\Sigma) &\overset{(11.87)}{=} O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \cdot (\text{bvec}(I_{hMN}))^{\mathsf{T}}\, \text{bvec}(\Sigma) \\
&= O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \cdot (\text{bvec}(I_{hMN}))^{\mathsf{T}}\, (I - \mathcal{F})^{-1}\text{bvec}(I_{hMN})
\end{aligned}
$$

(11.92)

Proof

But since $\mathcal{F}$ is a stable matrix, we can employ the expansion

$$
\begin{aligned}
(I-\mathcal{F})^{-1} &= I+\mathcal{F}+\mathcal{F}^2+\mathcal{F}^3+\ldots \\
&\overset{(11.56)}{=} I+\left(\mathcal{B}^{\mathsf{T}}\otimes_b\mathcal{B}^*\right)+\left(\left(\mathcal{B}^{\mathsf{T}}\right)^2\otimes_b(\mathcal{B}^*)^2\right)+\ldots \quad (11.93)
\end{aligned}
$$

and appeal to properties (11.86) and (11.87) again, to validate the identities:

$$
\left[\text{bvec}\left(\mathcal{Y}^{\mathsf{T}}\right)\right]^{\mathsf{T}}(I-\mathcal{F})^{-1}\text{bvec}\left(I_{hMN}\right) = \sum_{n=0}^{\infty}\text{Tr}\left[\mathcal{B}^n\mathcal{Y}(\mathcal{B}^*)^n\right] \quad (11.94)
$$

$$
\left(\text{bvec}(I_{hMN})\right)^{\mathsf{T}}(I-\mathcal{F})^{-1}\text{bvec}(I_{hMN}) = \sum_{n=0}^{\infty}\text{Tr}\left[(\mathcal{B}^*)^n\mathcal{B}^n\right] \quad (11.95)
$$

Proof

The two series that appear in the above expressions converge to the trace values of certain Lyapunov solutions. To see this, let

$$\mathcal{X}' \triangleq \sum_{n=0}^{\infty} (\mathcal{B}^*)^n \mathcal{B}^n, \qquad \mathcal{X} \triangleq \sum_{n=0}^{\infty} \mathcal{B}^n \mathcal{Y} (\mathcal{B}^*)^n \tag{11.96}$$

Then, these series correspond, respectively, to the unique solutions of the following Lyapunov equations (cf. Lemma F.2 from the appendix):

$$\mathcal{X}' - \mathcal{B}^* \mathcal{X}' \mathcal{B} = I_{hMN} \tag{11.97}$$

$$\mathcal{X} - \mathcal{B} \mathcal{X} \mathcal{B}^* = \mathcal{Y} \tag{11.98}$$

Proof

Moreover, the matrices $\mathcal{X}$ and $\mathcal{X}'$ so defined are Hermitian and nonnegative-definite (note for $\mathcal{X}$ that the matrix $\mathcal{Y}$ defined by (11.54) is Hermitian and non-negative definite). Therefore, we have established so far that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2 = \text{Tr}(\mathcal{X}) \;+\; \text{Tr}(\mathcal{X}')\cdot O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \tag{11.99}$$

We now verify that $\text{Tr}(\mathcal{X}') = O(1/\mu_{\max})$ — see (11.103); this result will permit us to assess the size of the second term on the right-hand side of (11.99) — see (11.104).

Recall#6: Size of Entries

Lemma 9.5 Assume the matrix P is primitive. For sufficiently small step-sizes, it holds that

$$(I-\mathcal{F})^{-1} = O(1/\mu_{\max}) \tag{9.242}$$

$$(I-\bar{\mathcal{F}})^{-1} = \left[\begin{array}{c|c} O(1/\mu_{\max}) & O(1) \\ \hline O(1) & O(1) \end{array}\right] \tag{9.243}$$

where the leading $(hM)^2 \times (hM)^2$ block in $(I-\bar{\mathcal{F}})^{-1}$ is $O(1/\mu_{\max})$.

Proof

Applying the bvec operation to both sides of (11.97) and using (11.86) we find that

$$\text{bvec}(\mathcal{X}') = (I - \mathcal{F})^{-1}\text{bvec}(I) \tag{11.100}$$

Then,

$$\begin{aligned} \|\text{bvec}(\mathcal{X}')\| &\leq \left\|(I - \mathcal{F})^{-1}\right\| \, \|\text{bvec}(I)\| \\ &\overset{\text{(a)}}{\leq} r \cdot \left\|(I - \mathcal{F})^{-1}\right\|_1 \, \|\text{bvec}(I)\| \\ &\overset{(9.243)}{=} O(1/\mu_{\max}) \end{aligned} \tag{11.101}$$

where in step (a) we used a positive constant r to account for the fact that matrix norms are equivalent (cf. (F.6) in the appendix). We can use this result to bound the trace of $\mathcal{X}'$ as follows.

Proof

Let $L \times L$ denote the dimensions of $\mathcal{X}'$; we know that $L = hNM$. Let further $\{x'_{nn},\ n = 1, 2, \ldots, L\}$ denote the diagonal entries of $\mathcal{X}'$. Since $\mathcal{X}' \geq 0$, we know that $x'_{nn} \geq 0$. We collect the diagonal entries of $\mathcal{X}'$ into the column vector $b = \text{col}\{x'_{nn}\}$. Then, for any two vectors a and b of compatible dimensions, we use the Cauchy-Schwartz inequality $(a^* b)^2 \leq \|a\|^2 \|b\|^2$ to conclude that

Proof

$$
\begin{aligned}
\left(\mathrm{Tr}(\mathcal{X}')\right)^2 &\stackrel{\Delta}{=} \left(\sum_{n=1}^{L} x'_{nn}\right)^2 \\
&= \left(\mathbb{1}^{\mathsf{T}} b\right)^2 \\
&\leq \|\mathbb{1}\|^2 \, \|b\|^2 \\
&= L \cdot \|b\|^2 \\
&\leq L \cdot \|\mathrm{bvec}(\mathcal{X}')\|^2 \\
&\stackrel{(11.101)}{=} O(1/\mu_{\max}^2) \qquad (11.102)
\end{aligned}
$$

and, therefore,

$$
\mathrm{Tr}(\mathcal{X}') = O\left(1/\mu_{\max}\right) \qquad (11.103)
$$

Proof

$\gamma' = \min\{2, \gamma\}$

It follows that

$$O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \cdot \text{Tr}(\mathcal{X}') \;=\; O\left(\mu_{\max}^{1+(\gamma'/2)}\right) \tag{11.104}$$

Returning to (11.99), we conclude that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2 \;=\; \text{Tr}(\mathcal{X}) \,+\, O\left(\mu_{\max}^{1+(\gamma'/2)}\right) \tag{11.105}$$

and, consequently, using (11.68), we obtain the following two equivalent characterizations for the network MSD:

Proof

$$O\left(\mu_{\max}^{3/2}\right) + O\left(\mu_{\max}^{1+\frac{1}{2}\min\{2,\gamma\}}\right)$$

$$\gamma_m \stackrel{\Delta}{=} \frac{1}{2}\min\{1,\gamma\}$$

$$\limsup_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^2 = \frac{1}{2N}\mathrm{Tr}(\mathcal{X}) + O\left(\mu_{\max}^{1+\gamma_m}\right) \tag{11.106}$$

$$= \frac{1}{2N}\sum_{n=0}^{\infty}\mathrm{Tr}\left[\mathcal{B}^n\mathcal{Y}(\mathcal{B}^*)^n\right] + O\left(\mu_{\max}^{1+\gamma_m}\right) \tag{11.107}$$

with γ_m replacing $\gamma'/2$. These results, along with the arguments leading to them, justify expressions (11.46) and (11.58)–(11.60). Observe in particular from (11.54) and (9.243) that the term on the left-hand side of (11.94) is $O(\mu_{\max})$ since $\|\mathcal{Y}\| = O(\mu_{\max}^2)$ and $\|(I-\mathcal{F})^{-1}\| = O(1/\mu_{\max})$. Therefore, the value of $\mathrm{Tr}(\mathcal{X})$ in (11.60) is $O(\mu_{\max})$, which dominates the factor $O(\mu_{\max}^{1+\gamma_m})$.

Proof

Similarly, if we start from (11.85) instead, and apply the same arguments we would arrive at the following equivalent expressions:

$$\liminf_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^2 = \frac{1}{2N}\text{Tr}(\mathcal{X}) \;-\; O\left(\mu_{\max}^{1+\gamma_m}\right) \tag{11.108}$$

$$= \frac{1}{2N}\sum_{n=0}^{\infty}\text{Tr}\left[\mathcal{B}^n\mathcal{Y}(\mathcal{B}^*)^n\right] - O\left(\mu_{\max}^{1+\gamma_m}\right) \tag{11.109}$$

This last result is not needed in the current derivation but is referred to later in Example 11.7.

Proof

We can also assess the mean-square performance of the *individual* agents in the network from (11.77). Let us introduce the $N \times N$ block diagonal matrix $\mathcal{J}_k$ defined by (11.57) with blocks of size $hM \times hM$, where all blocks on the diagonal are zero except for an identity matrix on the diagonal block of index k. Then, the error variance for agent k satisfies:

$$\limsup_{i\to\infty} \frac{1}{2}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^2_{\mathcal{J}_k} \;=\; \limsup_{i\to\infty} \frac{1}{2}\mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\mathcal{J}_k} \;+\; O(\mu_{\max}^{3/2}) \qquad (11.110)$$

The same argument that was used to obtain expression (11.46) for the network mean-square-error can then be repeated to give (11.45) and (11.61).

Proof

With regards to the convergence rate of $\mathbb{E}\|\widetilde{\boldsymbol{w}}_{k,i}\|^2$ towards the region (11.45), we substitute (11.76) into (11.70) to write for $i \gg 1$:

$$\mathbb{E}\,\|\boldsymbol{z}_i\|^2_{\Sigma} \;=\; \mathbb{E}\,\left(\|\boldsymbol{z}_{i-1}\|^2_{\mathcal{B}^*\Sigma\mathcal{B}}\right) + \mathrm{Tr}(\Sigma\mathcal{Y}) + \mathrm{Tr}(\Sigma)\cdot O\left(\mu_{\max}^{2+(\gamma'/2)}\right) \qquad (11.111)$$

Selecting origin of time at some large time and iterating from there:

$$\mathbb{E}\,\|\boldsymbol{z}_i\|^2 \;=\; \mathbb{E}\,\|\boldsymbol{z}_{-1}\|^2_{(\mathcal{B}^*)^{i+1}\mathcal{B}^{i+1}} \;+\; \sum_{n=0}^{i}\mathrm{Tr}\left[\mathcal{B}^n\mathcal{Y}(\mathcal{B}^*)^n\right] \;+\; o(\mu^2) \qquad (11.112)$$

Proof

The first-term on the right-hand side corresponds to a transient component that dies out with time. The rate of its convergence towards zero determines the rate of convergence of $\mathbb{E}\,\|\boldsymbol{z}_i\|^2$ towards its steady-state region. This rate can be characterized as follows. Note that, using properties (11.86)–(11.87) for block Kronecker products, we can express the weighted variance of $\boldsymbol{z}_{-1}$ as the following trace relation in terms of its un-weighted covariance matrix, which we denote by $R_z = \mathbb{E}\,\boldsymbol{z}_{-1}\boldsymbol{z}_{-1}^*$:

Proof

$$\begin{aligned}
\mathbb{E}\,\|\boldsymbol{z}_{-1}\|^2_{(\mathcal{B}^*)^{i+1}\mathcal{B}^{i+1}} &= \mathbb{E}\left(\boldsymbol{z}^*_{-1}\,(\mathcal{B}^*)^{i+1}\,\mathcal{B}^{i+1}\boldsymbol{z}_{-1}\right) \\
&= \mathrm{Tr}\left((\mathcal{B}^*)^{i+1}\,\mathcal{B}^{i+1}R_z\right) \\
&\overset{(11.87)}{=} \left[\mathrm{bvec}\left(R_z^{\mathsf{T}}\right)\right]^{\mathsf{T}}\mathrm{bvec}\left((\mathcal{B}^*)^{i+1}\,\mathcal{B}^{i+1}\right) \\
&\overset{(11.86)}{=} \left[\mathrm{bvec}(R_z^{\mathsf{T}})\right]^{\mathsf{T}}\left((\mathcal{B}^{\mathsf{T}})^{i+1}\otimes_b(\mathcal{B}^*)^{i+1}\right)\mathrm{bvec}(I)
\end{aligned} \tag{11.113}$$

Proof

It is clear now that the convergence rate of the transient component is dictated by the spectral radius of the matrix multiplying bvec(I), namely, by

$$\rho\left(\left(\mathcal{B}^{\mathsf{T}}\right)^{i+1} \otimes_b \left(\mathcal{B}^{*}\right)^{i+1}\right) = \left(\left[\rho(\mathcal{B})\right]^2\right)^{i+1} \tag{11.114}$$

We conclude that the convergence rate of $\mathbb{E}\,\|\boldsymbol{z}_i\|^2$ towards the steady-state regime is dictated by $[\rho(\mathcal{B})]^2$ since this value characterizes the slowest rate at which the transient term dies out. Therefore, using (9.173) and the relation $(1-x)^2 = 1 - 2x + O(x^2)$, we can approximate the convergence rate to first-order in μ as follows:

Proof

$$
\begin{aligned}
[\rho(\mathcal{B})]^2 &= \left[1 - \lambda_{\min}\left(\sum_{k=1}^{N} q_k H_k\right) + O\left(\mu_{\max}^{(N+1)/N}\right)\right]^2 \\
&= 1 - 2\lambda_{\min}\left(\sum_{k=1}^{N} q_k H_k\right) + O\left(\mu_{\max}^{(N+1)/N}\right) \qquad (11.115)
\end{aligned}
$$

$\square$

Example #11.1

Example 11.1 (Steady-state region for MSE networks). Let us consider the case of MSE networks, defined earlier in Example 6.3, where the data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ satisfy the linear regression model (6.14) and where the cost function associated with each agent is the mean-square-error cost, $J_k(w) = \mathbb{E}\,|\boldsymbol{d}_k(i) - \boldsymbol{u}_{k,i} w|^2$.

Example #11.1

We showed in Example 6.1 that in this case, all individual costs are minimized at the same location w^o. It follows that the reference vectors w^o and $w^\star$ will coincide and, therefore, the bias vector b^e that appears in the error recursion (10.2) will be zero (as is evident from the definition of its entries in (8.136)). Moreover, the matrices $\boldsymbol{H}_{k,i-1}$ and H_k defined by (10.6) and (10.9), respectively, will coincide with each other since the Hessian matrix $\nabla_w^2 J_k(w)$ will be constant for all w. Thus, in this case, we get:

$$\boldsymbol{H}_{k,i-1} \equiv H_k = \nabla_w^2 J_k(w^o) \qquad (11.116)$$

Example #11.1

We showed in Example 6.1 that in this case, all individual costs are minimized at the same location w^o. It follows that the reference vectors w^o and $w^\star$ will coincide and, therefore, the bias vector b^e that appears in the error recursion (10.2) will be zero (as is evident from the definition of its entries in (8.136)). Moreover, the matrices $\boldsymbol{H}_{k,i-1}$ and H_k defined by (10.6) and (10.9), respectively, will coincide with each other since the Hessian matrix $\nabla_w^2 J_k(w)$ will be constant for all w. Thus, in this case, we get:

$$\boldsymbol{H}_{k,i-1} \equiv H_k = \nabla_w^2 J_k(w^o) \tag{11.116}$$

Example #11.1

As a result, the perturbation term $\boldsymbol{c}_{i-1}$ in (10.13) will be identically zero and recursions (10.13) and (10.19) will therefore coincide (including having $b^e = 0$). Both models (i.e., the actual error recursion and the long-term error recursion) will then have the same MSD expressions. Therefore, we can rely on expression (11.68) without the need for the additional error factor $O(\mu_{\max}^{3/2})$. We know from the earlier result (4.16) that $\gamma = 2$ for mean-square-error costs. Using this value for γ in the derivation leading to (11.107), and ignoring the correction by $O(\mu_{\max}^{3/2})$, we arrive instead at

$$\limsup_{i\to\infty} \frac{1}{2N}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^2 \;=\; \frac{1}{2N}\sum_{n=0}^{\infty} \mathrm{Tr}\left[\mathcal{B}^n \mathcal{Y} (\mathcal{B}^*)^n\right] \;+\; O\left(\mu_{\max}^2\right) \qquad (11.117)$$

Example #11.1

with an approximation error in the order of $O(\mu_{\max}^2)$ rather than the term $O(\mu_{\max}^{3/2})$ that would result from (11.107) if we use $\gamma_m = 1/2$. We conclude that for MSE networks, the results of Theorem 11.2 are valid with the approximation error $O(\mu_{\max}^{1+\gamma_m})$ in (11.45)–(11.46) replaced by the smaller factor $O(\mu_{\max}^2)$.

■

End of Lecture