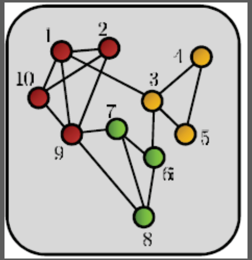
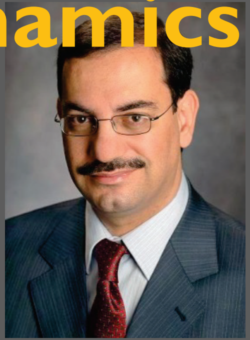


INFERENCE OVER NETWORKS

LECTURE #19: Long-Term Network Dynamics

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

Chapter 10 (Long-Term Network Dynamics, pp. 552-573):

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 move on to motivate a long-term model for the evolution of the network error dynamics, $\widetilde{\boldsymbol{w}}_i^e$, after sufficient iterations have passed, i.e., for $i \gg 1$. We examine the stability property of the model, the proximity of its trajectory from that of the original network dynamics, and subsequently employ the model to assess network MSD and ER performance metrics. To do so, we will need to recall the same smoothness condition used in establishing the mean-stability result of Theorem 9.6.

Setting

Assumption 10.1. (Smoothness condition on individual cost functions). It is assumed that each $J_k(w)$ satisfies a smoothness condition close to the limit point $w^\star$, defined by (8.55), in that the corresponding Hessian matrix is Lipschitz continuous in the proximity of $w^\star$ with some parameter $\kappa_d \geq 0$, i.e.,

$$\left\| \nabla_w^2 J_k(w^\star + \Delta w) - \nabla_w^2 J_k(w^\star) \right\| \;\leq\; \kappa_d \, \|\Delta w\| \tag{10.1}$$

for small perturbations $\|\Delta w\| \leq \epsilon$.

Long-Term Model

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.

Long-Term Error Model

We reconsider the network error recursion (9.12), namely,

$$\widetilde{\boldsymbol{w}}_i^e = \boldsymbol{\mathcal{B}}_{i-1}\widetilde{\boldsymbol{w}}_{i-1}^e + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e, \quad i \geq 0 \tag{10.2}$$

where, from the expressions in Lemma 8.1,

$$\boldsymbol{\mathcal{B}}_{i-1} = \mathcal{P}^{\mathsf{T}} - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{\mathcal{H}}_{i-1}\mathcal{A}_1^{\mathsf{T}} \tag{10.3}$$

$$\mathcal{P}^{\mathsf{T}} = \mathcal{A}_2^{\mathsf{T}}\mathcal{A}_o^{\mathsf{T}}\mathcal{A}_1^{\mathsf{T}} \tag{10.4}$$

$$\boldsymbol{\mathcal{H}}_{i-1} \triangleq \text{diag}\{\, \boldsymbol{H}_{1,i-1},\ \boldsymbol{H}_{2,i-1},\ \ldots,\ \boldsymbol{H}_{N,i-1}\,\} \tag{10.5}$$

$$\boldsymbol{H}_{k,i-1} \triangleq \int_0^1 \nabla_w^2 J_k(w^\star - t\widetilde{\boldsymbol{\phi}}_{k,i-1})dt \tag{10.6}$$

Long-Term Error Model

We again introduce the error matrix:

$$\widetilde{\boldsymbol{\mathcal{H}}}_{i-1} \;\stackrel{\Delta}{=}\; \mathcal{H} \;-\; \boldsymbol{\mathcal{H}}_{i-1} \tag{10.7}$$

which measures the deviation of $\boldsymbol{\mathcal{H}}_{i-1}$ from the constant matrix:

$$\mathcal{H} \;\stackrel{\Delta}{=}\; \text{diag}\{\, H_1,\; H_2,\; \ldots,\; H_N \,\} \tag{10.8}$$

with each H_k given by the value of the Hessian matrix at the limit point, namely,

$$H_k \;\stackrel{\Delta}{=}\; \nabla_w^2\, J_k(w^\star) \tag{10.9}$$

Long-Term Error Model

Then, using (9.166) in the expression for $\boldsymbol{\mathcal{B}}_{i-1}$, we can write

$$\boldsymbol{\mathcal{B}}_{i-1} \;=\; \mathcal{B} + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\widetilde{\boldsymbol{\mathcal{H}}}_{i-1}\mathcal{A}_1^{\mathsf{T}} \tag{10.10}$$

in terms of the constant coefficient matrix

$$\mathcal{B} \stackrel{\Delta}{=} \mathcal{P}^{\mathsf{T}} - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\mathcal{H}\mathcal{A}_1^{\mathsf{T}} \tag{10.11}$$

We established in Theorem 9.3 that, for sufficiently small step-sizes, the matrix $\mathcal{B}$ is stable and its spectral radius is given by

$$\rho(\mathcal{B}) \;=\; 1 - \lambda_{\min}\left(\sum_{k=1}^{N} q_k H_k\right) + O\left(\mu_{\max}^{(N+1)/N}\right) \tag{10.12}$$

Long-Term Error Model

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of its Hermitian matrix argument. Now, using (10.10), we can rewrite error recursion (10.2) as

$$\widetilde{\boldsymbol{w}}_i^e = \mathcal{B}\,\widetilde{\boldsymbol{w}}_{i-1}^e + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{c}_{i-1} \quad (10.13)$$

in terms of the random perturbation sequence:

$$\boldsymbol{c}_{i-1} \triangleq \widetilde{\boldsymbol{\mathcal{H}}}_{i-1}\mathcal{A}_1^{\mathsf{T}}\widetilde{\boldsymbol{w}}_{i-1}^e \quad (10.14)$$

Long-Term Error Model

By exploiting the smoothness condition (10.1), and following an argument similar to (9.280)–(9.283), we can verify that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{c}_{i-1}\| \;=\; O(\mu_{\max}) \qquad (10.15)$$

This is because

$$\begin{aligned}
\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{c}_{i-1}\| &\overset{(10.14)}{\leq} \|\mathcal{A}_1\| \left(\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{\mathcal{H}}}_{i-1}\|\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\| \right) \\
&\overset{(9.281)}{\leq} \frac{1}{2}\kappa'_d N \|\mathcal{A}_1\| \left(\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|^2 \right) \\
&\overset{(9.11)}{=} O(\mu_{\max}) \qquad (10.16)
\end{aligned}$$

Long-Term Error Model

Returning to (10.15), we deduce that $\|\boldsymbol{c}_{i-1}\| = O(\mu_{\max})$ asymptotically with *high probability* using the same argument that led to (4.53) in the single-agent case. Referring to recursion (10.13), this analysis suggests that we can assess the mean-square performance of the original error recursion (10.2) by considering instead the following long-term model, which holds with high probability after sufficient iterations:

$$\widetilde{\boldsymbol{w}}_i^e \;=\; \mathcal{B}\,\widetilde{\boldsymbol{w}}_{i-1}^e \;+\; \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) \;-\; \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e, \quad i \gg 1 \qquad (10.17)$$

Long-Term Error Model

In this model, the perturbation term $\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{c}_{i-1}$ that appears in (10.13) is removed. We may also consider an alternative long-term model where $\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{c}_{i-1}$ is instead replaced by a constant driving term in the order of $O(\mu_{\max}^2)$. However, the conclusions that will follow about the performance of the original recursion (10.2) will be the same whether we remove $\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{c}_{i-1}$ altogether or replace it by $O(\mu_{\max}^2)$. We therefore

Long-Term Error Model

continue our analysis by using model (10.17). Obviously, the iterates $\{\widetilde{\boldsymbol{w}}_i^e\}$ that are generated by (10.17) are generally different from the iterates that are generated by the original recursion (10.2). To highlight this fact more accurately, we rewrite the long-term recursion (10.17) more explicitly as follows for $i \gg 1$:

$$\widetilde{\boldsymbol{w}}_i^{e'} = \mathcal{B}\,\widetilde{\boldsymbol{w}}_{i-1}^{e'} + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \qquad (10.18)$$

Long-Term Error Model

with the iterates now denoted by $\widetilde{\boldsymbol{w}}_i^{e'}$ using the prime notation for the state of the long-term model. Note that the driving process $\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)$ in (10.18) is the *same* gradient noise process from the original recursion (10.2) and is therefore evaluated at $\boldsymbol{w}_{i-1}^e$. It is instructive to compare the following statement with the earlier Lemma 8.1.

Long-Term Error Model

Lemma 10.1 (Long-term network dynamics). 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. After sufficient iterations, $i \gg 1$, the error dynamics of the network relative to the limit point $w^\star$ defined by (8.55) is well-approximated by the following model (as confirmed by future result (10.29)):

$$\widetilde{\boldsymbol{w}}_i^{e'} = \mathcal{B}\, \widetilde{\boldsymbol{w}}_{i-1}^{e'} + \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}} \mathcal{M} b^e \qquad (10.19)$$

where

Long-Term Error Model

$$\mathcal{B} \triangleq \mathcal{A}_2^{\mathsf{T}} \left(\mathcal{A}_o^{\mathsf{T}} - \mathcal{M}\mathcal{H}\right) \mathcal{A}_1^{\mathsf{T}} \tag{10.20}$$

$$\mathcal{A}_o \triangleq A_o \otimes I_{2M}, \quad \mathcal{A}_1 \triangleq A_1 \otimes I_{2M}, \quad \mathcal{A}_2 \triangleq A_2 \otimes I_{2M} \tag{10.21}$$

$$\mathcal{M} \triangleq \text{diag}\{\, \mu_1 I_{2M},\ \mu_2 I_{2M},\ \ldots,\ \mu_N I_{2M} \,\} \tag{10.22}$$

$$\mathcal{H} \triangleq \text{diag}\{\, H_1,\ H_2,\ \ldots,\ H_N \,\} \tag{10.23}$$

$$H_k \triangleq \nabla_w^2\, J_k(w^\star) \tag{10.24}$$

where $\nabla_w^2 J_k(w)$ denotes the $2M \times 2M$ Hessian matrix of $J_k(w)$ relative to w.

Long-Term Error Model

In a manner similar to the partitioning of $\widetilde{\boldsymbol{w}}_i^e$ into its constituent elements in (8.143), we partition $\widetilde{\boldsymbol{w}}_i^{e'}$ into its $2M \times 1$ block entries as follows:

$$\widetilde{\boldsymbol{w}}_i^{e'} \triangleq \begin{bmatrix} \widetilde{\boldsymbol{w}}_{1,i}^{e'} \\ \widetilde{\boldsymbol{w}}_{2,i}^{e'} \\ \vdots \\ \widetilde{\boldsymbol{w}}_{N,i}^{e'} \end{bmatrix} \tag{10.25}$$

with each $\widetilde{\boldsymbol{w}}_{k,i}^{e'}$ at every agent in turn consisting of

$$\widetilde{\boldsymbol{w}}_{k,i}^{e'} = \begin{bmatrix} \widetilde{\boldsymbol{w}}_{k,i}^{'} \\ \left(\widetilde{\boldsymbol{w}}_{k,i}^{'*}\right)^{\mathsf{T}} \end{bmatrix} \tag{10.26}$$

Long-Term Error Model

We can view the long-term model (10.19) as a dynamic recursion that is fed by the gradient noise sequence, $\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)$. Therefore, assuming both the original system (10.2) and the long-term model (10.19) are launched from the same initial conditions, we observe by iterating (10.19) that $\widetilde{\boldsymbol{w}}_i^{e'}$ will still be determined by the past history of the original iterates $\{\boldsymbol{w}_j,\ j \leq i-1\}$ through its dependence on the gradient noise process $\{\boldsymbol{s}_j^e(\boldsymbol{w}_{j-1}^e),\ j \leq i\}$. Therefore, it continues to hold that the error vectors $\widetilde{\boldsymbol{w}}_{k,i}^{'}$ belong to the filtration $\boldsymbol{\mathcal{F}}_{i-1}$ that is determined by the history of all iterates $\{\boldsymbol{w}_{k,j},\ j \leq i-1,\ k = 1, 2, \ldots, N\}$ that are generated by the original distributed strategy (8.46).

Long-Term Error Model

Working with recursion (10.19) is much more tractable for performance analysis because its dynamics is driven by the constant matrix $\mathcal{B}$ as opposed to the random matrix $\boldsymbol{\mathcal{B}}_{i-1}$ in the original error recursion (10.2). We shall therefore follow the following route to evaluate the MSD of the stochastic-gradient distributed algorithm (8.46). We shall work with the long-term model (10.19) and evaluate its MSD.

Long-Term Error Model

Subsequently, we will argue that, under a bounding condition on the fourth-order moment of the gradient noise process, namely, condition (8.121), this MSD is within $O(\mu_{\max}^{3/2})$ from the true MSD expression that would have resulted had we worked directly with the original error recursion (10.2) without the approximation of ignoring $\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{c}_{i-1}$. This fact will then allow us to conclude that the MSD expression that is derived from the long-term model (10.19) provides an accurate representation for the MSD of the original stochastic-gradient distributed strategy (8.46) to first-order in $\mu_{\max}$.

Approximation Error

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.

Size of Approximation Error

We first examine how close the trajectories of the original error recursion (10.2) and the long-term model (10.19) are to each other. We reproduce both recursions below with the state variable for the long-term model denoted by $\widetilde{\boldsymbol{w}}_i^{e'}$:

$$\begin{aligned}
\widetilde{\boldsymbol{w}}_i^e &= \boldsymbol{\mathcal{B}}_{i-1}\widetilde{\boldsymbol{w}}_{i-1}^e + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \qquad (10.27)\\
\widetilde{\boldsymbol{w}}_i^{e'} &= \mathcal{B}\,\widetilde{\boldsymbol{w}}_{i-1}^{e'} + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \qquad (10.28)
\end{aligned}$$

Size of Approximation Error

Observe that both models are driven by the *same* gradient noise process; in this way, the evolution of the long-term model is coupled to the evolution of the original recursion (but not the other way around). The next result establishes that the mean-square difference between the trajectories $\{\widetilde{\boldsymbol{w}}_i^e, \widetilde{\boldsymbol{w}}_i^{e'}\}$ is asymptotically bounded by $O(\mu_{\max}^2)$.

Size of Approximation Error

Theorem 10.2 (Performance error is $O(\mu_{\max}^{3/2})$). 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). Then, it holds that, for sufficiently small step-sizes:

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e - \widetilde{\boldsymbol{w}}_i^{e'}\|^2 \;=\; O(\mu_{\max}^2) \tag{10.29}$$

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^2 \;=\; \limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'}\|^2 + O(\mu_{\max}^{3/2}) \tag{10.30}$$

Proof

Proof. To simplify the notation, we introduce the difference

$$\boldsymbol{z}_i \triangleq \widetilde{\boldsymbol{w}}_i^e - \widetilde{\boldsymbol{w}}_i^{e'} \qquad (10.31)$$

Using (10.10) and (10.14), and subtracting recursions (10.27) and (10.28) we then get

$$\boldsymbol{z}_i = \mathcal{B}\boldsymbol{z}_{i-1} + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{c}_{i-1} \qquad (10.32)$$

We also know from (9.173) that the matrix $\mathcal{B}$ is stable for sufficiently small step-sizes and, moreover, for $\mu_{\max} \ll 1$, it holds from (9.222) and (9.226) that

$$\rho(\mathcal{B}) = 1 - O(\mu_{\max}) \; = \; 1 - \sigma_b\mu_{\max} \qquad (10.33)$$

for some positive constant σ_b that is independent of $\mu_{\max}$.

Proof

We introduce the same Jordan canonical decomposition (9.24) for the matrix P, namely,

$$P \stackrel{\Delta}{=} V_\epsilon J V_\epsilon^{-1} \tag{9.189}$$

$$J = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & J_\epsilon \end{array}\right] \tag{9.190}$$

where the matrix J_ϵ consists of Jordan blocks of forms similar to (9.25) with $\epsilon > 0$ appearing on the lower diagonal. The value of ϵ can be chosen to be arbitrarily small and is independent of $\mu_{\max}$. The Jordan decomposition of the extended matrix $\mathcal{P} = P \otimes I_{2M}$ is given by

Proof

$$\mathcal{P} \;=\; (V_\epsilon \otimes I_{2M})(J \otimes I_{2M})(V_\epsilon^{-1} \otimes I_{2M}) \tag{9.191}$$

so that substituting into (9.170) we obtain

$$\mathcal{B} = \left((V_\epsilon^{-1})^{\mathsf{T}} \otimes I_{2M}\right)\left\{(J^{\mathsf{T}} \otimes I_{2M}) - \mathcal{D}^{\mathsf{T}}\right\}\left(V_\epsilon^{\mathsf{T}} \otimes I_{2M}\right) \tag{9.192}$$

where

$$\begin{aligned} \mathcal{D}^{\mathsf{T}} \;&\stackrel{\Delta}{=}\; \left(V_\epsilon^{\mathsf{T}} \otimes I_{2M}\right)\mathcal{A}_2^{\mathsf{T}}\mathcal{M}\mathcal{H}\mathcal{A}_1^{\mathsf{T}}\left((V_\epsilon^{-1})^{\mathsf{T}} \otimes I_{2M}\right) \\ &\equiv\; \begin{bmatrix} D_{11}^{\mathsf{T}} & D_{21}^{\mathsf{T}} \\ D_{12}^{\mathsf{T}} & D_{22}^{\mathsf{T}} \end{bmatrix} \end{aligned} \tag{9.193}$$

Proof

Using the partitioning (9.23)–(9.24) and the fact that

$$\mathcal{A}_1 = A_1 \otimes I_{2M}, \qquad \mathcal{A}_2 = A_2 \otimes I_{2M} \tag{9.194}$$

we find that the block entries $\{D_{mn}\}$ in (9.193) are given by

$$D_{11} = \sum_{k=1}^{N} q_k H_k^{\mathsf{T}} \tag{9.195}$$

$$D_{12} = (\mathbb{1}^{\mathsf{T}} \otimes I_{2M})\mathcal{H}^{\mathsf{T}}\mathcal{M}(A_2 V_R \otimes I_{2M}) \tag{9.196}$$

$$D_{21} = (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}^{\mathsf{T}}(q \otimes I_{2M}) \tag{9.197}$$

$$D_{22} = (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}^{\mathsf{T}}\mathcal{M}(A_2 V_R \otimes I_{2M}) \tag{9.198}$$

Proof

In a manner similar to the arguments used in the proof of Theorem 9.1, we can verify that

$$\begin{aligned} D_{11} &= O(\mu_{\max}) && (9.199)\\ D_{12} &= O(\mu_{\max}) && (9.200)\\ D_{21} &= O(\mu_{\max}) && (9.201)\\ D_{22} &= O(\mu_{\max}) && (9.202)\\ \rho(I_{2M} - D_{11}^{\mathsf{T}}) &= 1 - \sigma_{11}\mu_{\max} \;=\; 1 - O(\mu_{\max}) && (9.203) \end{aligned}$$

where σ_{11} is a positive scalar independent of $\mu_{\max}$.

Proof

Let

$$\mathcal{V}_\epsilon \stackrel{\Delta}{=} V_\epsilon \otimes I_{2M}\,, \qquad \mathcal{J}_\epsilon \stackrel{\Delta}{=} J_\epsilon \otimes I_{2M} \tag{9.204}$$

Then, using (9.192), we can write

$$\mathcal{B} = \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix} \mathcal{V}_\epsilon^{\mathsf{T}} \tag{9.205}$$

so that

$$\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{B} \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} = \begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix} \tag{9.206}$$

which shows that the matrix $\mathcal{B}$ is similar to, and therefore has the same eigenvalues as, the block matrix on the right-hand side, written as

Proof

We multiply both sides of (10.32) from the left by $\mathcal{V}_\epsilon^\mathsf{T}$ and use (9.57) and (9.206) to get for $i \gg 1$:

$$\begin{bmatrix} \bar{z}_i \\ \check{z}_i \end{bmatrix} = \underbrace{\begin{bmatrix} I_{2M} - D_{11}^\mathsf{T} & -D_{21}^\mathsf{T} \\ -D_{12}^\mathsf{T} & \mathcal{J}_\epsilon^\mathsf{T} - D_{22}^\mathsf{T} \end{bmatrix}}_{\triangleq \bar{\mathcal{B}}} \begin{bmatrix} \bar{z}_{i-1} \\ \check{z}_{i-1} \end{bmatrix} + \mathcal{V}_\epsilon^\mathsf{T} \mathcal{A}_2^\mathsf{T} \mathcal{M} c_{i-1} \qquad (10.34)$$

where the matrix

$$\bar{\mathcal{B}} = \mathcal{V}_\epsilon^\mathsf{T} \mathcal{B} \left(\mathcal{V}_\epsilon^{-1}\right)^\mathsf{T} \qquad (10.35)$$

Proof

is similar to $\mathcal{B}$ and is therefore stable by Theorem 9.3. We partition the vectors $\mathcal{V}_\epsilon^\mathsf{T} \boldsymbol{z}_i$ and $\mathcal{V}_\epsilon^\mathsf{T} \mathcal{A}_2^\mathsf{T} \mathcal{M} \boldsymbol{c}_{i-1}$ in recursion (10.34) into

$$\mathcal{V}_\epsilon^\mathsf{T} \boldsymbol{z}_i \triangleq \begin{bmatrix} \bar{\boldsymbol{z}}_i \\ \check{\boldsymbol{z}}_i \end{bmatrix}, \qquad \mathcal{V}_\epsilon^\mathsf{T} \mathcal{A}_2^\mathsf{T} \mathcal{M} \boldsymbol{c}_{i-1} \triangleq \begin{bmatrix} \bar{\boldsymbol{c}}_{i-1} \\ \check{\boldsymbol{c}}_{i-1} \end{bmatrix} \tag{10.36}$$

with the leading vectors, $\{\bar{\boldsymbol{z}}_i, \bar{\boldsymbol{c}}_{i-1}\}$, having dimensions $hM \times 1$ each. It follows that

Proof

$$\begin{bmatrix} \bar{\boldsymbol{z}}_i \\ \check{\boldsymbol{z}}_i \end{bmatrix} = \begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{z}}_{i-1} \\ \check{\boldsymbol{z}}_{i-1} \end{bmatrix} + \begin{bmatrix} \bar{\boldsymbol{c}}_{i-1} \\ \check{\boldsymbol{c}}_{i-1} \end{bmatrix} \qquad (10.37)$$

This recursion has a form that is similar to the earlier recursion (9.285) we encountered while studying the mean stability of the original error dynamics (10.2) with two minor difference. First, the variables $\{\bar{\boldsymbol{z}}_i, \check{\boldsymbol{z}}_i, \bar{\boldsymbol{c}}_{i-1}, \check{\boldsymbol{c}}_{i-1}\}$ are now stochastic in nature and, second, the rightmost $O(\mu_{\max})$ perturbation term in (9.285) is absent from (10.37). Nevertheless, from an argument similar to the one that led to (9.282), we can similarly establish that

$$\|\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \boldsymbol{c}_{i-1}\|^2 \leq r^2 \mu_{\max}^2 \|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4 \qquad (10.38)$$

Proof

and, hence,

$$\|\bar{\boldsymbol{c}}_{i-1}\|^2 \leq r^2\mu_{\max}^2\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4, \qquad \|\check{\boldsymbol{c}}_{i-1}\|^2 \leq r^2\mu_{\max}^2\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4 \tag{10.39}$$

Moreover, repeating the argument that led to (9.292) and (9.294) we find that these recursions, under expectation, are now replaced by the following relations:

$$\begin{aligned}\mathbb{E}\,\|\bar{\boldsymbol{z}}_i\|^2 \quad &\leq \quad (1-\sigma_{11}\mu_{\max})\mathbb{E}\,\|\bar{\boldsymbol{z}}_{i-1}\|^2 + \\ &\qquad \frac{2\sigma_{21}^2\mu_{\max}}{\sigma_{11}}\mathbb{E}\,\|\check{\boldsymbol{z}}_{i-1}\|^2 + \frac{2r^2\mu_{\max}}{\sigma_{11}}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4\end{aligned} \tag{10.40}$$

Proof

$$
\begin{aligned}
\mathbb{E}\,\|\check{\boldsymbol{z}}_i\|^2 \quad \leq \quad & \left(\rho(J_\epsilon) + \epsilon + \frac{3\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\left\|\check{\boldsymbol{z}}_{i-1}^e\right\|^2 + \\
& \left(\frac{3\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\left\|\bar{\boldsymbol{z}}_{i-1}^e\right\|^2 + \\
& \left(\frac{3r^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4 \qquad (10.41)
\end{aligned}
$$

If we now introduce the scalar coefficients

Proof

$$a = 1 - \sigma_{11}\mu_{\max} = 1 - O(\mu_{\max}) \tag{10.42}$$

$$b = \frac{2\sigma_{21}^2\mu_{\max}}{\sigma_{11}} = O(\mu_{\max}) \tag{10.43}$$

$$c = \frac{3\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = O(\mu_{\max}^2) \tag{10.44}$$

$$d = \rho(J_\epsilon) + \epsilon + \frac{3\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \tag{10.45}$$

$$e = \frac{2r^2\mu_{\max}}{\sigma_{11}} = O(\mu_{\max}) \tag{10.46}$$

$$f = \frac{3r^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = O(\mu_{\max}^2) \tag{10.47}$$

Proof

we can combine (10.40) and (10.41) into a single compact inequality recursion as follows:

$$\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{z}}_i\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{z}}_i\|^2 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\Gamma} \begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{z}}_{i-1}\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{z}}_{i-1}\|^2 \end{bmatrix} + \underbrace{\begin{bmatrix} e \\ f \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix}} \underbrace{\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|^4}_{O(\mu_{\max}^2)} \qquad (10.48)$$

in terms of the 2×2 coefficient matrix Γ indicated above. Using the fact that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_i\|^4 \overset{(9.107)}{=} O(\mu_{\max}^2) \qquad (10.49)$$

Proof

and relation (9.103) we conclude that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\bar{\boldsymbol{z}}_i\|^2 \;=\; O(\mu_{\max}^2), \qquad \limsup_{i\to\infty} \mathbb{E}\,\|\check{\boldsymbol{z}}_i\|^2 \;=\; O(\mu_{\max}^4) \tag{10.50}$$

and, hence,

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2 \;=\; O(\mu_{\max}^2) \tag{10.51}$$

It follows that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e - \widetilde{\boldsymbol{w}}_i^{e'}\|^2 \;=\; O(\mu_{\max}^2) \tag{10.52}$$

which establishes (10.29). Finally, note that

$$(I-\Gamma)^{-1} = \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(\mu_{\max}) & O(1) \end{bmatrix}$$

Proof

$$
\begin{aligned}
\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'}\|^2 &= \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'} - \widetilde{\boldsymbol{w}}_i^{e} + \widetilde{\boldsymbol{w}}_i^{e}\|^2 \\
&\leq \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'} - \widetilde{\boldsymbol{w}}_i^{e}\|^2 + \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e}\|^2 \;+\; 2\left|\mathbb{E}\,(\widetilde{\boldsymbol{w}}_i^{e'} - \widetilde{\boldsymbol{w}}_i^{e})^*\widetilde{\boldsymbol{w}}_i^{e}\right| \\
&\leq \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'} - \widetilde{\boldsymbol{w}}_i^{e}\|^2 + \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e}\|^2 \;+\; 2\sqrt{\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'} - \widetilde{\boldsymbol{w}}_i^{e}\|^2\,\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e}\|^2}
\end{aligned}
\tag{10.53}
$$

Proof

and, hence, from (9.11) and (10.29) we get

$$\limsup_{i\to\infty}\left(\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e'}\|^2 - \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^{e}\|^2\right) \;\leq\; O(\mu_{\max}^2) \;+\; \sqrt{O(\mu_{\max}^3)} \;=\; O(\mu_{\max}^{3/2}) \tag{10.54}$$

since $\mu_{\max}^2 < \mu_{\max}^{3/2}$ for small $\mu_{\max} \ll 1$, which establishes (10.30).

$\square$

Second-Order Stability

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.

Second-Order Error Moment

We already know from the result of Theorem 9.1 that the original error recursion (10.2) is mean-square stable in the sense that $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2$ tends asymptotically to a region that is bounded by $O(\mu_{\max})$. Before launching into the performance analysis of the stochastic-gradient distributed algorithm (8.46), we first remark that the long-term approximate model (10.19) is also mean-square stable.

Second-Order Error Moment

Lemma 10.3 (Mean-square stability of long-term model). 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 second-order moments of the gradient noise process satisfy the conditions of Assumption 8.1. Consider the iterates that are generated by the long-term model (10.19). Then, for sufficiently small step-sizes, it holds that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}'_{k,i}\|^2 \;=\; O(\mu_{\max}), \quad k = 1, 2, \ldots, N \qquad (10.55)$$

Proof

Proof. We multiply both sides of the long-term model (10.19) from the left by $\mathcal{V}_\epsilon^\mathsf{T}$ to get, for $i \gg 1$:

$$\underbrace{\begin{bmatrix} \bar{\boldsymbol{w}}_i^{e'} \\ \check{\boldsymbol{w}}_i^{e'} \end{bmatrix}}_{\triangleq\, \boldsymbol{z}_i} = \underbrace{\begin{bmatrix} I_{2M} - D_{11}^\mathsf{T} & -D_{21}^\mathsf{T} \\ -D_{12}^\mathsf{T} & \mathcal{J}_\epsilon^\mathsf{T} - D_{22}^\mathsf{T} \end{bmatrix}}_{\triangleq\, \bar{\mathcal{B}}} \underbrace{\begin{bmatrix} \bar{\boldsymbol{w}}_{i-1}^{e'} \\ \check{\boldsymbol{w}}_{i-1}^{e'} \end{bmatrix}}_{\triangleq\, \boldsymbol{z}_{i-1}} +$$

$$\mathcal{V}_\epsilon^\mathsf{T} \mathcal{A}_2^\mathsf{T} \mathcal{M} \boldsymbol{s}_i^e - \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \qquad (10.56)$$

Proof

where the matrix $\bar{\mathcal{B}}$ is stable by Theorem 9.3, and where we are denoting the transformed error vector by $\boldsymbol{z}_i$ for ease of reference:

$$\boldsymbol{z}_i \triangleq \mathcal{V}_\epsilon^{\mathsf{T}} \widetilde{\boldsymbol{w}}_i^{e'} = \begin{bmatrix} \bar{\boldsymbol{w}}_i^{e'} \\ \check{\boldsymbol{w}}_i^{e'} \end{bmatrix} \tag{10.57}$$

Proof

We are also dropping the argument $\boldsymbol{w}_{i-1}^e$ from $\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)$ and writing simply $\boldsymbol{s}_i^e$. The long-term model (10.56) represents a dynamic system that is driven by two components: a deterministic (constant) driving term represented by $\check{b}^e$, and a random term represented by $\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)$. To facilitate the mean-square stability analysis, we may examine the contribution of these driving terms separately. For this purpose, we introduce the following two auxiliary recursions, one driven by the deterministic term and the other driven by the stochastic term and running over $i > i_o$ for some large enough $i_o \gg 1$:

Proof

$$a_i = \bar{\mathcal{B}}\, a_{i-1} + \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \tag{10.58}$$

$$\boldsymbol{b}_i = \bar{\mathcal{B}}\, \boldsymbol{b}_{i-1} + \mathcal{V}_\epsilon^\mathsf{T} \mathcal{A}_2^\mathsf{T} \mathcal{M} \boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) \tag{10.59}$$

with initial conditions $a_{i_o} = 0$ and $\boldsymbol{b}_{i_o} = \boldsymbol{z}_{i_o}$ so that at any time instant $i > i_o$,

$$\boldsymbol{z}_i = \boldsymbol{b}_i - a_i \tag{10.60}$$

Consider first recursion (10.58) for a_i. Since $\bar{\mathcal{B}}$ is stable, the sequence a_i converges to

Proof

$$\lim_{i\to\infty} a_i = (I - \bar{\mathcal{B}})^{-1} \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \overset{(9.229)}{=} O(\mu_{\max}) \qquad (10.61)$$

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

since $\check{b}^e = O(\mu_{\max})$. It follows that

$$\limsup_{i\to\infty} \|a_i\| = O(\mu_{\max}) \qquad (10.62)$$

Consider next recursion (10.59) for $\boldsymbol{b}_i$. As was done earlier in (9.56) we partition the entries of $\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \boldsymbol{s}_i^e$ into:

Proof

$$\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) \;\stackrel{\Delta}{=}\; \begin{bmatrix} \bar{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \\ \check{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \end{bmatrix} \tag{10.63}$$

We also partition the entries of $\boldsymbol{b}_i$ in the following manner in conformity with the dimensions of $\{\bar{\boldsymbol{s}}_i^e, \check{\boldsymbol{s}}_i^e\}$:

$$\underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_i \\ \check{\boldsymbol{b}}_i \end{bmatrix}}_{=\boldsymbol{b}_i} = \underbrace{\begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix}}_{\stackrel{\Delta}{=}\, \bar{\mathcal{B}}} \underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_{i-1} \\ \check{\boldsymbol{b}}_{i-1} \end{bmatrix}}_{=\boldsymbol{b}_{i-1}} + \begin{bmatrix} \bar{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \\ \check{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \end{bmatrix} \tag{10.64}$$

Sketch of Proof

we can combine (10.68) and (10.69) into a single compact inequality recursion as follows:

$$\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^2 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\Gamma} \begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\| \end{bmatrix} + \underbrace{\begin{bmatrix} h & h \\ h & h \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}^2) & O(\mu_{\max}^2) \\ O(\mu_{\max}^2) & O(\mu_{\max}^2) \end{bmatrix}} \underbrace{\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^2 \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix}} + \underbrace{\begin{bmatrix} e \\ e \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}^2) \\ O(\mu_{\max}^2) \end{bmatrix}}$$

in terms of the 2×2 coefficient matrix Γ indicated above. Using result (9.105) and the derivation leading to it we can similarly conclude that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^2 = O(\mu_{\max}), \qquad \limsup_{i\to\infty} \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^2 = O(\mu_{\max}^2) \tag{10.78}$$

Sketch of Proof

and, hence,

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{b}_i\|^2 \;=\; O(\mu_{\max}) \tag{10.79}$$

From (10.60) we have that $\|\boldsymbol{z}_i\|^2 \leq 2\|a_i\|^2 + 2\|\boldsymbol{b}_i\|^2$ so that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2 \;=\; O(\mu_{\max}) \tag{10.80}$$

from which we conclude that (10.55) holds.

$\square$

Back to Detailed Proof

We also partition the entries of $\boldsymbol{b}_i$ in the following manner in conformity with the dimensions of $\{\bar{\boldsymbol{s}}_i^e, \check{\boldsymbol{s}}_i^e\}$:

$$\underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_i \\ \check{\boldsymbol{b}}_i \end{bmatrix}}_{=\boldsymbol{b}_i} = \underbrace{\begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_{\epsilon}^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix}}_{\triangleq \bar{\mathcal{B}}} \underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_{i-1} \\ \check{\boldsymbol{b}}_{i-1} \end{bmatrix}}_{=\boldsymbol{b}_{i-1}} + \begin{bmatrix} \bar{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \\ \check{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \end{bmatrix} \quad (10.64)$$

Proof

This recursion has a form similar to the earlier recursion we encountered in (9.60) while studying the mean-square stability of the original error dynamics (10.2), with three differences. First, the driving term involving $\check{b}^e$ in (9.60) is not present in (10.64). Second, the matrices $\{D_{11}, D_{12}, D_{21}, D_{22}\}$ in (10.64) are constant matrices; nevertheless, they satisfy the same bounds as the matrices $\{\boldsymbol{D}_{11,i-1}, \boldsymbol{D}_{12,i-1}, \boldsymbol{D}_{21,i-1}, \boldsymbol{D}_{22,i-1}\}$ in (9.60). And, third, the argument of the noise terms $\{\bar{\boldsymbol{s}}_i^e, \check{\boldsymbol{s}}_i^e\}$ in (10.64) is $\boldsymbol{w}_{i-1}^e$ and not $\boldsymbol{b}_i$. However, these noise terms still satisfy the same bound given by (9.91), namely,

$$\mathbb{E}\,\|\bar{\boldsymbol{s}}_i^e\|^2 + \mathbb{E}\,\|\check{\boldsymbol{s}}_i^e\|^2 \;\;\leq\;\; v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 \left[\mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^2 + \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^2\right] + v_1^2 \mu_{\max}^2 \sigma_s^2 \tag{10.65}$$

Proof

in terms of the transformed vectors $\{\bar{\boldsymbol{w}}^e_{i-1}, \check{\boldsymbol{w}}^e_{i-1}\}$ defined by (9.55). Therefore, repeating the same argument that led to (9.106) will show that relations (9.69) and (9.81) still hold for $\{\mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^2, \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^2\}$, namely,

$$\mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^2 \;\leq\; (1-\sigma_{11}\mu_{\max})\mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^2 \;+\; \left(\frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}\right)\mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\|^2 \;+\; \mathbb{E}\,\|\bar{\boldsymbol{s}}^e_i\|^2 \tag{10.66}$$

and

$$\mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^2 \;\leq\; \left(\rho(J_\epsilon)+\epsilon+\frac{2\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\left\|\check{\boldsymbol{b}}_{i-1}\right\|^2 + \left(\frac{2\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\left\|\bar{\boldsymbol{b}}_{i-1}\right\|^2 + \mathbb{E}\,\|\check{\boldsymbol{s}}^e_i\|^2 \tag{10.67}$$

Proof

Using (10.65) we find that the last two recursive inequalities can be replaced by

$$\begin{aligned}\mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^2 \;\le\; & (1-\sigma_{11}\mu_{\max})\,\mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^2 + \left(\frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}\right)\mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\|^2 + \\ & v_1^2\mu_{\max}^2\sigma_s^2 + v_1^2 v_2^2\beta_d^2\mu_{\max}^2\left[\mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^2 + \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^2\right]\end{aligned} \tag{10.68}$$

Proof

$$
\begin{aligned}
\mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^2 \quad \leq \quad & \left(\rho(J_\epsilon) + \epsilon + \frac{2\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\left\|\check{\boldsymbol{b}}_{i-1}\right\|^2 + \\
& \left(\frac{2\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\left\|\bar{\boldsymbol{b}}_{i-1}\right\|^2 + v_1^2\mu_{\max}^2\sigma_s^2 \ + \\
& v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 \left[\mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^2 \ + \ \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^2\right] \qquad (10.69)
\end{aligned}
$$

If we now introduce the scalar coefficients

Proof

$$a = 1 - \sigma_{11}\mu_{\max} = 1 - O(\mu_{\max}) \quad (10.70)$$

$$b = \frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}} = O(\mu_{\max}) \quad (10.71)$$

$$c = \frac{2\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = O(\mu_{\max}^2) \quad (10.72)$$

$$d = \rho(J_\epsilon) + \epsilon + \frac{3\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \quad (10.73)$$

$$e = v_1^2\mu_{\max}^2\sigma_s^2 = O(\mu_{\max}^2) \quad (10.74)$$

$$f = 0 \quad (10.75)$$

$$h = v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 = O(\mu_{\max}^2) \quad (10.76)$$

Proof

we can combine (10.68) and (10.69) into a single compact inequality recursion as follows:

$$\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^2 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\Gamma} \begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\| \end{bmatrix} + \underbrace{\begin{bmatrix} h & h \\ h & h \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}^2) & O(\mu_{\max}^2) \\ O(\mu_{\max}^2) & O(\mu_{\max}^2) \end{bmatrix}} \underbrace{\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^2 \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix}} + \underbrace{\begin{bmatrix} e \\ e \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}^2) \\ O(\mu_{\max}^2) \end{bmatrix}}$$

in terms of the 2×2 coefficient matrix Γ indicated above. Using result (9.105) and the derivation leading to it we can similarly conclude that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^2 = O(\mu_{\max}), \qquad \limsup_{i\to\infty} \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^2 = O(\mu_{\max}^2) \tag{10.78}$$

Proof

and, hence,

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{b}_i\|^2 \;=\; O(\mu_{\max}) \tag{10.79}$$

From (10.60) we have that $\|\boldsymbol{z}_i\|^2 \leq 2\|a_i\|^2 + 2\|\boldsymbol{b}_i\|^2$ so that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^2 \;=\; O(\mu_{\max}) \tag{10.80}$$

from which we conclude that (10.55) holds.

$\square$

Fourth-Order Stability

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.

Fourth-Order Error Moment

In the next chapter we will employ the long-term model (10.19) to assess the performance of the multi-agent network as $i \to \infty$ and for sufficiently small step-sizes. In preparation for that discussion, we establish here the stability of the fourth-order moment of the error in the long-term model (10.19) in a manner similar to what we did in Theorem 9.2 for the fourth-order moment of the error in the original recursion (10.2).

Fourth-Order Error Moment

Lemma 10.4 (Fourth-order moment stability of long-term model). 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). Then, the fourth-order moments of the error vectors generated by the long-term model (10.19) are stable for sufficiently small step-sizes, namely, it holds that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}'_{k,i}\|^4 \;=\; O(\mu_{\max}^2), \quad k = 1, 2, \ldots, N \tag{10.81}$$

Proof

Proof. We employ the same notation from the proof of Lemma 10.3 and reconsider recursions (10.58) and (10.64) for the auxiliary variables $\{a_i, \boldsymbol{b}_i\}$:

$$a_i \;=\; \bar{\mathcal{B}}\, a_{i-1} + \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \tag{10.82}$$

$$\underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_i \\ \check{\boldsymbol{b}}_i \end{bmatrix}}_{=\boldsymbol{b}_i} \;=\; \underbrace{\begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_{\epsilon}^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix}}_{\triangleq\, \bar{\mathcal{B}}} \underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_{i-1} \\ \check{\boldsymbol{b}}_{i-1} \end{bmatrix}}_{=\boldsymbol{b}_{i-1}} + \begin{bmatrix} \bar{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \\ \check{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \end{bmatrix} \tag{10.83}$$

Using (10.62), we readily conclude from (10.62) that

$$\limsup_{i\to\infty} \|a_i\|^4 \;=\; O(\mu_{\max}^4) \tag{10.84}$$

Sketch of Proof

Therefore, repeating the same argument that led to (9.153) we can similarly show that

$$\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^4 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^4 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\triangleq\, \Gamma'} \begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^4 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\|^4 \end{bmatrix} + \underbrace{\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}^2) & O(\mu_{\max}^3) \\ O(\mu_{\max}^4) & O(\mu_{\max}^2) \end{bmatrix}} \underbrace{\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\|^2 \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix}} +$$

$$\begin{bmatrix} a'' & b'' \\ c'' & d'' \end{bmatrix} \begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^2 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \qquad (10.89)$$

where

$$\begin{bmatrix} O(\mu_{\max}^2) & O(\mu_{\max}^2) \\ O(\mu_{\max}^2) & O(\mu_{\max}^2) \end{bmatrix} \qquad \begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix} \qquad \begin{bmatrix} O(\mu_{\max}^4) \\ O(\mu_{\max}^4) \end{bmatrix}$$

Proof

and

$$\Gamma' = \begin{bmatrix} 1 - O(\mu_{\max}) & O(\mu_{\max}) \\ O(\mu_{\max}^4) & \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \end{bmatrix} \qquad (10.102)$$

We again find that Γ' is a stable matrix for sufficiently small $\mu_{\max}$ and ϵ. Using results (9.105) and (10.78), and repeating the argument that led to (9.156) we conclude that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^4 = O(\mu_{\max}^2), \quad \limsup_{i\to\infty} \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^4 = O(\mu_{\max}^4) \qquad (10.103)$$

so that

Proof

$$
\begin{aligned}
\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{b}_i\|^4 &= \limsup_{i\to\infty} \mathbb{E}\left(\left\|\begin{bmatrix} \bar{\boldsymbol{b}}_i \\ \check{\boldsymbol{b}}_i \end{bmatrix}\right\|^2\right)^2 \\
&= \limsup_{i\to\infty} \mathbb{E}\left(\|\bar{\boldsymbol{b}}_i\|^2 + \|\check{\boldsymbol{b}}_i\|^2\right)^2 \\
&\leq 2\left(\limsup_{i\to\infty} \mathbb{E}\left(\|\bar{\boldsymbol{b}}_i\|^4 + \|\check{\boldsymbol{b}}_i\|^4\right)\right) \\
&= O(\mu_{\max}^2) \qquad (10.104)
\end{aligned}
$$

Proof

Now, from $\boldsymbol{z}_i = \boldsymbol{b}_i - a_i$ we have

$$\|\boldsymbol{z}_i\|^4 \;\leq\; 8\|\boldsymbol{b}_i\|^4 \;+\; 8\|a_i\|^4 \tag{10.105}$$

and, therefore,

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^4 \;=\; O(\mu_{\max}^2) \tag{10.106}$$

Consequently,

Proof

$$
\begin{aligned}
\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^4 &= \limsup_{i\to\infty}\ \mathbb{E}\left(\left\|\left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}}\begin{bmatrix}\bar{\boldsymbol{w}}_i^e\\ \check{\boldsymbol{w}}_i^e\end{bmatrix}\right\|^2\right)^2\\
&= \limsup_{i\to\infty}\ \mathbb{E}\left(\left\|\left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}}\boldsymbol{z}_i\right\|^2\right)^2\\
&\le \left\|\left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}}\right\|^4\left(\limsup_{i\to\infty}\ \mathbb{E}\,\|\boldsymbol{z}_i\|^4\right)\\
&= O(\mu_{\max}^2) \qquad\qquad (10.107)
\end{aligned}
$$

which leads to the desired result (10.81).

$\square$

Back to Detailed Proof

Proof. We employ the same notation from the proof of Lemma 10.3 and reconsider recursions (10.58) and (10.64) for the auxiliary variables $\{a_i, \boldsymbol{b}_i\}$:

$$a_i = \bar{\mathcal{B}}\, a_{i-1} + \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \tag{10.82}$$

$$\underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_i \\ \check{\boldsymbol{b}}_i \end{bmatrix}}_{=\boldsymbol{b}_i} = \underbrace{\begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix}}_{\triangleq\, \bar{\mathcal{B}}} \underbrace{\begin{bmatrix} \bar{\boldsymbol{b}}_{i-1} \\ \check{\boldsymbol{b}}_{i-1} \end{bmatrix}}_{=\boldsymbol{b}_{i-1}} + \begin{bmatrix} \bar{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \\ \check{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \end{bmatrix} \tag{10.83}$$

Using (10.62), we readily conclude from (10.62) that

$$\limsup_{i\to\infty} \|a_i\|^4 = O(\mu_{\max}^4) \tag{10.84}$$

Proof

With regards to the recursion involving $\{\bar{\boldsymbol{b}}_i^e, \check{\boldsymbol{b}}_i^e\}$, we can unfold it and write

$$\begin{aligned}
\bar{\boldsymbol{b}}_i^e &= (I_{2M} - D_{11}^{\mathsf{T}})\bar{\boldsymbol{b}}_{i-1}^e - D_{21}^{\mathsf{T}}\check{\boldsymbol{b}}_{i-1}^e + \bar{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \qquad (10.85)\\
\check{\boldsymbol{b}}_i^e &= (\mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}})\check{\boldsymbol{b}}_{i-1}^e - D_{12}^{\mathsf{T}}\bar{\boldsymbol{b}}_{i-1}^e + \check{\boldsymbol{s}}_i^e(\boldsymbol{w}_{i-1}^e) \qquad (10.86)
\end{aligned}$$

These relations have similar forms to the earlier relations (9.108)–(9.109) we encountered while studying the stability of the fourth-order moment of the original error recursion (10.2), with three differences. First, the driving term involving $\check{b}^e$ in (9.109) is not present in (10.86). Second, the matrices $\{D_{11}, D_{12}, D_{21}, D_{22}\}$ in (10.85)–(10.86) are constant

Proof

matrices; nevertheless, they satisfy the same bounds as the matrices $\{\boldsymbol{D}_{11,i-1}, \boldsymbol{D}_{12,i-1}, \boldsymbol{D}_{21,i-1}, \boldsymbol{D}_{22,i-1}\}$ in (9.108)–(9.109). And, third, the argument of the noise terms $\{\bar{\boldsymbol{s}}_i^e, \check{\boldsymbol{s}}_i^e\}$ in (10.85)–(10.86) is $\boldsymbol{w}_{i-1}^e$ and not $\boldsymbol{b}_i$. However, these noise terms still satisfy the same bounds given by (9.91) and (9.131), namely,

$$\mathbb{E}\,\|\bar{\boldsymbol{s}}_i^e\|^2 + \mathbb{E}\,\|\check{\boldsymbol{s}}_i^e\|^2 \leq v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 \left[\mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^2 + \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^2\right] \;+\; v_1^2 \mu_{\max}^2 \sigma_s^2 \tag{10.87}$$

and

$$\mathbb{E}\,\|\bar{\boldsymbol{s}}_i^e\|^4 + \mathbb{E}\,\|\check{\boldsymbol{s}}_i^e\|^4 \leq v_1^4 v_2^4 \beta_{d4}^4 \mu_{\max}^4 \left[\mathbb{E}\,\|\bar{\boldsymbol{w}}_{i-1}^e\|^4 + \mathbb{E}\,\|\check{\boldsymbol{w}}_{i-1}^e\|^4\right] + v_1^4 \mu_{\max}^4 \sigma_{s4}^4 \tag{10.88}$$

Proof

Therefore, repeating the same argument that led to (9.153) we can similarly show that

$$
\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^4 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^4 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\triangleq\, \Gamma'} \begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^4 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\|^4 \end{bmatrix} + \underbrace{\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}^2) & O(\mu_{\max}^3) \\ O(\mu_{\max}^4) & O(\mu_{\max}^2) \end{bmatrix}} \underbrace{\begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{b}}_{i-1}\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{b}}_{i-1}\|^2 \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix}} +
$$

$$
\begin{bmatrix} a'' & b'' \\ c'' & d'' \end{bmatrix} \begin{bmatrix} \mathbb{E}\,\|\bar{\boldsymbol{w}}^e_{i-1}\|^2 \\ \mathbb{E}\,\|\check{\boldsymbol{w}}^e_{i-1}\|^2 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \qquad (10.89)
$$

where

$$
\begin{bmatrix} O(\mu_{\max}^2) & O(\mu_{\max}^2) \\ O(\mu_{\max}^2) & O(\mu_{\max}^2) \end{bmatrix} \qquad \begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix} \qquad \begin{bmatrix} O(\mu_{\max}^4) \\ O(\mu_{\max}^4) \end{bmatrix}
$$

Proof

$$
\begin{aligned}
a &= 1 - \sigma_{11}\mu_{\max} + O(\mu_{\max}^2) && (10.90)\\
b &= O(\mu_{\max}) && (10.91)\\
c &= O(\mu_{\max}^4) && (10.92)\\
d &= \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) && (10.93)\\
a' &= O(\mu_{\max}^2) && (10.94)\\
b' &= O(\mu_{\max}^3) && (10.95)\\
c' &= O(\mu_{\max}^4) && (10.96)\\
d' &= O(\mu_{\max}^2) && (10.97)\\
a'' &= O(\mu_{\max}^2) && (10.98)\\
b'' &= O(\mu_{\max}^2) && (10.99)\\
c'' &= O(\mu_{\max}^2) && (10.100)\\
d'' &= O(\mu_{\max}^2) && (10.101)
\end{aligned}
$$

Proof

and

$$\Gamma' = \begin{bmatrix} 1 - O(\mu_{\max}) & O(\mu_{\max}) \\ O(\mu_{\max}^4) & \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \end{bmatrix} \tag{10.102}$$

We again find that Γ' is a stable matrix for sufficiently small $\mu_{\max}$ and ϵ. Using results (9.105) and (10.78), and repeating the argument that led to (9.156) we conclude that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\bar{\boldsymbol{b}}_i\|^4 = O(\mu_{\max}^2), \quad \limsup_{i\to\infty} \mathbb{E}\,\|\check{\boldsymbol{b}}_i\|^4 = O(\mu_{\max}^4) \tag{10.103}$$

so that

Proof

$$
\begin{aligned}
\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{b}_i\|^4 &= \limsup_{i\to\infty} \mathbb{E}\left(\left\|\begin{bmatrix} \bar{\boldsymbol{b}}_i \\ \check{\boldsymbol{b}}_i \end{bmatrix}\right\|^2\right)^2 \\
&= \limsup_{i\to\infty} \mathbb{E}\left(\|\bar{\boldsymbol{b}}_i\|^2 + \|\check{\boldsymbol{b}}_i\|^2\right)^2 \\
&\leq 2\left(\limsup_{i\to\infty} \mathbb{E}\left(\|\bar{\boldsymbol{b}}_i\|^4 + \|\check{\boldsymbol{b}}_i\|^4\right)\right) \\
&= O(\mu_{\max}^2) \qquad (10.104)
\end{aligned}
$$

Proof

Now, from $\boldsymbol{z}_i = \boldsymbol{b}_i - a_i$ we have

$$\|\boldsymbol{z}_i\|^4 \;\leq\; 8\|\boldsymbol{b}_i\|^4 \;+\; 8\|a_i\|^4 \tag{10.105}$$

and, therefore,

$$\limsup_{i\to\infty} \mathbb{E}\,\|\boldsymbol{z}_i\|^4 \;=\; O(\mu_{\max}^2) \tag{10.106}$$

Consequently,

Proof

$$
\begin{aligned}
\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i^e\|^4 &= \limsup_{i\to\infty} \mathbb{E}\left(\left\|\left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf T}\begin{bmatrix}\bar{\boldsymbol{w}}_i^e\\ \check{\boldsymbol{w}}_i^e\end{bmatrix}\right\|^2\right)^2\\
&= \limsup_{i\to\infty} \mathbb{E}\left(\left\|\left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf T}\boldsymbol{z}_i\right\|^2\right)^2\\
&\le \left\|\left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf T}\right\|^4\left(\limsup_{i\to\infty}\ \mathbb{E}\,\|\boldsymbol{z}_i\|^4\right)\\
&= O(\mu_{\max}^2)
\end{aligned}
\qquad (10.107)
$$

which leads to the desired result (10.81).

□

First-Order Stability

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.

First-Order Error Moment

Lemma 10.5 (Mean stability of long-term model)**.** 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 second-order moments of the gradient noise process satisfy the conditions of Assumption 8.1. Consider the iterates that are generated by the long-term model (10.19). Then, for sufficiently small step-sizes, it holds that

$$\limsup_{i\to\infty} \|\mathbb{E}\, \widetilde{\boldsymbol{w}}'_{k,i}\| = O(\mu_{\max}), \quad k = 1, 2, \ldots, N \tag{10.108}$$

Proof

Proof. Conditioning both sides of (10.19) on $\boldsymbol{\mathcal{F}}_{i-1}$, invoking the conditions on the gradient noise process from Assumption 8.1, and computing the conditional expectations we obtain:

$$\mathbb{E}\left[\widetilde{\boldsymbol{w}}_i^{e'} \,|\, \boldsymbol{\mathcal{F}}_{i-1}\right] = \mathcal{B}\widetilde{\boldsymbol{w}}_{i-1}^{e'} - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \tag{10.109}$$

where the term involving $\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e)$ is eliminated because $\mathbb{E}\left[\boldsymbol{s}_i^e|\boldsymbol{\mathcal{F}}_{i-1}\right] = 0$. Taking expectations again we arrive at

$$\mathbb{E}\,\widetilde{\boldsymbol{w}}_i^{e'} = \mathcal{B}\left(\mathbb{E}\,\widetilde{\boldsymbol{w}}_{i-1}^{e'}\right) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \tag{10.110}$$

Proof

We multiply both sides of this recursion from the left by $\mathcal{V}_\epsilon^{\mathsf{T}}$ to get

$$\underbrace{\begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^{e'} \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^{e'} \end{bmatrix}}_{\triangleq\, z_i} = \underbrace{\begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix}}_{\triangleq\, \bar{\mathcal{B}}} \underbrace{\begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_{i-1}^{e'} \\ \mathbb{E}\,\check{\boldsymbol{w}}_{i-1}^{e'} \end{bmatrix}}_{\triangleq\, z_{i-1}} - \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \qquad (10.111)$$

where the matrix $\bar{\mathcal{B}}$ is stable by Theorem 9.3. For simplicity, we denote the state variable in (10.111) by z_i, so that we can rewrite the recursion more compactly in the form

$$z_i = \bar{\mathcal{B}} z_{i-1} - \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \qquad (10.112)$$

Proof

This is a first-order recursion that is driven by a constant term. Since $\bar{\mathcal{B}}$ is stable and $\check{b}^e = O(\mu_{\max})$, we conclude from (10.112) that

$$\begin{aligned}
\lim_{i\to\infty} z_i \quad &= \quad -(I-\bar{\mathcal{B}})^{-1}\begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \\
&\overset{(9.229)}{=} \quad \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(1) & O(1) \end{bmatrix}\begin{bmatrix} 0 \\ O(\mu_{\max}) \end{bmatrix} \\
&= \quad O(\mu_{\max})
\end{aligned} \tag{10.113}$$

Proof

It follows that

$$\limsup_{i\to\infty} \|z_i\| \;=\; O(\mu_{\max}) \tag{10.114}$$

Consequently,

$$\limsup_{i\to\infty} \left\| \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^{e'} \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^{e'} \end{bmatrix} \right\| \;=\; O(\mu_{\max}) \tag{10.115}$$

and, hence,

Proof

$$
\begin{aligned}
\limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}'_{k,i}\| &\leq \limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_i^{e'}\| \\
&\leq \limsup_{i\to\infty} \left\| \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^{e'} \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^{e'} \end{bmatrix} \right\| \\
&\leq \left\| \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \right\| \left(\limsup_{i\to\infty} \left\| \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^{e'} \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^{e'} \end{bmatrix} \right\| \right) \\
&= O(\mu_{\max})
\end{aligned}
\qquad (10.116)
$$

as claimed.

□

Comparing Consensus & Diffusion

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.

Setting

Using results from the previous sections, we are able to compare some stability properties of diffusion and consensus networks. Recall from (8.7)–(8.10) that the consensus and diffusion strategies correspond to the following choices for $\{A_o, A_1, A_2\}$ in terms of a single combination matrix A in the general description (8.46):

$$\text{consensus:} \qquad A_o = A, \;\; A_1 = I_N = A_2 \qquad (10.117)$$

$$\text{CTA diffusion:} \qquad A_1 = A, \;\; A_2 = I_N = A_o \qquad (10.118)$$

$$\text{ATC diffusion:} \qquad A_2 = A, \;\; A_1 = I_N = A_o \qquad (10.119)$$

Example #10.1

Example 10.1 (Stabilizing effect of diffusion networks). We revisit the conclusion of Example 8.4, albeit now under more general costs. Thus, refer to the mean recursion (10.110), namely,

$$\mathbb{E}\,\tilde{\boldsymbol{w}}_i^{e'} = \mathcal{B}\left(\mathbb{E}\,\tilde{\boldsymbol{w}}_{i-1}^{e'}\right) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \tag{10.120}$$

which is driven by a constant matrix $\mathcal{B}$. Using the choices (10.117)–(10.119), the $\mathcal{B}$ matrix is given by the following expressions in terms of the $\mathcal{B}$ matrix for the non-cooperative strategy:

Example #10.1

$$\begin{aligned}
\mathcal{B}_{\text{ncop}} &= I_{hMN} - \mathcal{M}\mathcal{H} && \text{(non-cooperation)} && (10.121)\\
\mathcal{B}_{\text{cons}} &= \mathcal{B}_{\text{ncop}} + \left(\mathcal{A}^{\mathsf{T}} - I_{hMN}\right) && \text{(consensus)} && (10.122)\\
\mathcal{B}_{\text{atc}} &= \mathcal{A}^{\mathsf{T}}\mathcal{B}_{\text{ncop}} && \text{(ATC diffusion)} && (10.123)\\
\mathcal{B}_{\text{cta}} &= \mathcal{B}_{\text{ncop}}\mathcal{A}^{\mathsf{T}} && \text{(CTA diffusion)} && (10.124)
\end{aligned}$$

where $\mathcal{A} = A \otimes I_{hM}$ and $h = 1$ for real data and $h = 2$ for complex data. We encountered a similar structure in expressions (8.30)–(8.33) for the case of MSE networks in Example 8.3, where the mean error vector evolved instead according to the recursion:

$$\mathbb{E}\,\widetilde{\boldsymbol{w}}_i = \mathcal{B}\left(\mathbb{E}\,\widetilde{\boldsymbol{w}}_{i-1}\right) \tag{10.125}$$

Example #10.1

without the additional driving terms appearing in (10.120). Now, observe that the coefficient matrices $\{\mathcal{B}_{\rm atc}, \mathcal{B}_{\rm cta}\}$ shown in (10.123)–(10.124) for the diffusion strategies are expressed in terms of $\mathcal{B}_{\rm ncop}$ in a *multiplicative* manner, while $\mathcal{B}_{\rm cons}$ is related to $\mathcal{B}_{\rm ncop}$ in an *additive* manner. These structures have an important implication on mean stability in view of the following matrix result.

Example #10.1

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be any left-stochastic matrices with blocks of size $hM \times hM$, and let $\mathcal{D}$ be any Hermitian block-diagonal positive-definite matrix also with blocks of size $hM \times hM$. Then, it holds from property (F.24) in the appendix that $\rho(\mathcal{X}_2^{\mathsf{T}} \mathcal{D} \mathcal{X}_1^{\mathsf{T}}) \leq \rho(\mathcal{D})$. That is, multiplication of $\mathcal{D}$ by left-stochastic transformations generally reduces the spectral radius. This result can be used to establish the stability of the diffusion dynamics (i.e., of $\mathcal{B}_{\rm diff}$) whenever the non-cooperative strategy is stable (i.e., $\mathcal{B}_{\rm ncop}$) and regardless of the combination policy, A. Indeed, note that $\mathcal{B}_{\rm ncop}$ has a Hermitian block-diagonal structure similar to $\mathcal{D}$ and that it is stable for any $\mu_{\max} < 2/\rho(\mathcal{H})$:

Example #10.1

$$\mathcal{B}_{\text{ncop}} \text{ stable} \iff \mu_{\max} < \frac{2}{\rho(\mathcal{H})} \tag{10.126}$$

The matrix $\mathcal{A}$ in (10.123)–(10.124) plays the role of $\mathcal{X}_1$ or $\mathcal{X}_2$. Therefore, it follows that, whenever (10.126) holds, it will also hold that $\rho(\mathcal{B}_{\text{atc}}) < 1$ and $\rho(\mathcal{B}_{\text{cta}}) < 1$ for any $\mathcal{A}$. The same conclusion does not generally hold for $\mathcal{B}_{\text{cons}}$ [248]. Note further that since $\rho(\mathcal{B}_{\text{atc}}) \leq \rho(\mathcal{B}_{\text{ncop}})$ and $\rho(\mathcal{B}_{\text{cta}}) \leq \rho(\mathcal{B}_{\text{ncop}})$, it follows that diffusion strategies have a stabilizing effect.

■

Example #10.2

Example 10.2 (Two interacting agents). We illustrate further the conclusion of Example 10.1 by considering the case of an MSE network (cf. Example 8.2) consisting of two interacting agents shown in Figure 10.1 [248], with

$$R_{u,1} = \sigma_{u,1}^2 I_M, \quad R_{u,2} = \sigma_{u,2}^2 I_M \tag{10.127}$$

Without loss of generality, we assume

$$\mu_1 \sigma_{u,1}^2 \leq \mu_2 \sigma_{u,2}^2 \tag{10.128}$$

Example #10.2

Agent 1 uses combination weights $\{1-a, a\}$, while agent 2 uses combination weights $\{1-b, b\}$ with $a, b \in (0,1)$. The combination matrix A is therefore given by

$$A = \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} \tag{10.129}$$

which is left-stochastic. If desired, a symmetric A can be obtained by setting $a = b$.

Example #10.2

The agents run either the consensus LMS strategy (7.14) or the diffusion LMS strategy (7.22) or (7.23). We already know from (8.28) in Example 8.2 that the mean error recursion for the non-cooperative, diffusion, and consensus LMS strategies running over complex data evolve according to the following dynamics:

$$\mathbb{E}\,\widetilde{\boldsymbol{w}}_i \;=\; \mathcal{B}\left(\mathbb{E}\,\widetilde{\boldsymbol{w}}_{i-1}\right), \qquad i \geq 0 \tag{10.130}$$

with the coefficient matrix $\mathcal{B}$ given by the following expressions for the various strategies under consideration (we are only showing the $\mathcal{B}$ matrix for the ATC strategy since the argument is similar for CTA):

Example #10.2

$$\mathcal{B}_{\rm ncop} = \begin{bmatrix} 1-\mu_1\sigma_{u,1}^2 & 0 \\ 0 & 1-\mu_2\sigma_{u,2}^2 \end{bmatrix} \otimes I_M \qquad (10.131)$$

$$\mathcal{B}_{\rm atc} = \begin{bmatrix} (1-a)\,(1-\mu_1\sigma_{u,1}^2) & a\,(1-\mu_2\sigma_{u,2}^2) \\ b\,(1-\mu_1\sigma_{u,1}^2) & (1-b)\,(1-\mu_2\sigma_{u,2}^2) \end{bmatrix} \otimes I_M \qquad (10.132)$$

$$\mathcal{B}_{\rm cons} = \begin{bmatrix} (1-a)-\mu_1\sigma_{u,1}^2 & a \\ b & (1-b)-\mu_2\sigma_{u,2}^2 \end{bmatrix} \otimes I_M \qquad (10.133)$$

Example #10.2

We first assume that

$$0 < \mu_1 \sigma_{u,1}^2 \leq \mu_2 \sigma_{u,2}^2 < 2 \tag{10.134}$$

so that each of the individual agents is stable in the mean and, hence, the matrix $\mathcal{B}_{\text{ncop}}$ given above is stable. Then, from the conclusion of Example 10.1 above we know that the diffusion network will also be stable in the mean for any choice of the parameters $\{a, b\}$. This is because the stability of $\mathcal{B}_{\text{ncop}}$ guarantees the stability of $\mathcal{B}_{\text{atc}}$.

Example #10.2

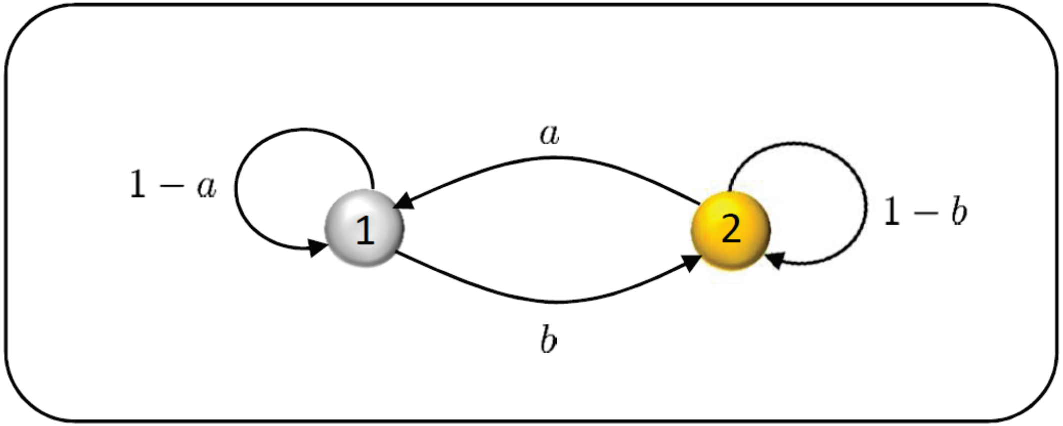


Figure 10.1: A two-agent MSE network with agent 1 using combination weights $\{a, 1-a\}$ and agent 2 using combination weights $\{b, 1-b\}$.

Example #10.2

We now verify that there are choices for the combination parameters $\{a, b\}$ that will destabilize the consensus network (even though the individual agents are themselves stable in the mean). Specifically, we verify below that if the parameters $\{a, b\} \in (0, 1)$ are chosen to satisfy

$$a + b \geq 2 - \mu_1 \sigma_{u,1}^2 > 0 \tag{10.135}$$

then consensus will lead to unstable mean behavior, i.e., $\mathbb{E}\,\widetilde{\boldsymbol{w}}_i$ will grow unbounded. Indeed, note first that the minimum eigenvalue of $\mathcal{B}_{\rm cons}$ can be found to be

$$\lambda_{\min}(\mathcal{B}_{\rm cons}) = \frac{1}{2}\left(\,(2 - a - b - \mu_1\sigma_{u,1}^2 - \mu_2\sigma_{u,2}^2) - \sqrt{\tau}\,\right) \tag{10.136}$$

Example #10.2

where

$$\begin{aligned}\tau &\triangleq (b - a - \mu_1\sigma_{u,1}^2 + \mu_2\sigma_{u,2}^2)^2 + 4ab \\ &= (b + a + \mu_1\sigma_{u,1}^2 - \mu_2\sigma_{u,2}^2)^2 + 4b(\mu_2\sigma_{u,2}^2 - \mu_1\sigma_{u,1}^2) \qquad (10.137)\end{aligned}$$

From the first equality in (10.137), we conclude that $\tau \geq 0$ and, hence, that $\lambda_{\min}(\mathcal{B}_{\text{cons}})$ is real. Moreover, using (10.134)–(10.135), we have that

$$b + a + \mu_1\sigma_{u,1}^2 - \mu_2\sigma_{u,2}^2 \geq 0 \qquad (10.138)$$

$$4b(\mu_2\sigma_{u,2}^2 - \mu_1\sigma_{u,1}^2) \geq 0 \qquad (10.139)$$

Example #10.2

It follows that

$$\begin{aligned}
\lambda_{\min}(\mathcal{B}_{\text{cons}}) &\leq \frac{1}{2}\left((2 - a - b - \mu_1\sigma_{u,1}^2 - \mu_2\sigma_{u,2}^2) - (b + a + \mu_1\sigma_{u,1}^2 - \mu_2\sigma_{u,2}^2)\right) \\
&= 1 - b - a - \mu_1\sigma_{u,1}^2 \\
&\leq -1 \qquad (10.140)
\end{aligned}$$

where (10.140) follows from (10.135). We conclude that the consensus network is unstable since the eigenvalues of $\mathcal{B}_{\text{cons}}$ do not lie strictly inside the unit circle.

Example #10.2

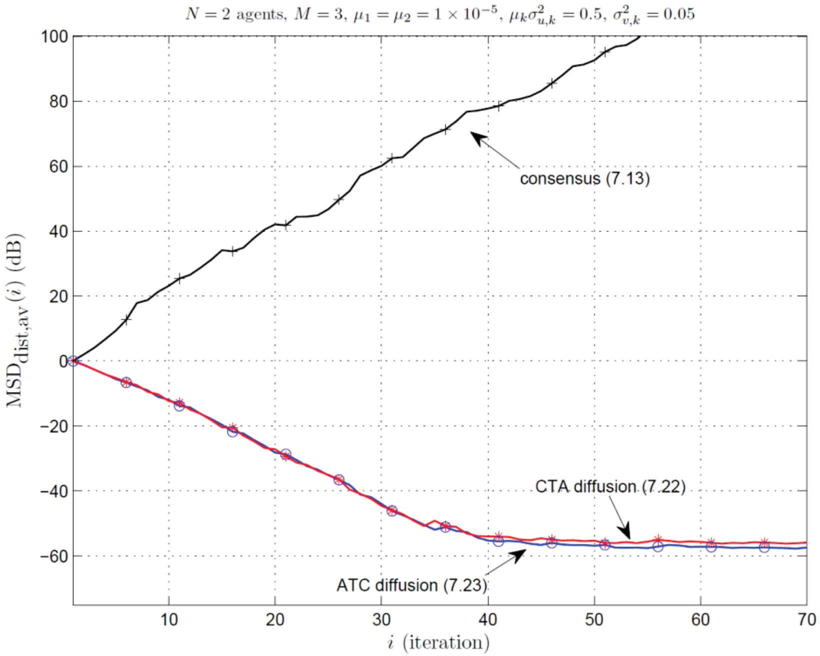


Figure 10.2: Evolution of the learning curves for the diffusion and consensus strategies for the numerical values $\mu_1 = \mu_2 = 1 \times 10^{-5}$, $\mu_1\sigma^2_{u,1} = \mu_2\sigma^2_{u,2} = 0.5$, and $(a, b) = (0.8, 0.8)$. These numerical values satisfy (10.135) for which the consensus solution becomes unstable.

Example #10.2

Figure 10.2 illustrates these results for the two-agent MSE network of Figure 10.1 dealing with complex-valued data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$ satisfying the model $\boldsymbol{d}_k(i) = \boldsymbol{u}_{k,i} w^o + \boldsymbol{v}_k(i)$ with $M = 3$. The unknown vector w^o is generated randomly and its norm is normalized to one. The figure plots the evolution of the ensemble-average learning curves, $\frac{1}{2}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$, for consensus, ATC diffusion, and CTA diffusion using $\mu_1 = \mu_2 = 1 \times 10^{-5}$. The measure $\frac{1}{2}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2$ corresponds to the average mean-square-deviation (MSD) of the agents at time i since

$$\frac{1}{2}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_i\|^2 = \frac{1}{2}\left(\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{1,i}\|^2 + \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{2,i}\|^2\right) \tag{10.141}$$

and $\widetilde{\boldsymbol{w}}_{k,i} = w^o - \boldsymbol{w}_{k,i}$. The learning curves are obtained by averaging the

Example #10.2

trajectories $\{\frac{1}{2}\|\widetilde{\boldsymbol{w}}_i\|^2\}$ over 100 repeated experiments. The label on the vertical axis in the figure refers to the learning curves $\frac{1}{2}\mathbb{E}\|\widetilde{\boldsymbol{w}}_i\|^2$ by writing $\text{MSD}_{\text{dist,av}}(i)$, with an iteration index i and where the subscripts "dist" and "av" are meant to indicate that this is an average performance measure for the distributed solution. Each experiment in this simulation involves running the consensus (7.13) or diffusion (7.22)–(7.23) LMS recursions with $h = 2$ on the complex-valued data $\{\boldsymbol{d}_k(i), \boldsymbol{u}_{k,i}\}$. The simulations use $\sigma_{v,1}^2 = \sigma_{v,2}^2 = 0.05$, $\mu_1\sigma_{u,1}^2 = \mu_2\sigma_{u,2}^2 = 0.5$, and $(a, b) = (0.8, 0.8)$. These numerical values ensure that (10.134) and (10.135) are satisfied so that the individual agents and the diffusion strategy are both mean stable, while the consensus strategy becomes unstable in the mean.

Example #10.2

The small step-sizes ensure that the networks are mean-square stable. It is seen in the figure that the learning curve of the consensus strategy grows unbounded while the learning curve of the diffusion strategies tend towards steady-state values.

Next, we consider an example satisfying

$$0 < \mu_1 \sigma_{u,1}^2 < 2 \leq \mu_2 \sigma_{u,2}^2 \tag{10.142}$$

so that, for the non-cooperative mode of operation, agent 1 is still stable while agent 2 is unstable. From the first equality of (10.137), we again conclude that

Example #10.2

$$
\begin{aligned}
\lambda_{\min}(\mathcal{B}_{\text{cons}}) &\leq \frac{1}{2}\left((2-a-b-\mu_1\sigma_{u,1}^2-\mu_2\sigma_{u,2}^2)-|b-a-\mu_1\sigma_{u,1}^2+\mu_2\sigma_{u,2}^2|\right)\\
&= \begin{cases} 1-a-\mu_1\sigma_{u,1}^2, & \text{if } b+\mu_2\sigma_{u,2}^2 \leq a+\mu_1\sigma_{u,1}^2\\ 1-b-\mu_2\sigma_{u,2}^2, & \text{otherwise}\end{cases}\\
&\leq 1-b-\mu_2\sigma_{u,2}^2\\
&\leq 1-\mu_2\sigma_{u,2}^2\\
&\leq -1 \qquad (10.143)
\end{aligned}
$$

Example #10.2

That is, in this second case, no matter how we choose the parameters $\{a, b\}$, the consensus network is always unstable. In contrast, the diffusion network is able to stabilize the network, i.e., there are choices for $\{a, b\}$ that lead to stable behavior. To see this, we set $b = 1 - a$ so that the eigenvalues of $\mathcal{B}_{\text{atc}}$ are

$$\lambda\left(\mathcal{B}_{\text{atc}}\right) \in \{0,\ 1 - \mu_1\sigma_{u,1}^2 - (\mu_2\sigma_{u,2}^2 - \mu_1\sigma_{u,1}^2)a\} \tag{10.144}$$

Some straightforward algebra shows that the magnitude of the nonzero eigenvalue will be bounded by one and, hence, the diffusion network will be stable in the mean if a satisfies:

$$0 \leq a < \frac{2 - \mu_1\sigma_{u,1}^2}{\mu_2\sigma_{u,2}^2 - \mu_1\sigma_{u,1}^2} \tag{10.145}$$

∎

End of Lecture

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.