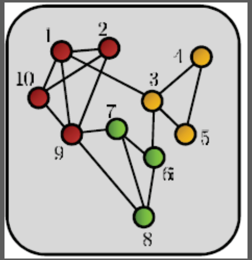
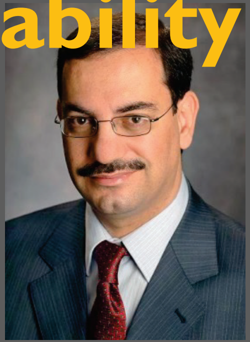


INFERENCE OVER NETWORKS

LECTURE #18: Mean Error Network Stability



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 9 (Stability of Multi-Agent Networks, pp. 507-551):</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.

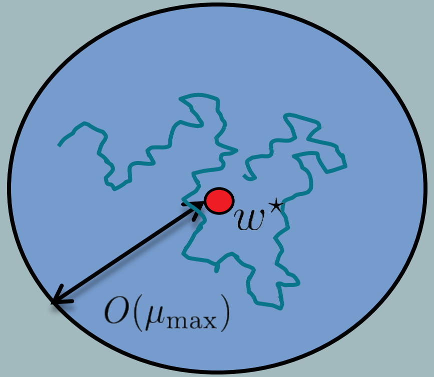
Network Stability

Theorems 9.1, 9.2, 9.6: For sufficiently small step-sizes:

$$\limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}\| = O(\mu_{\max})$$

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2 = O(\mu_{\max})$$

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^4 = O(\mu_{\max}^2)$$



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

Using the fact that $(\mathbb{E}\,\boldsymbol{a})^2 \leq \mathbb{E}\,\boldsymbol{a}^2$ for any real-valued random variable $\boldsymbol{a}$, we can readily conclude from (9.11), by using $\boldsymbol{a} = \|\widetilde{\boldsymbol{w}}_{k,i}\|$, that

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\| \;=\; O(\mu_{\max}^{1/2}), \quad k = 1, 2, \ldots, N \tag{9.158}$$

so that the first-order moment of the error vector tends to a bounded region in the order of $O(\mu_{\max}^{1/2})$. However, a smaller upper bound on $\|\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}\|$ can be derived with $O(\mu_{\max}^{1/2})$ replaced by $O(\mu_{\max})$, as shown in (9.1) and as we proceed to verify in this section. To do so, we examine the evolution of the mean-error vector more closely.

First-Order Error Moment

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

$$\widetilde{\boldsymbol{w}}_i^e = \boldsymbol{\mathcal{B}}_{i-1}\widetilde{\boldsymbol{w}}_{i-1}^e + \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{s}_i^e(\boldsymbol{w}_{i-1}^e) - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e, \quad i \geq 0 \qquad (9.159)$$

where, from the expressions in Lemma 8.1:

$$\boldsymbol{\mathcal{B}}_{i-1} = \mathcal{P}^{\mathsf{T}} - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\boldsymbol{\mathcal{H}}_{i-1}\mathcal{A}_1^{\mathsf{T}} \qquad (9.160)$$

$$\mathcal{P}^{\mathsf{T}} = \mathcal{A}_2^{\mathsf{T}}\mathcal{A}_o^{\mathsf{T}}\mathcal{A}_1^{\mathsf{T}} \qquad (9.161)$$

$$\boldsymbol{\mathcal{H}}_{i-1} \stackrel{\Delta}{=} \text{diag}\{\, \boldsymbol{H}_{1,i-1},\ \boldsymbol{H}_{2,i-1},\ \ldots,\ \boldsymbol{H}_{N,i-1}\,\} \qquad (9.162)$$

$$\boldsymbol{H}_{k,i-1} \stackrel{\Delta}{=} \int_0^1 \nabla_w^2 J_k(w^\star - t\widetilde{\boldsymbol{\phi}}_{k,i-1})dt \qquad (9.163)$$

First-Order Error Moment

Conditioning both sides of (9.159) 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:

$$\mathbb{E}\left[\,\widetilde{\boldsymbol{w}}_i^e\,|\,\boldsymbol{\mathcal{F}}_{i-1}\right] \;=\; \boldsymbol{\mathcal{B}}_{i-1}\widetilde{\boldsymbol{w}}_{i-1}^e \;-\; \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \qquad (9.164)$$

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

$$\mathbb{E}\,\widetilde{\boldsymbol{w}}_i^e \;=\; \mathbb{E}\left[\boldsymbol{\mathcal{B}}_{i-1}\widetilde{\boldsymbol{w}}_{i-1}^e\right] \;-\; \mathcal{A}_2^{\mathsf{T}}\mathcal{M}b^e \qquad (9.165)$$

First-Order Error Moment

Let

$$\widetilde{\boldsymbol{\mathcal{H}}}_{i-1} \stackrel{\Delta}{=} \mathcal{H} - \boldsymbol{\mathcal{H}}_{i-1} \tag{9.166}$$

where, in a manner similar to (9.162), we define the constant matrix

$$\mathcal{H} \stackrel{\Delta}{=} \text{diag}\{ H_1, H_2, \ldots, H_N \} \tag{9.167}$$

with each $H_{k,i-1}$ given by the value of the Hessian matrix at the limit point defined by (8.55), namely,

$$H_k \stackrel{\Delta}{=} \nabla_w^2 J_k(w^\star) \tag{9.168}$$

First-Order Error Moment

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

$$\begin{aligned} \boldsymbol{\mathcal{B}}_{i-1} &= \mathcal{P}^{\mathsf{T}} - \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \mathcal{H} \mathcal{A}_1^{\mathsf{T}} + \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \widetilde{\boldsymbol{\mathcal{H}}}_{i-1} \mathcal{A}_1^{\mathsf{T}} \\ &\triangleq \mathcal{B} + \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \widetilde{\boldsymbol{\mathcal{H}}}_{i-1} \mathcal{A}_1^{\mathsf{T}} \end{aligned} \qquad (9.169)$$

in terms of the constant coefficient matrix

$$\mathcal{B} \triangleq \mathcal{P}^{\mathsf{T}} - \mathcal{A}_2^{\mathsf{T}} \mathcal{M} \mathcal{H} \mathcal{A}_1^{\mathsf{T}} \qquad (9.170)$$

In this way, the mean-error relation (9.165) becomes

$$\mathbb{E}\, \widetilde{\boldsymbol{w}}_i^e = \mathcal{B} \left(\mathbb{E}\, \widetilde{\boldsymbol{w}}_{i-1}^e \right) - \mathcal{A}_2^{\mathsf{T}} \mathcal{M} b^e + \mathcal{A}_2^{\mathsf{T}} \mathcal{M} c_{i-1} \qquad (9.171)$$

First-Order Error Moment

in terms of a deterministic perturbation sequence defined by

$$c_{i-1} \triangleq \mathbb{E}\left(\widetilde{\boldsymbol{\mathcal{H}}}_{i-1}\mathcal{A}_1^{\mathsf{T}}\widetilde{\boldsymbol{w}}_{i-1}^{e}\right) \tag{9.172}$$

The constant matrix $\mathcal{B}$ defined by (9.170), and which drives the mean-error recursion (9.171), will play a critical role in characterizing the performance of multi-agent networks in future chapters. It also plays an important role in characterizing the mean-error stability of the network in this section. We therefore establish several important properties for $\mathcal{B}$ and subsequently use these properties to establish result (9.1) later in Theorem 9.6.

Stability of Coefficient Matrix

Theorem 9.3 (Stability of $\mathcal{B}$). 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) satisfies condition (6.13) in Assumption 6.1. Then, the constant matrix $\mathcal{B}$ defined by (9.170) is stable for sufficiently small step-sizes and its spectral radius is given by

$$\rho(\mathcal{B}) \;=\; 1 - \lambda_{\min}\left(\sum_{k=1}^{N} q_k H_k\right) \;+\; O\left(\mu_{\max}^{(N+1)/N}\right) \qquad (9.173)$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of its Hermitian matrix argument.

Proof

Proof. We first establish the result for diffusion and consensus networks and then extend the conclusion to the general distributed structure (8.46) with three combination matrices $\{A_1, A_o, A_2\}$. The arguments used in steps (a) and (b) below are justified when all step-sizes in $\mathcal{M}$ are strictly positive, which is the situation under study. The more general argument under step (c) below is applicable even to situations where some of the step-sizes are zero (a scenario we shall encounter later in Chapter 13).

Proof

(a) Diffusion strategies. For the case of diffusion strategies, the stability argument follows directly by examining the expression for the matrix $\mathcal{B}$. Recall that different choices for $\{A_o, A_1, A_2\}$ correspond to different strategies, as already shown by (8.7)–(8.10). In particular, for ATC and CTA diffusion, we set $A_1 = A$ or $A_2 = A$, for some left-stochastic matrix A, and the matrix A_o disappears from $\mathcal{B}$ since $A_o = I_N$ for these strategies. Specifically, the expression for $\mathcal{B}$ becomes

$$\mathcal{B}_{\rm atc} = \mathcal{A}^{\sf T}\left(I_{2MN} - \mathcal{M}\mathcal{H}\right) \tag{9.174}$$

$$\mathcal{B}_{\rm cta} = \left(I_{2MN} - \mathcal{M}\mathcal{H}\right)\mathcal{A}^{\sf T} \tag{9.175}$$

Proof

where $\mathcal{A} = A \otimes I_{2M}$ is left-stochastic and

$$\mathcal{M} \triangleq \text{diag}\{ \mu_1 I_{2M},\ \mu_2 I_{2M},\ \ldots,\ \mu_N I_{2M} \} \tag{9.176}$$

$$\mathcal{H} \triangleq \text{diag}\{ H_1,\ H_2,\ \ldots,\ H_N \} \tag{9.177}$$

The important fact to note from (9.174) and (9.175) is that the combination matrix $\mathcal{A}^{\mathsf{T}}$ appears *multiplying* (from left or right) the block diagonal matrix $I_{2MN} - \mathcal{M}\mathcal{H}$. We can then immediately call upon result (F.24) from the appendix, and employ the block maximum norm with blocks of size $2M \times 2M$ each, to conclude that

Proof

$$\rho\left(\mathcal{B}_{\text{atc}}\right) \;\leq\; \rho\left(I_{2MN} - \mathcal{M}\mathcal{H}\right) \tag{9.178}$$

$$\rho\left(\mathcal{B}_{\text{cta}}\right) \;\leq\; \rho\left(I_{2MN} - \mathcal{M}\mathcal{H}\right) \tag{9.179}$$

Therefore, for both cases of ATC and CTA diffusion, the respective coefficient matrices $\mathcal{B}$ become stable whenever the block-diagonal matrix $I_{2MN} - \mathcal{M}\mathcal{H}$ is stable. It is easily seen that this latter condition is guaranteed for step-sizes μ_k satisfying

$$\mu_k \;<\; \frac{2}{\rho(H_k)}, \quad k = 1, 2, \ldots, N \tag{9.180}$$

from which we conclude that sufficiently small step-sizes stabilize $\mathcal{B}_{\text{atc}}$ or $\mathcal{B}_{\text{cta}}$.

Recall#1: Weyl's Theorem

Weyl's Theorem [113, 260], shows how the eigenvalues of a Hermitian matrix are disturbed through additive perturbations to the entries of the matrix. Thus, let $\{A', A, \Delta A\}$ denote arbitrary $N \times N$ Hermitian matrices with ordered eigenvalues $\{\lambda_m(A'), \lambda_m(A), \lambda_m(\Delta A)\}$, i.e.,

$$\lambda_1(A) \;\geq\; \lambda_2(A) \;\geq\; \ldots \;\geq\; \lambda_N(A) \qquad \text{(F.31)}$$

Recall#1: Weyl's Theorem

and similarly for the eigenvalues of $\{A', \Delta A\}$, with the subscripts 1 and N representing the largest and smallest eigenvalues, respectively. Weyl's Theorem states that if A is perturbed to

$$A' = A + \Delta A \tag{F.32}$$

then the eigenvalues of the new matrix are bounded as follows:

$$\lambda_n(A) + \lambda_N(\Delta A) \leq \lambda_n(A') \leq \lambda_n(A) + \lambda_1(\Delta A) \tag{F.33}$$

Continuing with Proof

(b) Consensus strategy. For the consensus strategy, we set $A_1 = A_2 = I_N$ and $A_o = A$. In this case, the expression for $\mathcal{B}$ becomes

$$\mathcal{B}_{\text{cons}} \quad = \quad \mathcal{A}^{\mathsf{T}} - \mathcal{M}\mathcal{H} \tag{9.181}$$

where $\mathcal{A}$ now appears as an additive term. A condition on the step-sizes to ensure the stability of $\mathcal{B}_{\text{cons}}$ can be deduced from Weyl's Theorem (F.33) in the appendix if we additionally assume that the left-stochastic matrix A is *symmetric* [248], in which case it will also be doubly stochastic. Since A is then both symmetric and left-stochastic, its eigenvalues will be real and lie inside the interval $[-1, 1]$. Hence, $(I_{2MN} - \mathcal{A}^{\mathsf{T}}) \geq 0$. Moreover, since the matrices

smallest eigenvalue is zero

Proof

$\mathcal{M}$ and $\mathcal{H}$ are block-diagonal Hermitian and commute with each other, i.e., $\mathcal{H}\mathcal{M} = \mathcal{M}\mathcal{H}$, it follows that $\mathcal{B}_{\rm cons}$ in (9.181) is Hermitian, as well as the matrix $\mathcal{B}_{\rm ncop} = I_{2MN} - \mathcal{M}\mathcal{H}$. Now note that we can write the following two trivial equalities (by adding and subtracting equal terms):

$$\mathcal{B}_{\rm ncop} = \mathcal{B}_{\rm cons} + (I_{2MN} - \mathcal{A}^{\mathsf{T}}) \tag{9.182}$$

$$\mathcal{B}_{\rm cons} = (\lambda_{\min}(A) \cdot I_{2MN} - \mathcal{M}\mathcal{H}) + (\mathcal{A}^{\mathsf{T}} - \lambda_{\min}(A) \cdot I_{2MN}) \tag{9.183}$$

so that by applying Weyl's Theorem (F.33) to both representations, we obtain the following eigenvalue relations:

smallest eigenvalue is zero

Proof

$$\lambda_\ell(\mathcal{B}_{\rm cons}) \;\leq\; \lambda_\ell(\mathcal{B}_{\rm ncop}) \tag{9.184}$$

$$\lambda_\ell(\mathcal{B}_{\rm cons}) \;\geq\; \lambda_\ell\left\{\lambda_{\min}(A)\cdot I_{2MN} - \mathcal{M}\mathcal{H}\right\} \tag{9.185}$$

for $\ell = 1, 2, \ldots, 2MN$ and where we are assuming ordered eigenvalues, namely, $\lambda_1 \geq \lambda_2 \geq \ldots$, for any of the matrix arguments. It follows that the matrix $\mathcal{B}_{\rm cons}$ will be stable, namely, $-1 < \lambda_\ell(\mathcal{B}_{\rm cons}) < 1$ for all ℓ if

$$\lambda_\ell(\mathcal{B}_{\rm ncop}) \;<\; 1 \tag{9.186}$$

$$\lambda_\ell\left\{\lambda_{\min}(A)\cdot I_{2MN} - \mathcal{M}\mathcal{H}\right\} \;>\; -1 \tag{9.187}$$

Proof

The first condition is automatically satisfied due to the form of the matrix $\mathcal{B}_{\rm ncop}$ and since $\mathcal{M}\mathcal{H} > 0$. For the second condition, it will be satisfied by step-sizes $\{\mu_k\}$ such that

$$\mu_k < \frac{1 + \lambda_{\min}(A)}{\rho(H_k)}, \quad k = 1, 2, \ldots, N \tag{9.188}$$

Since we are dealing with strongly-connected networks, the matrix A is primitive and, therefore, it has a single eigenvalue matching its spectral radius, which is equal to one. That eigenvalue occurs at $+1$ so that

Proof

$\lambda_{\min}(A) > -1$ and the upper bound in (9.188) is positive. We therefore conclude that sufficiently small step-sizes stabilize $\mathcal{B}$ for consensus strategies with a *symmetric* combination policy A. If A is not symmetric, then the next argument would apply to this case.

Proof

(c) General case (eigenvalue perturbation analysis). For the general case, when the matrix A_o is not necessarily the identity matrix or symmetric, and when all three matrices $\{A_o, A_1, A_2\}$ or subsets thereof may be present, the argument is more demanding. The argument that follows is based on an eigenvalue perturbation analysis in the small step-size regime similar to [277]. We establish the result for the general case of complex data and, therefore, $h = 2$ throughout this derivation.

Proof

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

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

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

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

$$\begin{aligned}
D_{11} &= \sum_{k=1}^{N} q_k H_k^{\mathsf{T}} && (9.195)\\
D_{12} &= (\mathbb{1}^{\mathsf{T}} \otimes I_{2M})\mathcal{H}^{\mathsf{T}}\mathcal{M}(A_2 V_R \otimes I_{2M}) && (9.196)\\
D_{21} &= (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}^{\mathsf{T}}(q \otimes I_{2M}) && (9.197)\\
D_{22} &= (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}^{\mathsf{T}}\mathcal{M}(A_2 V_R \otimes I_{2M}) && (9.198)
\end{aligned}$$

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

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

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

$$\mathcal{B} \sim \begin{bmatrix} I_{2M} - O(\mu_{\max}) & O(\mu_{\max}) \\ O(\mu_{\max}) & \mathcal{J}_\epsilon^\mathsf{T} + O(\mu_{\max}) \end{bmatrix} \tag{9.207}$$

Now recall that J_ϵ is $(N-1) \times (N-1)$ and has a Jordan structure. For ease of presentation, and without any loss of generality, let us assume that J_ϵ consists of two Jordan blocks, say, as

$$J_\epsilon = \left[\begin{array}{cc|ccc} \lambda_a & & & & \\ \epsilon & \lambda_a & & & \\ \hline & & \lambda_b & & \\ & & \epsilon & \lambda_b & \\ & & & \epsilon & \lambda_b \end{array}\right] \tag{9.208}$$

Proof

Then, the matrix $\mathcal{J}_\epsilon = J_\epsilon \otimes I_{2M}$ has dimensions $2M(N-1) \times 2M(N-1)$ and is given by

$$\mathcal{J}_\epsilon = J_\epsilon \otimes I_{2M} \left[\begin{array}{cc|ccc} \lambda_a I_{2M} & & & & \\ \epsilon I_{2M} & \lambda_a I_{2M} & & & \\ \hline & & \lambda_b I_{2M} & & \\ & & \epsilon I_{2M} & \lambda_b I_{2M} & \\ & & & \epsilon I_{2M} & \lambda_b I_{2M} \end{array}\right] \tag{9.209}$$

More generically, for multiple Jordan blocks, it is clear that we can express $\mathcal{J}_\epsilon$ in the following lower-triangular form:

Proof

$$\mathcal{J}_\epsilon = \begin{bmatrix} \lambda_{a,2} I_{2M} & & & \\ & \lambda_{a,3} I_{2M} & & \\ \mathcal{K} & & \ddots & \\ & & & \lambda_{a,L} I_{2M} \end{bmatrix} \qquad (9.210)$$

with scalars $\{\lambda_{a,\ell}\}$ on the diagonal, all of which have norms strictly less than one, and where the entries of the strictly lower-triangular matrix $\mathcal{K}$ are either ϵ or zero. In the above representation, we are assuming that J_ϵ consists of several Jordan blocks. It follows that

Proof

$$\mathcal{J}_\epsilon^\mathsf{T} + O(\mu_{\max}) = \begin{bmatrix} \lambda_{a,2} I_{2M} + O(\mu_{\max}) & & \mathcal{K}^\mathsf{T} + O(\mu_{\max}) \\ & \ddots & \\ O(\mu_{\max}) & & \lambda_{a,L} I_{2M} + O(\mu_{\max}) \end{bmatrix} \tag{9.211}$$

We introduce the eigen-decomposition of the Hermitian positive-definite matrix D_{11}^T and denote it by:

$$D_{11}^\mathsf{T} \triangleq U \Lambda U^* \tag{9.212}$$

where U is unitary and Λ has positive-diagonal entries $\{\lambda_k\}$; the matrices U and Λ are $2M \times 2M$. Using U, we further introduce the following block-diagonal similarity transformation:

Proof

$$\mathcal{T} \triangleq \text{diag}\left\{\mu_{\max}^{1/N} U,\ \mu_{\max}^{2/N} I_{2M},\ \ldots,\ \mu_{\max}^{(N-1)/N} I_{2M},\ \mu_{\max} I_{2M}\ \right\} \tag{9.213}$$

where all block entries are defined in terms of I_{2M}, except for the first entry defined in terms of U. We now use (9.205) to get

$$\mathcal{T}^{-1}\left(\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{B} \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}}\right) \mathcal{T}\ = \tag{9.214}$$

Proof

$$\left[\begin{array}{c|cc}
B & \multicolumn{2}{c}{O\left(\mu_{\max}^{(N+1)/N}\right)} \\
\hline
 & \lambda_{a,2} I_{2M} + O(\mu_{\max}) & O\left(\mu_{\max}^{1/N}\right) \\
O(\mu_{\max}^{1/N}) & & \ddots \\
 & O\left(\mu_{\max}^{1/N}\right) & \lambda_{a,L} I_{2M} + O(\mu_{\max})
\end{array}\right]$$

where we introduced the $2M \times 2M$ diagonal matrix

$$B \stackrel{\Delta}{=} I_{2M} - \Lambda \tag{9.215}$$

Recall#2: Gershgorin's Theorem

Gershgorin's Theorem [48, 94, 101, 104, 113, 254, 264], specifies circular regions within which the eigenvalues of a matrix are located. Thus, consider an $N \times N$ matrix A with scalar entries $\{a_{\ell k}\}$. With each diagonal entry $a_{\ell\ell}$ we associate a disc in the complex plane centered at $a_{\ell\ell}$ and with

$$r_\ell \triangleq \sum_{k \neq \ell, k=1}^{N} |a_{\ell k}| \tag{F.35}$$

Recall#2: Gershgorin's Theorem

That is, r_ℓ is equal to the sum of the magnitudes of the non-diagonal entries on the same row as $a_{\ell\ell}$. We denote the disc by D_ℓ; it consists of all points that satisfy

$$D_\ell = \left\{ z \in \mathbb{C}^N \text{ such that } |z - a_{\ell\ell}| \leq r_\ell \right\} \tag{F.36}$$

The theorem states that the spectrum of A (i.e., the set of all its eigenvalues, denoted by $\lambda(A)$) is contained in the union of all N Gershgorin discs:

$$\lambda(A) \subset \bigcup_{\ell=1}^{N} D_\ell \tag{F.37}$$

Recall#2: Gershgorin's Theorem

A stronger statement of the Gershgorin theorem covers the situation in which some of the Gershgorin discs happen to be disjoint. Specifically, if the union of L of the discs is disjoint from the union of the remaining $N - L$ discs, then the theorem further asserts that L eigenvalues of A will lie in the first union of L discs and the remaining $N - L$ eigenvalues of A will lie in the second union of $N - L$ discs.

Continuing with Proof

It follows from (9.214) that all off-diagonal entries of the above transformed matrix are at least $O(\mu_{\max}^{1/N})$. Although the factor $\mu_{\max}^{1/N}$ decays slower than $\mu_{\max}$, it nevertheless becomes small for sufficiently small $\mu_{\max}$. Then, calling upon Gershgorin's Theorem (F.37) from the appendix, we conclude from (9.214) that the eigenvalues of $\mathcal{B}$ are are either located in the Gershgorin circles that are centered at the eigenvalues of B with radii $O(\mu_{\max}^{(N+1)/N})$ or in the Gershgorin circles that are centered at the $\{\lambda_{a,\ell}\}$ with radii $O(\mu_{\max}^{1/N})$, namely,

Proof

$$|\lambda(\mathcal{B}) - \lambda(B)| \le O\left(\mu_{\max}^{(N+1)/N}\right) \quad \text{or} \quad |\lambda(\mathcal{B}) - \lambda_{a,\ell}| \le O\left(\mu_{\max}^{1/N}\right) \tag{9.216}$$

where $\lambda(\mathcal{B})$ and $\lambda(B)$ denote any of the eigenvalues of $\mathcal{B}$ and B, and $\ell = 2, \ldots, L$. It follows that

$$\rho(\mathcal{B}) \le \rho(B) + O\left(\mu_{\max}^{(N+1)/N}\right) \quad \text{or} \quad \rho(\mathcal{B}) \le \rho(J_\epsilon) + O(\mu_{\max}^{1/N}) \tag{9.217}$$

Now since J_ϵ is a stable matrix, we know that $\rho(J_\epsilon) < 1$. We express this spectral radius as

$$\rho(J_\epsilon) \;=\; 1 - \delta_J \tag{9.218}$$

Proof

where δ_J is positive and independent of $\mu_{\max}$. We also know from (9.203) that

$$\rho(B) = 1 - \sigma_{11}\mu_{\max} \;\; < 1 \tag{9.219}$$

since $B = U^*(I_{2M} - D_{11}^{\mathsf{T}})U$. We conclude from (9.217) that

$$\rho(\mathcal{B}) \leq 1 - \sigma_{11}\mu_{\max} + O\left(\mu_{\max}^{(N+1)/N}\right) \;\; \text{or} \;\; \rho(\mathcal{B}) \leq 1 - \delta_J + O(\mu_{\max}^{1/N}) \tag{9.220}$$

If we now select $\mu_{\max} \ll 1$ small enough such that

$$O\left(\mu_{\max}^{(N+1)/N}\right) < \sigma_{11}\mu_{\max} \quad \text{and} \quad O\left(\mu_{\max}^{1/N}\right) + O(\mu_{\max}) < \delta_J \tag{9.221}$$

Proof

then we would be able to conclude that $\rho(\mathcal{B}) < 1$ so that $\mathcal{B}$ is stable for sufficiently small step-sizes. Both conditions in (9.221) can be satisfied simultaneously and they will ensure

$$\rho(\mathcal{B}) = 1 - O(\mu_{\max}) \tag{9.222}$$

Proof

With regards to expression (9.173) for the spectral radius of $\mathcal{B}$, we call upon the stronger statement of Gershgorin's theorem mentioned after (F.37) in the appendix and which relates to how the eigenvalues of a matrix are distributed over disjoint Gershgorin sets. To begin with, note from (9.203) that for $\mu_{\max} \ll 1$, all eigenvalues of $B = I_{2M} - \Lambda$ are real-valued and positive. We then conclude from (9.222) that all eigenvalues of B lie inside the open interval

$$\lambda(B) \in (1 - O(\mu_{\max}), 1) \tag{9.223}$$

It further follows from this result that the eigenvalues of B are at most $O(\mu_{\max})$ apart from each other.

$B \sim I_{2M} - D_{11}^{\mathsf{T}}, \quad c_1\mu_{\max} \leq \lambda_\ell(D_{11}^{\mathsf{T}}) \leq c_2\mu_{\max}$

Proof

Now, referring to (9.216), the condition on the left describes a region in space that consists of the union of $2M$ Gershgorin discs: each disc is centered at one of the eigenvalues of B with radius $O(\mu_{\max}^{(N+1)/N})$. We can then choose $\mu_{\max}$ small enough such that the discs that are centered at distinct eigenvalues of B remain disjoint from each other. The union of these discs will be contained within the circle that is centered at one and with radius $O(\mu_{\max})$ — see the region described by the smaller circle on the right in Figure 9.1.

Proof

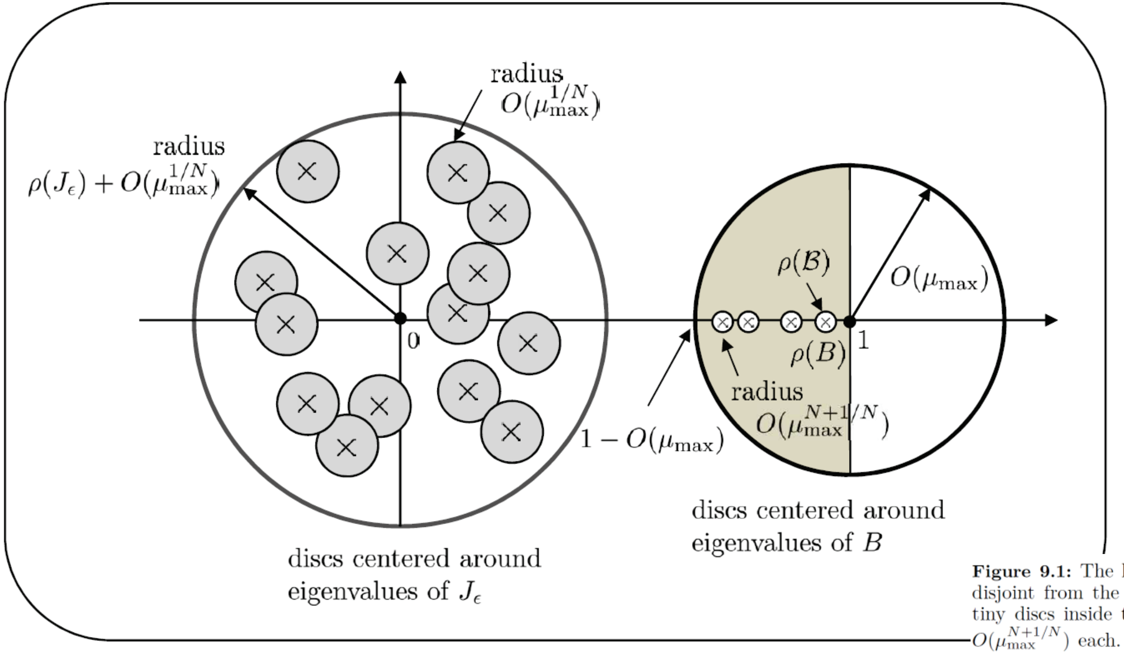


Figure 9.1: The larger circle on the left has radius $\rho(J_\epsilon) + O(\mu_{\max}^{1/N})$ and is disjoint from the smaller circle on the right whose radius is $O(\mu_{\max})$. The tiny discs inside the smaller circle on the right are disjoint and have radii $O(\mu_{\max}^{N+1/N})$ each. The eigenvalue corresponding to the spectral radius of $\mathcal{B}$ lies inside the rightmost smaller disc centered around $\rho(B)$.

Proof

Let us now examine the rightmost condition in (9.216). This condition describes a region in space that consists of the union of $2M(N-1)$ Gershgorin discs: each disc is now centered at an eigenvalue of $\mathcal{J}_\epsilon$ with radius $O(\mu_{\max}^{1/N})$. Therefore, again for $\mu_{\max} \ll 1$, the union of these discs is contained within a circle centered at the origin and with radius $\rho(J_\epsilon) + O(\mu_{\max}^{1/N})$; this radius is smaller than $1 - O(\mu_{\max})$ by virtue of the second condition in (9.221) — see the region described by the larger circle on the left in Figure 9.1. It follows that the two circular regions that we identified are disjoint from each other:

Proof

one region is determined by the circle on the left that is centered at the origin with radius smaller than $1 - O(\mu_{\max})$, while the other region is determined by the circle on the right that is centered at one and has radius $O(\mu_{\max})$. The $2M$ discs that appear within this smaller circle are disjoint from the discs that appear inside the larger circle on the left. We conclude that $2M$ of the eigenvalues of $\mathcal{B}$ are located inside the discs in the rightmost circle. The eigenvalue that attains the spectral radius of $\mathcal{B}$ occurs inside this region so that

$$\rho(\mathcal{B}) \quad = \quad \rho(B) + O\left(\mu_{\max}^{(N+1)/N}\right) \qquad (9.224)$$

Proof

Since it is assumed that $\mu_{\max} \ll 1$, and by referring back to expression (9.195) for D_{11}, we have

$$\rho(B) \;=\; \rho(I_{2M} - D_{11}^{\mathsf{T}}) \;=\; 1 - \lambda_{\min}\left(\sum_{k=1}^{N} q_k H_k\right) \qquad (9.225)$$

Combining this relation with (9.224), we arrive at (9.173).

$\square$

Size of Entries of $\mathcal{B}$

We can further exploit the structure revealed by expression (9.205) for $\mathcal{B}$ to examine the size of the entries of $(I - \mathcal{B})^{-1}$. In our derivations, the matrix $\mathcal{B}$ also appears transformed under the similarity transformation:

$$\bar{\mathcal{B}} \triangleq \mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{B} \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \overset{(9.206)}{=} \begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix} \qquad (9.226)$$

where, according to (9.204),

$$\mathcal{V}_\epsilon \triangleq V_\epsilon \otimes I_{hM} \qquad (9.227)$$

We therefore examine both matrices. The following result clarifies the size of the entries of $(I - \mathcal{B})^{-1}$ and $(I - \bar{\mathcal{B}})^{-1}$.

Size of Entries of $\mathcal{B}$

Lemma 9.4 (Similarity transformation). Assume the matrix P is primitive. It holds that for sufficiently small step-sizes:

$$(I - \mathcal{B})^{-1} = O(1/\mu_{\max}) \tag{9.228}$$

$$(I - \bar{\mathcal{B}})^{-1} = \left[\begin{array}{c|c} O(1/\mu_{\max}) & O(1) \\ \hline O(1) & O(1) \end{array}\right] \tag{9.229}$$

where the leading $(1,1)$ block in $(I - \bar{\mathcal{B}})^{-1}$ has dimensions $hM \times hM$.

Proof

Proof. We carry out the derivation for the complex case $h = 2$ without loss of generality following arguments similar to [69, 278]. We first remark that, by similarity, the matrix $\bar{\mathcal{B}}$ is stable by Theorem 9.3. Let

$$\begin{aligned} \mathcal{X} = I - \bar{\mathcal{B}} &= \begin{bmatrix} D_{11}^{\mathsf{T}} & D_{21}^{\mathsf{T}} \\ D_{12}^{\mathsf{T}} & I - \mathcal{J}_{\epsilon}^{\mathsf{T}} + D_{22}^{\mathsf{T}} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \mathcal{X}_{11} & \mathcal{X}_{12} \\ \mathcal{X}_{21} & \mathcal{X}_{22} \end{bmatrix} \end{aligned} \tag{9.230}$$

where, from (9.199)–(9.202),

Proof

$$\mathcal{X}_{11} = O(\mu_{\max}) \tag{9.231}$$

$$\mathcal{X}_{12} = O(\mu_{\max}) \tag{9.232}$$

$$\mathcal{X}_{21} = O(\mu_{\max}) \tag{9.233}$$

$$\mathcal{X}_{22} = O(1) \tag{9.234}$$

The matrix $\mathcal{X}$ is invertible since $I - \bar{\mathcal{B}}$ is invertible. Moreover, $\mathcal{X}_{11}$ is invertible since $D_{11} > 0$. We now appeal to the useful block matrix inversion formula [113, 206]:

Proof

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} A^{-1}B\Delta^{-1}CA^{-1} & -A^{-1}B\Delta^{-1} \\ -\Delta^{-1}CA^{-1} & \Delta^{-1} \end{bmatrix} \tag{9.235}$$

for matrices $\{A, B, C, D\}$ of compatible dimensions with invertible A and invertible Schur complement Δ defined by

$$\Delta = D - CA^{-1}B \tag{9.236}$$

Using this formula we can write

$$\mathcal{X}^{-1} = \begin{bmatrix} \mathcal{X}_{11}^{-1} + \mathcal{X}_{11}^{-1}\mathcal{X}_{12}\Delta^{-1}\mathcal{X}_{21}\mathcal{X}_{11}^{-1} & -\mathcal{X}_{11}^{-1}\mathcal{X}_{12}\Delta^{-1} \\ -\Delta^{-1}\mathcal{X}_{21}\mathcal{X}_{11}^{-1} & \Delta^{-1} \end{bmatrix} \tag{9.237}$$

Proof

where Δ denotes the Schur complement of $\mathcal{X}$ relative to $\mathcal{X}_{11}$:

$$\Delta \stackrel{\Delta}{=} \mathcal{X}_{22} - \mathcal{X}_{21}\mathcal{X}_{11}^{-1}\mathcal{X}_{12} \; = \; O(1) \tag{9.238}$$

We then use (9.231)–(9.234) and (9.238) to deduce that

$$\mathcal{X}^{-1} \; = \; \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(1) & O(1) \end{bmatrix} \tag{9.239}$$

as claimed.

$\square$

Recall#3: Block Kronecker Products

$$\mathcal{K} \triangleq \mathcal{A} \otimes_b \mathcal{B} = \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1n} \\ K_{21} & K_{22} & \dots & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \dots & K_{nn} \end{bmatrix} \tag{F.2}$$

where each block K_{ij} is $mp^2 \times mp^2$ and is constructed as follows:

Recall#3: Block Kronecker Products

$$K_{ij} = \begin{bmatrix} A_{ij} \otimes B_{11} & A_{ij} \otimes B_{12} & \dots & A_{ij} \otimes B_{1m} \\ A_{ij} \otimes B_{21} & A_{ij} \otimes B_{22} & \dots & A_{ij} \otimes B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{ij} \otimes B_{m1} & A_{ij} \otimes B_{m2} & \dots & A_{ij} \otimes B_{mm} \end{bmatrix} \tag{F.3}$$

Recall#3: Block Kronecker Products

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.	$\mathrm{bvec}(xy^{\mathsf{T}}) = y \otimes_b x$

Low-Rank Approximation

We can establish similar results for the matrix

$$\mathcal{F} \triangleq \mathcal{B}^{\mathsf{T}} \otimes_b \mathcal{B}^* \tag{9.240}$$

which is defined in terms of the block Kronecker product operation using blocks of size $hM \times hM$, where $h = 1$ for real data and $h = 2$ for complex data. The matrix $\mathcal{F}$ will play a critical role in characterizing the performance and convergence rate of distributed algorithms, as will be revealed by future Theorem 11.2. In our derivations, the matrix $\mathcal{F}$ will also sometimes appear transformed under the similarity transformation:

$$\bar{\mathcal{F}} \triangleq (\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon)^{-1} \mathcal{F} (\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon) \tag{9.241}$$

Low-Rank Approximation

Lemma 9.5 (Low-rank approximation). 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 $O(1/\mu_{\max})$. Moreover, we can also write

$$(I-\mathcal{F})^{-1} = \left[(p\otimes p)(\mathbb{1}\otimes\mathbb{1})^{\mathsf{T}}\right]\otimes Z^{-1} + O(1) \tag{9.244}$$

Second-Order Error Moment

in terms of the regular Kronecker product operation, where the matrix Z has dimensions $(hM)^2 \times (hM)^2$ and consists of blocks of size $hM \times hM$ each:

$$Z \triangleq \sum_{k=1}^{N} q_k \left[(I_{hM} \otimes H_k) + (H_k^{\mathsf{T}} \otimes I_{hM}) \right] \qquad (9.245)$$

where the vectors $\{p, q\}$ were defined earlier by (9.7)–(9.9). In addition, $Z = O(\mu_{\max})$.

Proof

Proof. We again carry out the derivation for the complex case $h = 2$ without loss of generality by extending an argument from [278] to the current context. We recall from (9.170) the expression for $\mathcal{B}$:

$$\mathcal{B} \;=\; \mathcal{P}^{\mathsf{T}} - \mathcal{A}_2^{\mathsf{T}}\mathcal{M}\mathcal{R}\mathcal{A}_1^{\mathsf{T}} \;=\; \mathcal{A}_2^{\mathsf{T}}\left(\mathcal{A}_o^{\mathsf{T}} - \mathcal{M}\mathcal{H}\right)\mathcal{A}_1^{\mathsf{T}} \tag{9.246}$$

where $\mathcal{P} = P \otimes I_{2M}$ and $P = A_1 A_o A_2$. Since the matrices $\{\mathcal{A}_o, \mathcal{A}_1, \mathcal{A}_2, \mathcal{M}\}$ are real-valued, and $\mathcal{H}$ is Hermitian, we have

$$\mathcal{B}^{\mathsf{T}} \;=\; \mathcal{A}_1(\mathcal{A}_o - \mathcal{H}^{\mathsf{T}}\mathcal{M})\mathcal{A}_2 \tag{9.247}$$

$$\mathcal{B}^{*} \;=\; \mathcal{A}_1(\mathcal{A}_o - \mathcal{H}\mathcal{M})\mathcal{A}_2 \tag{9.248}$$

Proof

We introduce the same Jordan canonical decomposition (9.21)–(9.24) and verify, in a manner similar to (9.53), that

$$\mathcal{B}^* \quad = \quad (V_\epsilon \otimes I_{2M}) \begin{bmatrix} I_{2M} - E_{11} & -E_{12} \\ -E_{21} & (J_\epsilon \otimes I_{2M}) - E_{22} \end{bmatrix} (V_\epsilon^{-1} \otimes I_{2M}) \tag{9.249}$$

where the block matrices $\{E_{mn}\}$ are given by

Proof

$$E_{11} = \sum_{k=1}^{N} q_k H_k = O(\mu_{\max}) \tag{9.250}$$

$$E_{12} = (\mathbb{1}^{\mathsf{T}} \otimes I_{2M})\mathcal{H}\mathcal{M}(A_2 V_R \otimes I_{2M}) = O(\mu_{\max}) \tag{9.251}$$

$$E_{21} = (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}(q \otimes I_{2M}) = O(\mu_{\max}) \tag{9.252}$$

$$E_{22} = (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}\mathcal{M}(A_2 V_R \otimes I_{2M}) = O(\mu_{\max}) \tag{9.253}$$

Proof

and their entries are in the order of $\mu_{\max}$; this fact can be verified in the same manner that we assessed the size of the block matrices $\{\boldsymbol{D}_{11,i-1}, \boldsymbol{D}_{12,i-1}, \boldsymbol{D}_{21,i-1}, \boldsymbol{D}_{22,i-1}\}$ in the proof of the earlier Theorem 9.1. Moreover, the dimensions of E_{11} are $2M \times 2M$.

In a similar manner, we find that

$$\mathcal{B}^{\mathsf{T}} = (V_\epsilon \otimes I_{2M}) \begin{bmatrix} I_{2M} - D_{11} & -D_{12} \\ -D_{21} & (J_\epsilon \otimes I_{2M}) - D_{22} \end{bmatrix} (V_\epsilon^{-1} \otimes I_{2M}) \tag{9.254}$$

Proof

where the block matrices $\{D_{mn}\}$ are given by

$$
\begin{aligned}
D_{11} &= \sum_{k=1}^{N} q_k H_k^{\mathsf{T}} = O(\mu_{\max}) && (9.255)\\
D_{12} &= (\mathbb{1}^{\mathsf{T}} \otimes I_{2M})\mathcal{H}^{\mathsf{T}}\mathcal{M}(A_2 V_R \otimes I_{2M}) = O(\mu_{\max}) && (9.256)\\
D_{21} &= (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}^{\mathsf{T}}(q \otimes I_{2M}) = O(\mu_{\max}) && (9.257)\\
D_{22} &= (V_L^{\mathsf{T}} A_1 \otimes I_{2M})\mathcal{H}^{\mathsf{T}}\mathcal{M}(A_2 V_R \otimes I_{2M}) = O(\mu_{\max}) && (9.258)
\end{aligned}
$$

Proof

and D_{11} has dimensions $2M \times 2M$. Substituting expressions (9.249) and (9.254) into (9.240), and using the second property for block Kronecker products from Table F.2 in the appendix, we obtain

$$\mathcal{F} \;=\; (\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon)\, \mathcal{X}\, (\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon)^{-1} \tag{9.259}$$

where the block Kronecker product operation is relative to blocks of size $2M \times 2M$, and where we introduced

$$\mathcal{X} \triangleq \begin{bmatrix} I_{2M} - D_{11} & -D_{12} \\ -D_{21} & (J_\epsilon \otimes I_{2M}) - D_{22} \end{bmatrix} \otimes_b \begin{bmatrix} I_{2M} - E_{11} & -E_{12} \\ -E_{21} & (J_\epsilon \otimes I_{2M}) - E_{22} \end{bmatrix} \tag{9.260}$$

Proof

We conclude that

$$(I - \mathcal{F})^{-1} \;=\; (\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon)\,(I - \mathcal{X})^{-1}\,(\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon)^{-1} \tag{9.261}$$

We partition $\mathcal{X}$ into the following block structure:

$$\mathcal{X} \;=\; \begin{bmatrix} \mathcal{X}_{11} & \mathcal{X}_{12} \\ \mathcal{X}_{21} & \mathcal{X}_{22} \end{bmatrix} \tag{9.262}$$

where, for example, $\mathcal{X}_{11}$ is $(2M)^2 \times (2M)^2$ and is given by

$$\mathcal{X}_{11} = (I_{2M} - D_{11}) \otimes (I_{2M} - E_{11}) \tag{9.263}$$

Proof

It follows that

$$I - \mathcal{X} = \begin{bmatrix} I_{(2M)^2} - \mathcal{X}_{11} & -\mathcal{X}_{12} \\ -\mathcal{X}_{21} & I - \mathcal{X}_{22} \end{bmatrix} \tag{9.264}$$

and, in a manner similar to the way we assessed the size of the block matrices $\{\boldsymbol{D}_{11,i-1}, \boldsymbol{D}_{12,i-1}, \boldsymbol{D}_{21,i-1}, \boldsymbol{D}_{22,i-1}\}$ in the proof of Theorem 9.1, we can likewise verify that

$$I_{(2M)^2} - \mathcal{X}_{11} = O(\mu_{\max}) \tag{9.265}$$

$$\mathcal{X}_{12} = O(\mu_{\max}) \tag{9.266}$$

$$\mathcal{X}_{21} = O(\mu_{\max}) \tag{9.267}$$

$$I - \mathcal{X}_{22} = O(1) \tag{9.268}$$

Proof

In particular, note that

$$\begin{aligned} I_{(2M)^2} - \mathcal{X}_{11} &= I_{(2M)^2} - (I_{2M} - D_{11}) \otimes (I_{2M} - E_{11}) \\ &= (I_{2M} \otimes E_{11}) + (D_{11} \otimes I_{2M}) - (D_{11} \otimes E_{11}) \\ &= O(\mu_{\max}) \end{aligned} \tag{9.269}$$

and

$$\begin{aligned} I - \mathcal{X}_{22} &= I - ((J_\epsilon \otimes I_{2M}) - D_{22}) \otimes_b ((J_\epsilon \otimes I_{2M}) - E_{22}) \\ &= I - (J_\epsilon \otimes I_{2M}) \otimes_b (J_\epsilon \otimes I_{2M}) + O(\mu_{\max}) \\ &= O(1) \end{aligned} \tag{9.270}$$

Proof

To proceed, we call again upon the useful block matrix inversion formula (9.235). The matrix $I - \mathcal{X}$ is invertible since $I - \mathcal{F}$ is invertible; this is because $\rho(\mathcal{F}) = [\rho(\mathcal{B})]^2 < 1$. Therefore, applying (9.235) to $I - \mathcal{X}$ we get

$$(I-\mathcal{X})^{-1} = \begin{bmatrix} (I_{(2M)^2} - \mathcal{X}_{11})^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \qquad (9.271)$$

$$\begin{bmatrix} (I-\mathcal{X}_{11})^{-1}\mathcal{X}_{12}\Delta^{-1}\mathcal{X}_{21}(I-\mathcal{X}_{11})^{-1} & (I-\mathcal{X}_{11})^{-1}\mathcal{X}_{12}\Delta^{-1} \\ \Delta^{-1}\mathcal{X}_{21}(I-\mathcal{X}_{11})^{-1} & \Delta^{-1} \end{bmatrix}$$

Proof

It is seen from (9.269) that the entries of $(I - \mathcal{X}_{11})^{-1}$ are $O(1/\mu_{\max})$, while the entries in the second matrix on the right-hand side of equality (9.271) are $O(1)$ when the step-sizes are small. That is, we can write

$$(I - \mathcal{X})^{-1} = \left[\begin{array}{c|c} O(1/\mu_{\max}) & O(1) \\ \hline O(1) & O(1) \end{array}\right] \tag{9.272}$$

where the leading $(2M)^2 \times (2M)^2$ block is $O(1/\mu_{\max})$. Moreover, since $O(1/\mu_{\max})$ dominates $O(1)$ for sufficiently small $\mu_{\max}$, we can also write

Proof

$$
\begin{aligned}
(I - \mathcal{X})^{-1} &= \begin{bmatrix} (I_{(2M)^2} - \mathcal{X}_{11})^{-1} & 0 \\ 0 & 0 \end{bmatrix} + O(1) \qquad (9.273)\\
&= \begin{bmatrix} \{(I_{2M} \otimes E_{11}) + (D_{11} \otimes I_{2M})\}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + O(1)\\
&= \begin{bmatrix} I_{(2M)^2} \\ 0 \end{bmatrix} Z^{-1} \begin{bmatrix} I_{(2M)^2} & 0 \end{bmatrix} + O(1)
\end{aligned}
$$

where we used the fact from (9.245) that, for $h = 2$,

$$
Z = (I_{2M} \otimes E_{11}) + (D_{11} \otimes I_{2M}) \qquad (9.274)
$$

Proof

Substituting (9.273) into (9.261) and using expressions (9.250) and (9.255) for D_{11} and E_{11} we arrive at the following low-rank approximation:

$$
\begin{aligned}
&(I - \mathcal{F})^{-1} \\
&= \left(p \otimes I_{2M}\right) \otimes_b \left(p \otimes I_{2M}\right) Z^{-1} \left(\mathbb{1}^{\mathsf{T}} \otimes I_{2M}\right) \otimes_b \left(\mathbb{1}^{\mathsf{T}} \otimes I_{2M}\right) + O(1) \\
&\overset{(a)}{=} \left[(p \otimes p) \otimes (I_{2M} \otimes I_{2M})\right] (1 \otimes Z^{-1}) \left[(\mathbb{1} \otimes \mathbb{1})^{\mathsf{T}} \otimes (I_{2M} \otimes I_{2M})\right] + O(1) \\
&= \left[(p \otimes p) \otimes I_{4M^2}\right] (1 \otimes Z^{-1}) \left[(\mathbb{1} \otimes \mathbb{1})^{\mathsf{T}} \otimes I_{4M^2}\right] + O(1) \\
&= \left[(p \otimes p) \otimes Z^{-1}\right] \left[(\mathbb{1} \otimes \mathbb{1})^{\mathsf{T}} \otimes I_{4M^2}\right] + O(1) \\
&= \left[(p \otimes p)(\mathbb{1} \otimes \mathbb{1})^{\mathsf{T}}\right] \otimes Z^{-1} + O(1) \qquad (9.275)
\end{aligned}
$$

Proof

where step (a) uses the third property from Table F.2 in the appendix. Observe that the matrix $(p \otimes p)(\mathbb{1} \otimes \mathbb{1})^{\mathsf{T}}$ has rank one and, therefore, the above representation for $(I - \mathcal{F})^{-1})$ amounts to a low-rank approximation. Moreover, since $Z = O(\mu_{\max})$, we conclude from (9.275) that (9.243) holds. We also conclude that (9.242) holds since

$$(I - \bar{\mathcal{F}})^{-1} = (\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon)^{-1} (I - \mathcal{F})^{-1} (\mathcal{V}_\epsilon \otimes_b \mathcal{V}_\epsilon) = (I - \mathcal{X})^{-1} \tag{9.276}$$

$\square$

Mean-Error Stability

We now return to examine the mean-error stability of recursion (9.171). For this purpose, we need to introduce a smoothness condition on the Hessian matrices of the individual costs. This condition was not needed while establishing the stability of the second and fourth-order moments, $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^2$ and $\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{k,i}\|^4$, in the earlier sections. This same smoothness condition will be adopted in the next two chapters when we study the long-term behavior of the network and its performance.

Mean-Error Stability

Theorem 9.6 (Network mean-error stability). 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 Assumption 6.1. Assume additionally that each $J_k(w)$ satisfies a smoothness condition relative to the limit point $w^\star$, defined by (8.55), of the following form:

$$\left\|\nabla_w^2\, J_k(w^\star + \Delta w) - \nabla_w^2\, J_k(w^\star)\right\| \;\leq\; \kappa_d \left\|\Delta w\right\| \qquad (9.277)$$

for small perturbations $\|\Delta w\| \leq \epsilon$ and for some $\kappa_d \geq 0$. Assume further that the first and second-order moments of the gradient noise process satisfy the conditions of Assumption 8.1. Then, the first-order moment of the network errors satisfy

$$\limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}\| \;=\; O(\mu_{\max}), \quad k = 1, 2, \ldots, N \qquad (9.278)$$

Proof

Proof. We multiply both sides of the error recursion (9.171) from the left by $\mathcal{V}_\epsilon^{\mathsf{T}}$ and use (9.57) and (9.206) to get

$$\underbrace{\begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^e \end{bmatrix}}_{\triangleq\, z_i} = \underbrace{\begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_\epsilon^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix}}_{\triangleq\, \bar{\mathcal{B}}} \underbrace{\begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_{i-1}^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_{i-1}^e \end{bmatrix}}_{\triangleq\, z_{i-1}} - \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} + \mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} c_{i-1} \tag{9.279}$$

where the matrix $\bar{\mathcal{B}}$ from (9.226) is stable. We already know from (9.59) that $\|\check{b}^e\| = O(\mu_{\max})$. We now verify that the limit superior of $\|\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} c_{i-1}\|$ is $O(\mu_{\max}^2)$.

Proof

Indeed, in view of result (E.61) from the appendix, we know that condition (9.277) also holds globally for any Δw with κ_d replaced by some constant κ'_d. Then, for each agent k:

$$
\begin{aligned}
\|\widetilde{\boldsymbol{H}}_{k,i-1}\| &\overset{\Delta}{=} \|H - \boldsymbol{H}_{k,i-1}\| \\
&\leq \int_0^1 \left\| \nabla_w^2 J_k(w^\star) - \nabla_w^2 J_k(w^\star - t\widetilde{\boldsymbol{\phi}}_{k,i-1}) \right\| dt \\
&\overset{(9.277)}{\leq} \int_0^1 \kappa'_d \|t\widetilde{\boldsymbol{\phi}}_{k,i-1}\| dt \\
&= \frac{1}{2}\kappa'_d \|\widetilde{\boldsymbol{\phi}}_{k,i-1}\|
\end{aligned}
$$

Proof

$$
\begin{aligned}
&\leq \quad \frac{1}{2}\kappa'_d \left\| \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \widetilde{\boldsymbol{w}}_{\ell,i-1} \right\| \\
&\overset{(F.26)}{\leq} \quad \frac{1}{2}\kappa'_d \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \left\| \widetilde{\boldsymbol{w}}_{\ell,i-1} \right\| \\
&\leq \quad \frac{1}{2}\kappa'_d \sum_{\ell \in \mathcal{N}_k} \left\| \widetilde{\boldsymbol{w}}_{\ell,i-1} \right\| \\
&\leq \quad \frac{1}{2}\kappa'_d \sum_{\ell \in \mathcal{N}_k} \left\| \widetilde{\boldsymbol{w}}^e_{\ell,i-1} \right\| \\
&\leq \quad \frac{1}{2}\kappa'_d N \left\| \widetilde{\boldsymbol{w}}^e_{i-1} \right\| \qquad (9.280)
\end{aligned}
$$

Proof

so that

$$\|\widetilde{\boldsymbol{\mathcal{H}}}_{i-1}\| = \max_{1\le k\le N} \|\widetilde{\boldsymbol{H}}_{k,i-1}\| \le \frac{1}{2}\kappa'_d N \|\widetilde{\boldsymbol{w}}^e_{i-1}\| \tag{9.281}$$

and, consequently,

$$\begin{aligned}
\|\mathcal{V}_\epsilon^{\mathsf{T}}\mathcal{A}_2^{\mathsf{T}}\mathcal{M}c_{i-1}\| &\overset{(9.172)}{\le} \|\mathcal{V}_\epsilon\|\,\|\mathcal{A}_2\|\,\|\mathcal{M}\|\,\|\mathbb{E}\,\widetilde{\boldsymbol{\mathcal{H}}}_{i-1}\,\mathcal{A}_1^{\mathsf{T}}\,\widetilde{\boldsymbol{w}}^e_{i-1}\| \\
&\le \|\mathcal{V}_\epsilon\|\,\|\mathcal{A}_2\|\,\|\mathcal{M}\|\,\|\mathcal{A}_1\|\,\mathbb{E}\left[\|\widetilde{\boldsymbol{\mathcal{H}}}_{i-1}\|\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|\right] \\
&\le \frac{1}{2}\kappa'_d N\|\mathcal{V}_\epsilon\|\,\|\mathcal{A}_2\|\,\|\mathcal{M}\|\,\|\mathcal{A}_1\|\,\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|^2 \\
&\overset{\Delta}{=} r\mu_{\max}\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|^2
\end{aligned} \tag{9.282}$$

Proof

for some constant r that is independent of $\mu_{\max}$. It then follows from (9.11) that

$$\limsup_{i\to\infty} \|\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} c_{i-1}\| \;=\; O(\mu_{\max}^2) \tag{9.283}$$

as claimed, where one $\mu_{\max}$ arises from $\mathcal{M}$ and the other $\mu_{\max}$ arises from (9.11).

Returning to (9.279), we partition the vectors z_i and $\mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} c_{i-1}$ into

$$z_i \;\triangleq\; \begin{bmatrix} \bar{z}_i \\ \check{z}_i \end{bmatrix}, \quad \mathcal{V}_\epsilon^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{M} c_{i-1} \;\triangleq\; \begin{bmatrix} \bar{c}_{i-1} \\ \check{c}_{i-1} \end{bmatrix} \tag{9.284}$$

Proof

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

$$\begin{bmatrix} \bar{z}_i \\ \check{z}_i \end{bmatrix} = \begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_{\epsilon}^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \bar{z}_{i-1} \\ \check{z}_{i-1} \end{bmatrix} + \begin{bmatrix} \bar{c}_{i-1} \\ \check{c}_{i-1} \end{bmatrix} + \begin{bmatrix} 0 \\ O(\mu_{\max}) \end{bmatrix} \tag{9.285}$$

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 two differences. First, the matrices $\{D_{11}, D_{12}, D_{21}, D_{22}\}$ in (9.285) 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). In particular, it continues to hold that

Proof

$$\|I_{2M} - D_{11}^{\mathsf{T}}\| \overset{(9.47)}{\leq} 1 - \sigma_{11}\mu_{\max} \tag{9.286}$$

$$\|D_{12}\| \overset{(9.51)}{\leq} \sigma_{12}\mu_{\max} \tag{9.287}$$

$$\|D_{21}\| \overset{(9.50)}{\leq} \sigma_{21}\mu_{\max} \tag{9.288}$$

$$\|D_{22}\| \overset{(9.51)}{\leq} \sigma_{22}\mu_{\max} \tag{9.289}$$

for some positive constants $\{\sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22}\}$ that are independent of $\mu_{\max}$. Second, the gradient noise terms that appeared in (9.60) are now replaced by

Sketch of Argument

$$\begin{bmatrix} \|\bar{z}_i\|^2 \\ \|\check{z}_i\|^2 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\Gamma} \begin{bmatrix} \|\bar{z}_{i-1}\|^2 \\ \|\check{z}_{i-1}\|^2 \end{bmatrix} + \underbrace{\begin{bmatrix} e \\ f \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}) \\ O(\mu^2_{\max}) \end{bmatrix}} \underbrace{\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|^4}_{O(\mu^2_{\max})} + \begin{bmatrix} 0 \\ O(\mu^2_{\max}) \end{bmatrix} \tag{9.301}$$

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

Sketch of Argument

We conclude that, as $i \to \infty$,

$$\limsup_{i\to\infty} \|\bar{z}_i\|^2 = O(\mu_{\max}^2), \qquad \limsup_{i\to\infty} \mathbb{E}\,\|\check{z}_i\|^2 = O(\mu_{\max}^2) \tag{9.303}$$

and, hence,

$$\limsup \|z_i\|^2 = O(\mu_{\max}^2) \tag{9.304}$$

It follows that

$$\limsup_{i\to\infty} \|z_i\| = O(\mu_{\max}) \tag{9.305}$$

Consequently,

$$\limsup_{i\to\infty} \left\| \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^e \end{bmatrix} \right\| = O(\mu_{\max}) \tag{9.306}$$

Sketch of Argument

and, hence,

$$
\begin{aligned}
\limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}\| &\le \limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_i^e\| \\
&\le \limsup_{i\to\infty} \left\| \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^e \end{bmatrix} \right\| \\
&\le \left\| \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \right\| \left(\limsup_{i\to\infty} \left\| \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^e \end{bmatrix} \right\| \right) \\
&= O(\mu_{\max})
\end{aligned}
\tag{9.307}
$$

as claimed.

□

Detailed Proof

$$\begin{bmatrix} \bar{z}_i \\ \check{z}_i \end{bmatrix} = \begin{bmatrix} I_{2M} - D_{11}^{\mathsf{T}} & -D_{21}^{\mathsf{T}} \\ -D_{12}^{\mathsf{T}} & \mathcal{J}_{\epsilon}^{\mathsf{T}} - D_{22}^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \bar{z}_{i-1} \\ \check{z}_{i-1} \end{bmatrix} + \begin{bmatrix} \bar{c}_{i-1} \\ \check{c}_{i-1} \end{bmatrix} + \begin{bmatrix} 0 \\ O(\mu_{\max}) \end{bmatrix} \tag{9.285}$$

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 two differences. First, the matrices $\{D_{11}, D_{12}, D_{21}, D_{22}\}$ in (9.285) 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).

Proof

From (9.282) and using the fact that $(\mathbb{E}\,\boldsymbol{a})^2 \leq \mathbb{E}\,\boldsymbol{a}^2$ for any real random variable $\boldsymbol{a}$, we have that

$$\|\mathcal{V}_\epsilon^\mathsf{T} \mathcal{A}_2^\mathsf{T} \mathcal{M} c_{i-1}\|^2 \leq r^2 \mu_{\max}^2 \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4 \tag{9.290}$$

and, hence,

$$\|\bar{c}_{i-1}\|^2 \leq r^2 \mu_{\max}^2 \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4, \qquad \|\check{c}_{i-1}\|^2 \leq r^2 \mu_{\max}^2 \mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4 \tag{9.291}$$

Now, if we repeat the argument that led to (9.106), with proper adjustments, we can show that relations similar to (9.69) and (9.81) continue to hold for $\{\|\bar{z}_i\|^2, \|\check{z}_i\|^2\}$. The argument is as follows.

Proof

We first appeal to Jensen's inequality (F.26) from the appendix and apply it to the function $f(x) = \|x\|^2$ to obtain the bound:

$$
\begin{aligned}
\|\bar{z}_i\|^2 &= \left\|(1-t)\frac{1}{1-t}(I_{2M} - D_{11}^{\mathsf{T}})\bar{z}_{i-1} + t\frac{1}{t}\left(-D_{21}^{\mathsf{T}}\check{z}_{i-1} + \bar{c}_{i-1}\right)\right\|^2 \\
&\leq \frac{1}{1-t}(1-\sigma_{11}\mu_{\max})^2\|\bar{z}_{i-1}\|^2 + \frac{2}{t}\left(\sigma_{21}^2\mu_{\max}^2\|\check{z}_{i-1}\|^2 + \|\bar{c}_{i-1}\|^2\right) \\
&\leq (1-\sigma_{11}\mu_{\max})\|\bar{z}_{i-1}\|^2 + \frac{2}{\sigma_{11}\mu_{\max}}\left(\sigma_{21}^2\mu_{\max}^2\|\check{z}_{i-1}\|^2 + \|\bar{c}_{i-1}\|^2\right) \\
&\leq (1-\sigma_{11}\mu_{\max})\|\bar{z}_{i-1}\|^2 + \frac{2\sigma_{21}^2\mu_{\max}}{\sigma_{11}}\|\check{z}_{i-1}\|^2 + \frac{2r^2\mu_{\max}}{\sigma_{11}}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4
\end{aligned}
\tag{9.292}
$$

Proof

for any arbitrary positive number $t \in (0,1)$. We selected $t = \sigma_{11}\mu_{\max}$ in the above derivation. We repeat a similar argument for $\|\check{z}_i\|^2$. Thus, using Jensen's inequality again we have

$$\begin{aligned}
\|\check{z}_i\|^2 \quad &= \quad \left\| t\frac{1}{t}\mathcal{J}_\epsilon^{\mathsf{T}}\check{z}_{i-1} - (1-t)\frac{1}{1-t}\left[-D_{22}^{\mathsf{T}}\check{z}_{i-1} - D_{12}^{\mathsf{T}}\bar{z}_{i-1} + \check{c}_{i-1} + O(\mu_{\max})\right]\right\|^2 \\
&\overset{(9.76)}{\leq} \quad \frac{1}{t}\left(\rho(J_\epsilon)+\epsilon\right)^2\|\check{z}_{i-1}\|^2 + \\
&\qquad \frac{4}{1-t}\left[\sigma_{22}^2\mu_{\max}^2\|\check{z}_{i-1}\|^2 + \sigma_{12}^2\mu_{\max}^2\|\bar{z}_{i-1}\|^2 + \|\check{c}_{i-1}\|^2 + O(\mu_{\max}^2)\right]
\end{aligned} \tag{9.293}$$

Proof

for any arbitrary positive number $t \in (0,1)$. Since we know that $\rho(J_\epsilon) \in (0,1)$, then we can select ϵ small enough to ensure $t = \rho(J_\epsilon) + \epsilon \in (0,1)$ and rewrite (9.293) as

$$
\begin{aligned}
\|\check{z}_i\|^2 \;\; \leq \;\; & \left(\rho(J_\epsilon) + \epsilon + \frac{4\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\|\check{z}_{i-1}\|^2 + \\
& \left(\frac{4\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\|\bar{z}_{i-1}\|^2 + \\
& \left(\frac{4r^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon}\right)\mathbb{E}\,\|\widetilde{\boldsymbol{w}}_{i-1}^e\|^4 + O(\mu_{\max}^2)
\end{aligned}
\qquad (9.294)
$$

Proof

If we now introduce the scalar coefficients

$$
\begin{aligned}
a &= 1 - \sigma_{11}\mu_{\max} = 1 - O(\mu_{\max}) && (9.295)\\
b &= \frac{2\sigma_{21}^2\mu_{\max}}{\sigma_{11}} = O(\mu_{\max}) && (9.296)\\
c &= \frac{4\sigma_{12}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = O(\mu_{\max}^2) && (9.297)\\
d &= \rho(J_\epsilon) + \epsilon + \frac{4\sigma_{22}^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = \rho(J_\epsilon)+\epsilon+O(\mu_{\max}^2) && (9.298)\\
e &= \frac{2r^2\mu_{\max}}{\sigma_{11}} = O(\mu_{\max}) && (9.299)\\
f &= \frac{4r^2\mu_{\max}^2}{1-\rho(J_\epsilon)-\epsilon} = O(\mu_{\max}^2) && (9.300)
\end{aligned}
$$

Proof

we can combine (9.292) and (9.294) into a single compact inequality recursion as follows:

$$\begin{bmatrix} \|\bar{z}_i\|^2 \\ \|\check{z}_i\|^2 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\Gamma} \begin{bmatrix} \|\bar{z}_{i-1}\|^2 \\ \|\check{z}_{i-1}\|^2 \end{bmatrix} + \underbrace{\begin{bmatrix} e \\ f \end{bmatrix}}_{\begin{bmatrix} O(\mu_{\max}) \\ O(\mu^2_{\max}) \end{bmatrix}} \underbrace{\mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_{i-1}\|^4}_{O(\mu^2_{\max})} + \begin{bmatrix} 0 \\ O(\mu^2_{\max}) \end{bmatrix} \tag{9.301}$$

in terms of the 2×2 coefficient matrix Γ indicated above. We know from the argument (9.102) that Γ is stable for sufficiently small step-sizes. If we now recall the result

$$\limsup_{i\to\infty} \mathbb{E}\,\|\widetilde{\boldsymbol{w}}^e_i\|^4 \overset{(9.107)}{=} O(\mu^2_{\max}) \tag{9.302}$$

Proof

We conclude that, as $i \to \infty$,

$$\limsup_{i\to\infty} \|\bar{z}_i\|^2 = O(\mu_{\max}^2), \qquad \limsup_{i\to\infty} \mathbb{E}\,\|\check{z}_i\|^2 = O(\mu_{\max}^2) \tag{9.303}$$

and, hence,

$$\limsup \|z_i\|^2 = O(\mu_{\max}^2) \tag{9.304}$$

It follows that

$$\limsup_{i\to\infty} \|z_i\| = O(\mu_{\max}) \tag{9.305}$$

Consequently,

$$\limsup_{i\to\infty} \left\| \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^e \end{bmatrix} \right\| = O(\mu_{\max}) \tag{9.306}$$

Proof

and, hence,

$$\begin{aligned}
\limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_{k,i}\| &\leq \limsup_{i\to\infty} \|\mathbb{E}\,\widetilde{\boldsymbol{w}}_i^e\| \\
&\leq \limsup_{i\to\infty} \left\| \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^e \end{bmatrix} \right\| \\
&\leq \left\| \left(\mathcal{V}_\epsilon^{-1}\right)^{\mathsf{T}} \right\| \left(\limsup_{i\to\infty} \left\| \begin{bmatrix} \mathbb{E}\,\bar{\boldsymbol{w}}_i^e \\ \mathbb{E}\,\check{\boldsymbol{w}}_i^e \end{bmatrix} \right\| \right) \\
&= O(\mu_{\max}) \qquad (9.307)
\end{aligned}$$

as claimed.

□

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.