

INFERENCE OVER NETWORKS

LECTURE #17: Stability of Multi-Agent Networks

Professor Ali H. Sayed
UCLA Electrical Engineering



Part V: Multi-Agent Network Stability and Performance



Reference

3

Lecture #17: *Stability of Multi-Agent Networks*

EE210B: *Inference over Networks* (A. H. Sayed)

Chapter 8 (Evolution of Multi-Agent Networks, pp. 496-506):
Chapter 9 (Stability of Multi-Agent Networks, pp. 507-551):

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.

Distributed Strategies



Unified Description

Table 8.2: Update equations for non-cooperative, diffusion, and consensus strategies.

algorithm	update equations
non-cooperative	$\mathbf{w}_{k,i} = \mathbf{w}_{k,i-1} - \mu_k \widehat{\nabla_{w^*} J}_k (\mathbf{w}_{k,i-1})$
consensus	$\begin{cases} \boldsymbol{\psi}_{k,i-1} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \mathbf{w}_{\ell,i-1} \\ \mathbf{w}_{k,i} = \boldsymbol{\psi}_{k,i-1} - \mu_k \widehat{\nabla_{w^*} J}_k (\mathbf{w}_{k,i-1}) \end{cases}$
CTA diffusion	$\begin{cases} \boldsymbol{\psi}_{k,i-1} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \mathbf{w}_{\ell,i-1} \\ \mathbf{w}_{k,i} = \boldsymbol{\psi}_{k,i-1} - \mu_k \widehat{\nabla_{w^*} J}_k (\boldsymbol{\psi}_{k,i-1}) \end{cases}$
ATC diffusion	$\begin{cases} \boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} - \mu_k \widehat{\nabla_{w^*} J}_k (\mathbf{w}_{k,i-1}) \\ \mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \boldsymbol{\psi}_{\ell,i} \end{cases}$

Unified Description



In a manner similar to (8.5), we can again describe these strategies by means of a single unifying description as follows:

$$\left\{ \begin{array}{lcl} \phi_{k,i-1} & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \mathbf{w}_{\ell,i-1} \\ \psi_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \phi_{\ell,i-1} - \mu_k \widehat{\nabla_{w^*} J}_k(\phi_{k,i-1}) \\ \mathbf{w}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \psi_{\ell,i} \end{array} \right. \quad (8.46)$$



Unified Description

7

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where $\{\phi_{k,i-1}, \psi_{k,i}\}$ denote $M \times 1$ intermediate variables, while the nonnegative entries of the $N \times N$ matrices $A_o = [a_{o,\ell k}]$, $A_1 = [a_{1,\ell k}]$, and $A_2 = [a_{2,\ell k}]$ satisfy the same conditions (7.10) and, hence, the matrices $\{A_o, A_1, A_2\}$ are left-stochastic

$$A_o^T \mathbf{1} = \mathbf{1}, \quad A_1^T \mathbf{1} = \mathbf{1}, \quad A_2^T \mathbf{1} = \mathbf{1} \quad (8.47)$$

We assume that each of these combination matrices defines an underlying *connected* network topology so that none of their rows are identically zero.



Unified Description

8

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Again, different choices for $\{A_o, A_1, A_2\}$ correspond to different distributed strategies, as indicated earlier by (8.7)–(8.10), and where the left-stochastic matrix P represents the product:

$$P \triangleq A_1 A_o A_2 \quad (8.48)$$

Perron Eigenvector



9

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

- (a) The matrix P has a *single* eigenvalue at one.
- (b) All other eigenvalues of P are strictly inside the unit circle so that $\rho(P) = 1$.
- (c) With proper sign scaling, all entries of the right-eigenvector of P corresponding to the single eigenvalue at one are *positive*. Let p denote this right-eigenvector, with its entries $\{p_k\}$ normalized to add up to one, i.e.,

$$Pp = p, \quad \mathbb{1}^\top p = 1, \quad p_k > 0, \quad k = 1, 2, \dots, N \quad (8.49)$$

We refer to p as the *Perron eigenvector* of P .

Weighted Aggregate Cost



10

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Following [68–70], we next introduce the vector:

$$q \triangleq \text{diag}\{\mu_1, \mu_2, \dots, \mu_N\} A_2 p \quad (8.50)$$

It is clear that all entries of q are strictly positive since each $\mu_k > 0$ and the entries of $A_2 p$ are all positive. The latter statement follows from the fact that each entry of $A_2 p$ is a linear combination of the positive entries of p . Therefore, if we denote the individual entries of the vector q by $\{q_k\}$, then it holds that

$$q_k > 0, \quad k = 1, 2, \dots, N \quad (8.51)$$

Weighted Aggregate Cost



11

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We also represent the step-sizes as scaled multiples of the same factor μ_{\max} , namely,

$$\mu_k \stackrel{\Delta}{=} \tau_k \mu_{\max}, \quad k = 1, 2, \dots, N \quad (8.52)$$

where $0 < \tau_k \leq 1$. In this way, it becomes clear that all step-sizes become smaller as μ_{\max} is reduced in size.

We further introduce the weighted aggregate cost

$$J^{\text{glob},*}(w) \stackrel{\Delta}{=} \sum_{k=1}^N q_k J_k(w) \quad (8.53)$$



Weighted Aggregate Cost

12

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$J^{\text{glob}}(w) \triangleq \sum_{k=1}^N J_k(w) \rightarrow$$

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

$$J^{\text{glob},\star}(w) \triangleq \sum_{k=1}^N q_k J_k(w) \rightarrow$$

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

(is also **strongly convex**)

Error Vectors



From this point onwards, we shall therefore measure the performance of the distributed strategy (8.46) by using w^* as the reference vector (instead of w^o) and define the error vectors as:

$$\tilde{\mathbf{w}}_{k,i} \triangleq w^* - \mathbf{w}_{k,i} \quad (8.106)$$

$$\tilde{\boldsymbol{\psi}}_{k,i} \triangleq w^* - \boldsymbol{\psi}_{k,i} \quad (8.107)$$

$$\tilde{\boldsymbol{\phi}}_{k,i-1} \triangleq w^* - \boldsymbol{\phi}_{k,i-1} \quad (8.108)$$



Gradient Noise + Bias Vector

14

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Moreover, with each agent k , we associate a gradient noise vector in addition to a mismatch (or bias) vector, namely,

$$s_{k,i}(\phi_{k,i-1}) \triangleq \widehat{\nabla_{w^*} J_k}(\phi_{k,i-1}) - \nabla_{w^*} J_k(\phi_{k,i-1}) \quad (8.109)$$

and

$$b_k \triangleq -\nabla_{w^*} J_k(w^*) \quad (8.110)$$

Example #8.11



Example 8.11 (Gradient noise over MSE networks). Let us continue with the setting of Example 8.8, which deals with a *variation* of MSE networks where the data model at each agent is instead assumed to be given by

$$\mathbf{d}_k(i) = \mathbf{u}_{k,i} w_k^o + \mathbf{v}_k(i) \quad (8.126)$$

with the model vectors, w_k^o , being possibly different at the various agents. In a manner similar to (8.15), we can verify that if the distributed strategy (8.5) is employed at the agents, then the resulting gradient noise process at each agent k is now given by:

$$s_{k,i}(\phi_{k,i-1}) = \frac{2}{h} (R_{u,k} - \mathbf{u}_{k,i}^* \mathbf{u}_{k,i}) (w_k^o - \phi_{k,i-1}) - \frac{2}{h} \mathbf{u}_{k,i}^* \mathbf{v}_k(i) \quad (8.127)$$

Extended Error Dynamics

Error Dynamics



17

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We explained earlier after (8.45) that because the Hessian matrices, $\nabla_w^2 J_k(w)$, are not generally block diagonal, we will need to introduce extended versions of the error quantities $\{\tilde{w}_{k,i}, \tilde{\psi}_{k,i}, \tilde{\phi}_{k,i-1}\}$ in order to fully capture the dynamics of the network in the general case. This is in contrast to the mean-square-error case studied in Example 8.1 where these errors were sufficient to arrive at the state recursions (8.22) or (8.25) for the evolution of the network dynamics.



Error Dynamics

18

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

To motivate the need for extended error vectors, let us first introduce some notation. Thus, note that if we express any column vector $w \in \mathbb{C}^M$ in terms of its real and imaginary parts $x, y \in \mathbb{R}^M$, then

$$w = x + jy \quad (\text{a column vector}) \quad (8.128)$$

$$w^* = x^\top - jy^\top \quad (\text{a row vector}) \quad (8.129)$$

$$(w^*)^\top = x - jy \quad (\text{a column vector}) \quad (8.130)$$

In other words, the quantity $(w^*)^\top$ is again a *column* vector, just like w , except that its complex representation is obtained by replacing j by $-j$.

Error Dynamics



19

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

The reason why we need to introduce the quantity $(w^*)^\top$ is because, as the discussion will reveal, we will need to track the evolution of both quantities $w_{k,i}$ and $(w_{k,i}^*)^\top$ in the general case in order to examine how the network is performing. Thus, using equations (8.46), we can deduce similar relations for the evolution of the complex conjugate iterates, namely,

Error Dynamics



$$\left\{ \begin{array}{lcl} \left(\phi_{k,i-1}^* \right)^\top & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \left(\boldsymbol{w}_{\ell,i-1}^* \right)^\top \\ \left(\psi_{k,i}^* \right)^\top & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \left(\phi_{\ell,i-1}^* \right)^\top - \mu_k \widehat{\nabla_{w^\top J}}_k \left(\phi_{k,i-1} \right) \\ \left(\boldsymbol{w}_{k,i}^* \right)^\top & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \left(\psi_{\ell,i}^* \right)^\top \end{array} \right. \quad (8.131)$$

Error Dynamics



21

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Observe how the gradient vector approximation that appears in the second equation now involves differentiation relative to w^T and not w^* . Representations (8.46) and (8.131) can be grouped together into a single set of equations by introducing extended vectors of dimensions $2M \times 1$ as follows:

Error Dynamics



$$\left\{ \begin{array}{lcl} \begin{bmatrix} \phi_{k,i-1} \\ (\phi_{k,i-1}^*)^\top \end{bmatrix} & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \begin{bmatrix} \mathbf{w}_{\ell,i-1} \\ (\mathbf{w}_{\ell,i-1}^*)^\top \end{bmatrix} \\ \begin{bmatrix} \psi_{k,i} \\ (\psi_{k,i}^*)^\top \end{bmatrix} & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \begin{bmatrix} \phi_{\ell,i-1} \\ (\phi_{\ell,i-1}^*)^\top \end{bmatrix} - \mu_k \begin{bmatrix} \widehat{\nabla_{w^*} J}_k(\phi_{k,i-1}) \\ \widehat{\nabla_{w^\top} J}_k(\phi_{k,i-1}) \end{bmatrix} \\ \begin{bmatrix} \mathbf{w}_{k,i} \\ (\mathbf{w}_{k,i}^*)^\top \end{bmatrix} & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \begin{bmatrix} \psi_{\ell,i} \\ (\psi_{\ell,i}^*)^\top \end{bmatrix} \end{array} \right. \quad (8.132)$$



Error Dynamics

23

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We therefore extend the error vectors into size $2M \times 1$ and introduce

$$\tilde{\mathbf{w}}_{k,i}^e \triangleq \begin{bmatrix} \tilde{\mathbf{w}}_{k,i} \\ (\tilde{\mathbf{w}}_{k,i}^*)^\top \end{bmatrix}, \quad \tilde{\boldsymbol{\psi}}_{k,i}^e \triangleq \begin{bmatrix} \tilde{\boldsymbol{\psi}}_{k,i} \\ (\tilde{\boldsymbol{\psi}}_{k,i}^*)^\top \end{bmatrix}, \quad \tilde{\boldsymbol{\phi}}_{k,i-1}^e \triangleq \begin{bmatrix} \tilde{\boldsymbol{\phi}}_{k,i-1} \\ (\tilde{\boldsymbol{\phi}}_{k,i-1}^*)^\top \end{bmatrix} \quad (8.133)$$

where we are using the superscript “ e ” to refer to extended quantities of size $2M \times 1$. We also introduce extended versions of the limit vector, the gradient noise vector, and the bias vector:



Error Dynamics

24

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$(w^*)^e \triangleq \begin{bmatrix} w^* \\ ((w^*)^*)^\top \end{bmatrix}, \quad s_{k,i}^e \triangleq \begin{bmatrix} s_{k,i}(\phi_{k,i-1}) \\ (s_{k,i}^*(\phi_{k,i-1}))^\top \end{bmatrix}, \quad b_k^e \triangleq \begin{bmatrix} b_k \\ (b_k^*)^\top \end{bmatrix} \quad (8.134)$$

where the vector $s_{k,i}^e$ in (8.134) should have been written more explicitly as $s_{k,i}^e(\phi_{k,i-1})$; we are dropping the argument for compactness of notation. Now, subtracting $(w^*)^e$ from both sides of the equations in (8.132) and using (8.109) gives



Error Dynamics

25

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\left\{ \begin{array}{lcl} \tilde{\phi}_{k,i-1}^e & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \tilde{w}_{\ell,i-1}^e \\ \\ \tilde{\psi}_{k,i}^e & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \tilde{\phi}_{\ell,i-1}^e + \mu_k \begin{bmatrix} \nabla_{w^*} J_k(\phi_{k,i-1}) \\ \nabla_{w^\top} J_k(\phi_{k,i-1}) \end{bmatrix} + \mu_k s_{k,i}^e \\ \\ \tilde{w}_{k,i}^e & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \tilde{\psi}_{\ell,i}^e \end{array} \right. \quad (8.135)$$



Error Dynamics

26

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We observe that the gradient vectors in (8.135) are being evaluated at the intermediate variable, $\phi_{k,i-1}$, and not at any of the *error* variables. For this reason, equation (8.135) is still not an actual recursion. To transform it into a recursion that only involves error variables, we call upon the mean-value theorem (D.20) from the appendix, which allows us to write:

$$\begin{bmatrix} \nabla_{w^*} J_k(\phi_{k,i-1}) \\ \nabla_{w^\top} J_k(\phi_{k,i-1}) \end{bmatrix} = \underbrace{\begin{bmatrix} \nabla_{w^*} J_k(w^*) \\ \nabla_{w^\top} J_k(w^*) \end{bmatrix}}_{\triangleq -b_k^e} - \underbrace{\left[\int_0^1 \nabla_w^2 J_k(w^* - t\tilde{\phi}_{k,i-1}) dt \right]}_{\triangleq \mathbf{H}_{k,i-1}} \tilde{\phi}_{k,i-1}^e \quad (8.136)$$

Error Dynamics



That is,

$$\begin{bmatrix} \nabla_{w^*} J_k(\phi_{k,i-1}) \\ \nabla_{w^\top} J_k(\phi_{k,i-1}) \end{bmatrix} = -b_k^e - \mathbf{H}_{k,i-1} \tilde{\phi}_{k,i-1}^e \quad (8.137)$$

in terms of a $2M \times 2M$ stochastic matrix $\mathbf{H}_{k,i-1}$ defined in terms of the integral of the $2M \times 2M$ Hessian matrix of agent k :

$$\mathbf{H}_{k,i-1} \triangleq \int_0^1 \nabla_w^2 J_k(w^* - t\tilde{\phi}_{k,i-1}) dt \quad (8.138)$$

Substituting (8.137) into (8.135) leads to



Error Dynamics

28

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\left\{ \begin{array}{lcl} \tilde{\phi}_{k,i-1}^e & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \tilde{w}_{\ell,i-1}^e \\ \\ \tilde{\psi}_{k,i}^e & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \tilde{\phi}_{\ell,i-1}^e - \mu_k \mathbf{H}_{k,i-1} \tilde{\phi}_{k,i-1}^e - \mu_k b_k^e + \mu_k s_{k,i}^e \\ \\ \tilde{w}_{k,i}^e & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \tilde{\psi}_{\ell,i}^e \end{array} \right. \quad (8.139)$$



Error Dynamics

29

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

These equations describe the evolution of the error quantities at the individual agents for $k = 1, 2, \dots, N$. Observe that when the matrix $\mathbf{H}_{k,i-1}$ happens to be block diagonal, which occurs when the Hessian matrix function itself is block diagonal (as happened in (8.4) with the quadratic costs in Example 8.1), then the last term in (8.137) decouples into two separate terms in the variables

$$\left\{ \tilde{\phi}_{k,i-1}, \left(\tilde{\phi}_{k,i-1}^* \right)^T \right\} \quad (8.140)$$



Error Dynamics

30

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

since then

$$\mathbf{H}_{k,i-1} \tilde{\phi}_{k,i-1}^e \equiv \begin{bmatrix} \mathbf{H}_{k,i-1}^{11} & 0 \\ 0 & \mathbf{H}_{k,i-1}^{22} \end{bmatrix} \begin{bmatrix} \tilde{\phi}_{k,i-1} \\ (\tilde{\phi}_{k,i-1}^*)^\top \end{bmatrix} \quad (8.141)$$

In that case, it becomes unnecessary to propagate the extended vectors $\{\tilde{w}_{k,i}^e, \tilde{\psi}_{k,i}^e, \tilde{\phi}_{k,i-1}^e\}$ using (8.139); the dynamics of the network can be studied by examining solely the evolution of the original error vectors $\{\tilde{w}_{k,i}, \tilde{\psi}_{k,i}, \tilde{\phi}_{k,i-1}\}$, namely,



Error Dynamics

31

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\left\{ \begin{array}{lcl} \tilde{\phi}_{k,i-1} & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \tilde{w}_{\ell,i-1} \\ \\ \tilde{\psi}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \tilde{\phi}_{\ell,i-1} - \mu_k \mathbf{H}_{k,i-1}^{11} \tilde{\phi}_{k,i-1} - \mu_k b_k + \mu_k s_{k,i} \\ \\ \tilde{w}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \tilde{\psi}_{\ell,i} \end{array} \right. \quad (8.142)$$



Error Dynamics

32

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We continue our discussion by treating the general case (8.139). We collect the extended error vectors from all agents into the following $N \times 1$ block error vectors (whose individual entries are of size $2M \times 1$ each):

$$\tilde{w}_i^e \triangleq \begin{bmatrix} \tilde{w}_{1,i}^e \\ \tilde{w}_{2,i}^e \\ \vdots \\ \tilde{w}_{N,i}^e \end{bmatrix}, \quad \tilde{\phi}_{i-1}^e \triangleq \begin{bmatrix} \tilde{\phi}_{1,i-1}^e \\ \tilde{\phi}_{2,i-1}^e \\ \vdots \\ \tilde{\phi}_{N,i-1}^e \end{bmatrix}, \quad \tilde{\psi}_i^e \triangleq \begin{bmatrix} \tilde{\psi}_{1,i}^e \\ \tilde{\psi}_{2,i}^e \\ \vdots \\ \tilde{\psi}_{N,i}^e \end{bmatrix} \quad (8.143)$$

Error Dynamics



We also define the following block gradient noise and bias vectors:

$$\mathbf{s}_i^e \triangleq \begin{bmatrix} \mathbf{s}_{1,i}^e \\ \mathbf{s}_{2,i}^e \\ \vdots \\ \mathbf{s}_{N,i}^e \end{bmatrix}, \quad b^e \triangleq \begin{bmatrix} b_1^e \\ b_2^e \\ \vdots \\ b_N^e \end{bmatrix} \quad (8.144)$$

Error Dynamics



Now recall from the explanation after (8.134) that each entry, $s_{k,i}^e$, in (8.144) is dependent on $\phi_{k,i-1}$. Recall also from the distributed algorithm (8.46) that $\phi_{k,i-1}$ is a combination of various $\{\mathbf{w}_{\ell,i-1}\}$. Therefore, the block gradient vector, \mathbf{s}_i^e , defined in (8.144) is dependent on the network vector, \mathbf{w}_{i-1}^e , namely,



Error Dynamics

$$\mathbf{w}_{i-1}^e \triangleq \begin{bmatrix} \mathbf{w}_{1,i-1}^e \\ \mathbf{w}_{2,i-1}^e \\ \vdots \\ \mathbf{w}_{N,i-1}^e \end{bmatrix}, \quad \mathbf{w}_{k,i-1}^e \triangleq \begin{bmatrix} \mathbf{w}_{k,i-1} \\ (\mathbf{w}_{k,i-1}^*)^\top \end{bmatrix} \quad (8.145)$$

For this reason, we shall also write $\mathbf{s}_i^e(\mathbf{w}_{i-1}^e)$ rather than simply \mathbf{s}_i^e when it is desired to highlight the dependency of \mathbf{s}_i^e on \mathbf{w}_{i-1}^e .



Error Dynamics

36

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We further introduce the Kronecker products

$$\left\{ \begin{array}{l} \mathcal{A}_o \triangleq A_o \otimes I_{2M} \\ \mathcal{A}_1 \triangleq A_1 \otimes I_{2M} \\ \mathcal{A}_2 \triangleq A_2 \otimes I_{2M} \end{array} \right. \quad (8.146)$$

The matrix \mathcal{A}_o is an $N \times N$ *block* matrix whose (ℓ, k) -th block is equal to $a_{o,\ell k} I_{2M}$. Similarly, for \mathcal{A}_1 and \mathcal{A}_2 . Likewise, we introduce the following $N \times N$ *block diagonal* matrices, whose individual entries are of size $2M \times 2M$ each:



Error Dynamics

37

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mathcal{M} \triangleq \text{diag}\{\mu_1 I_{2M}, \mu_2 I_{2M}, \dots, \mu_N I_{2M}\} \quad (8.147)$$

$$\mathcal{H}_{i-1} \triangleq \text{diag}\{\mathbf{H}_{1,i-1}, \mathbf{H}_{2,i-1}, \dots, \mathbf{H}_{N,i-1}\} \quad (8.148)$$

We then conclude from (8.139) that the following relations hold for the *network variables*:

$$\begin{cases} \tilde{\phi}_{i-1}^e &= \mathcal{A}_1^\top \tilde{w}_{i-1}^e \\ \tilde{\psi}_i^e &= [\mathcal{A}_o^\top - \mathcal{M}\mathcal{H}_{i-1}] \tilde{\phi}_{i-1}^e + \mathcal{M}s_i^e(\tilde{w}_{i-1}^e) - \mathcal{M}b^e \\ \tilde{w}_i^e &= \mathcal{A}_2^\top \tilde{\psi}_i^e \end{cases} \quad (8.149)$$



Error Dynamics

38

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

so that the network weight error vector, $\tilde{\mathbf{w}}_i^e$, ends up evolving according to the following *stochastic* recursion over $i \geq 0$:

$$\tilde{\mathbf{w}}_i^e = \mathcal{A}_2^\top \left(\mathcal{A}_o^\top - \mathcal{M} \mathcal{H}_{i-1} \right) \mathcal{A}_1^\top \tilde{\mathbf{w}}_{i-1}^e + \mathcal{A}_2^\top \mathcal{M} \mathbf{s}_i^e(\mathbf{w}_{i-1}^e) - \mathcal{A}_2^\top \mathcal{M} \mathbf{b}^e \quad (8.150)$$

For comparison purposes, if each agent operates individually and uses the non-cooperative strategy, then the weight error vectors across all N agents would instead evolve according to the following stochastic recursion:



Error Dynamics

39

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\tilde{\mathbf{w}}_i^e = (I_{2MN} - \mathcal{M}\mathcal{H}_{i-1})\tilde{\mathbf{w}}_{i-1}^e + \mathcal{M}\mathbf{s}_i^e(\mathbf{w}_{i-1}^e) - \mathcal{M}\mathbf{b}^e \quad (8.151)$$

where the matrices $\{\mathcal{A}_o, \mathcal{A}_1, \mathcal{A}_2\}$ do not appear since, in this case, $A_o = A_1 = A_2 = I_N$. We summarize the discussion so far in the following statement for complex data (we show how these results simplify for real data in the example after the lemma).

Error Dynamics



Lemma 8.1 (Network error dynamics). Consider a network of N interacting agents running the distributed strategy (8.46). The evolution of the error dynamics across the network relative to the reference vector w^* defined by (8.55) is described by the following recursion:

$$\tilde{w}_i^e = \mathcal{B}_{i-1} \tilde{w}_{i-1}^e + \mathcal{A}_2^\top \mathcal{M} s_i^e(w_{i-1}^e) - \mathcal{A}_2^\top \mathcal{M} b^e, \quad i \geq 0 \quad (8.152)$$

where

Error Dynamics



$$\mathcal{B}_{i-1} \triangleq \mathcal{A}_2^\top (\mathcal{A}_o^\top - \mathcal{M}\mathcal{H}_{i-1}) \mathcal{A}_1^\top \quad (8.153)$$

$$\mathcal{A}_o \triangleq A_o \otimes I_{2M}, \quad \mathcal{A}_1 \triangleq A_1 \otimes I_{2M}, \quad \mathcal{A}_2 \triangleq A_2 \otimes I_{2M} \quad (8.154)$$

$$\mathcal{M} \triangleq \text{diag}\{\mu_1 I_{2M}, \mu_2 I_{2M}, \dots, \mu_N I_{2M}\} \quad (8.155)$$

$$\mathcal{H}_{i-1} \triangleq \text{diag}\{\mathbf{H}_{1,i-1}, \mathbf{H}_{2,i-1}, \dots, \mathbf{H}_{N,i-1}\} \quad (8.156)$$

$$\mathbf{H}_{k,i-1} \triangleq \int_0^1 \nabla_w^2 J_k(w^\star - t\tilde{\phi}_{k,i-1}) dt \quad (8.157)$$

where $\nabla_w^2 J_k(w)$ denotes the $2M \times 2M$ Hessian matrix of $J_k(w)$ relative to w . Moreover, the extended vectors $\{\tilde{\mathbf{w}}_i^e, \mathbf{s}_i^e, b^e\}$ are defined by (8.143) and (8.144).



Example #8.12

42

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Example 8.12 (Mean-square-error costs). Let us re-consider the scenario studied in Example 8.1 and verify that result (8.152) collapses to (8.25). Indeed, in this case, we have $w^* = w^o$ and the bias vector, b_k^e , will be zero for all agents $k = 1, 2, \dots, N$. Moreover since the Hessian matrix is now block diagonal, we can easily verify from the definition (8.137) that

$$\mathbf{H}_{k,i-1} = \begin{bmatrix} R_{u,k} & 0 \\ 0 & R_{u,k}^\top \end{bmatrix} \quad (8.158)$$

Substituting these facts into the expressions in Lemma 8.1 we recover (8.25). ■

Example #8.13



Example 8.13 (Simplifications in the real case). The network error model of Lemma 8.1 can be simplified in the case of real data. This is because when $w \in \mathbb{R}^M$ is real-valued, we do not need to introduce the extended vectors (8.133) and (8.134) any longer. The simplifications that occur are described below.

To begin with, the distributed strategy (8.46) will be given by

$$\left\{ \begin{array}{lcl} \phi_{k,i-1} & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \mathbf{w}_{\ell,i-1} \\ \psi_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \phi_{\ell,i-1} - \mu_k \widehat{\nabla_{w^\top} J}_k (\phi_{k,i-1}) \\ \mathbf{w}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \psi_{\ell,i} \end{array} \right. \quad (8.159)$$

Example #8.13



where the gradient vector approximation in the second equation is now relative to w^\top and not w^* . Subtracting the limit vector w^* directly from both sides of the above equations gives

$$\left\{ \begin{array}{lcl} \tilde{\phi}_{k,i-1} & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \tilde{w}_{\ell,i-1} \\ \tilde{\psi}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \tilde{\phi}_{\ell,i-1} + \mu_k \nabla_{w^\top} J_k(\phi_{k,i-1}) + \mu_k s_{k,i} \\ \tilde{w}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \tilde{\psi}_{\ell,i} \end{array} \right. \quad (8.160)$$

Example #8.13



where now

$$\mathbf{s}_{k,i} \triangleq \widehat{\nabla_{w^\top} J_k}(\boldsymbol{\phi}_{k,i-1}) - \nabla_{w^\top} J_k(\boldsymbol{\phi}_{k,i-1}) \quad (8.161)$$

and the error vectors are measured relative to the same limit vector w^* :

$$\tilde{\mathbf{w}}_{k,i} = w^* - \mathbf{w}_{k,i}, \quad \tilde{\boldsymbol{\psi}}_{k,i} = w^* - \boldsymbol{\psi}_{k,i}, \quad \tilde{\boldsymbol{\phi}}_{k,i-1} = w^* - \boldsymbol{\phi}_{k,i-1} \quad (8.162)$$

We then call upon the real-version of the mean-value theorem, namely, expression (D.9) in the appendix, to write

Example #8.13



$$\begin{aligned}
 \nabla_{w^\top} J_k(\phi_{k,i-1}) &= \underbrace{\nabla_{w^\top} J_k(w^*)}_{\stackrel{\Delta}{=} -b_k} - \underbrace{\left[\int_0^1 \nabla_w^2 J_k(w^* - t\tilde{\phi}_{k,i-1}) dt \right]}_{\stackrel{\Delta}{=} \mathbf{H}_{k,i-1}} \tilde{\phi}_{k,i-1} \\
 &= -b_k - \mathbf{H}_{k,i-1} \tilde{\phi}_{k,i-1} \tag{8.163}
 \end{aligned}$$

where we introduced the $M \times 1$ constant vector b_k and the (now) $M \times M$ stochastic matrix $\mathbf{H}_{k,i}$. Substituting (8.163) into (8.160) leads to

Example #8.13



$$\left\{ \begin{array}{lcl} \tilde{\phi}_{k,i-1} & = & \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \tilde{w}_{\ell,i-1} \\ \tilde{\psi}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{o,\ell k} \tilde{\phi}_{\ell,i-1} - \mu_k \mathbf{H}_{k,i-1} \tilde{\phi}_{k,i-1} - \mu_k b_k + \mu_k s_{k,i} \\ \tilde{w}_{k,i} & = & \sum_{\ell \in \mathcal{N}_k} a_{2,\ell k} \tilde{\psi}_{\ell,i} \end{array} \right. \quad (8.164)$$

so that the network error vector

$$\tilde{w}_i = \text{col}\{\tilde{w}_{1,i}, \tilde{w}_{2,i}, \dots, \tilde{w}_{N,i}\} \quad (8.165)$$

evolves according to the recursion

$$\tilde{w}_i = \mathcal{B}_{i-1} \tilde{w}_{i-1} + \mathcal{A}_2^\top \mathcal{M} s_i(w_{i-1}) - \mathcal{A}_2^\top \mathcal{M} b, \quad i \geq 0 \quad (8.166)$$

where now

Example #8.13



$$\mathcal{B}_{i-1} \triangleq \mathcal{A}_2^\top (\mathcal{A}_o^\top - \mathcal{M} \mathcal{H}_{i-1}) \mathcal{A}_1^\top \quad (8.167)$$

$$\mathcal{A}_o \triangleq A_o \otimes I_M, \quad \mathcal{A}_1 \triangleq A_1 \otimes I_M, \quad \mathcal{A}_2 \triangleq A_2 \otimes I_M \quad (8.168)$$

$$\mathcal{M} \triangleq \text{diag}\{\mu_1 I_M, \mu_2 I_M, \dots, \mu_N I_M\} \quad (8.169)$$

$$\mathcal{H}_{i-1} \triangleq \text{diag}\{\mathbf{H}_{1,i-1}, \mathbf{H}_{2,i-1}, \dots, \mathbf{H}_{N,i-1}\} \quad (8.170)$$

$$\mathbf{H}_{k,i-1} \triangleq \int_0^1 \nabla_w^2 J_k(w^\star - t\tilde{\phi}_{k,i-1}) dt \quad (8.171)$$

$$\mathbf{w}_{i-1} \triangleq \text{col}\{\mathbf{w}_{1,i-1}, \mathbf{w}_{2,i-1}, \dots, \mathbf{w}_{N,i-1}\} \quad (8.172)$$

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



Gradient Noise Model



Gradient Noise + Bias Vector

Moreover, with each agent k , we associate a gradient noise vector in addition to a mismatch (or bias) vector, namely,

$$s_{k,i}(\phi_{k,i-1}) \triangleq \widehat{\nabla_{w^*} J_k}(\phi_{k,i-1}) - \nabla_{w^*} J_k(\phi_{k,i-1}) \quad (8.109)$$

and

$$b_k \triangleq -\nabla_{w^*} J_k(w^*) \quad (8.110)$$



Gradient Noise Model

In the special case when all individual costs, $J_k(w)$, have the same minimizer at $w_k^o \equiv w^o$ (which is the situation considered in Example 8.1 over MSE networks), then $w^* = w^o$ and the vector b_k will be identically zero. In general, though, the vector b_k is nonzero. Let \mathcal{F}_{i-1} represent the collection of all random events generated by the processes $\{\mathbf{w}_{k,j}\}$ at all agents $k = 1, 2, \dots, N$ up to time $i - 1$:

$$\mathcal{F}_{i-1} \triangleq \text{filtration}\{\mathbf{w}_{k,-1}, \mathbf{w}_{k,0}, \mathbf{w}_{k,1}, \dots, \mathbf{w}_{k,i-1}, \text{ all } k\} \quad (8.111)$$



Gradient Noise Assumptions

Assumption 8.1 (Conditions on gradient noise). It is assumed that the first and second-order conditional moments of the individual gradient noise processes, $s_{k,i}(\phi)$, satisfy the following conditions for any iterates $\phi \in \mathcal{F}_{i-1}$ and for all $k, \ell = 1, 2, \dots, N$:

$$\mathbb{E} [s_{k,i}(\phi) | \mathcal{F}_{i-1}] = 0 \quad (8.112)$$

$$\mathbb{E} [s_{k,i}(\phi) s_{\ell,i}^*(\phi) | \mathcal{F}_{i-1}] = 0, \quad k \neq \ell \quad (8.113)$$

$$\mathbb{E} [s_{k,i}(\phi) s_{\ell,i}^\top(\phi) | \mathcal{F}_{i-1}] = 0, \quad k \neq \ell \quad (8.114)$$

$$\mathbb{E} [\|s_{k,i}(\phi)\|^2 | \mathcal{F}_{i-1}] \leq (\bar{\beta}_k/h)^2 \|\phi\|^2 + \bar{\sigma}_{s,k}^2 \quad (8.115)$$

almost surely, for some nonnegative scalars $\bar{\beta}_k^2$ and $\bar{\sigma}_{s,k}^2$ and where $h = 1$ for real data and $h = 2$ for complex data.



Gradient Noise Model

Using the above conditions, and in a manner similar to the derivation (3.28), it is straightforward to verify that the gradient noise processes satisfy:

$$\mathbb{E} \left[\mathbf{s}_{k,i}(\phi_{k,i-1}) \mid \mathcal{F}_{i-1} \right] = 0 \quad (8.116)$$

$$\mathbb{E} \left[\|\mathbf{s}_{k,i}(\phi_{k,i-1})\|^2 \mid \mathcal{F}_{i-1} \right] \leq (\beta_k^2/h^2) \|\tilde{\phi}_{k,i-1}\|^2 + \sigma_{s,k}^2 \quad (8.117)$$

$$\mathbb{E} \|\mathbf{s}_{k,i}(\phi_{k,i-1})\|^2 \leq (\beta_k^2/h^2) \mathbb{E} \|\tilde{\phi}_{k,i-1}\|^2 + \sigma_{s,k}^2 \quad (8.118)$$

in terms of the scalars

$$\beta_k^2 \triangleq 2\bar{\beta}_k^2 \quad (8.119)$$

$$\sigma_{s,k}^2 \triangleq 2(\bar{\beta}_k/h)^2 \|w^\star\|^2 + \bar{\sigma}_{s,k}^2 \quad (8.120)$$

Gradient Noise Model



We shall use conditions (8.116)–(8.118) more frequently in lieu of (8.112)–(8.115). We could have required these conditions directly in the statement of Assumption 8.1. We instead opted to state conditions (8.112)–(8.115) in that manner, in terms of a generic $\phi \in \mathcal{F}_{i-1}$ rather than $\tilde{w}_{k,i-1}$, so that the upper bound in (8.115) is independent of the unknown w^* .



Gradient Noise Model

Conditions (8.116)–(8.118) will be useful in establishing the mean-square stability of the second-order moment of the error vector, $\mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2$, in the next chapter. Later, in Sec. 9.2, when we examine the stability of the fourth-order moment of the same error vector, $\mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^4$, we will need to replace the bound (8.115) by a condition similar to (5.36) on the fourth-order moments of the individual gradient noise processes, namely, by the following condition:

$$\mathbb{E} \left[\|\mathbf{s}_{k,i}(\phi)\|^4 \mid \mathcal{F}_{i-1} \right] \leq (\bar{\beta}_k/h)^4 \|\phi\|^4 + \bar{\sigma}_{s,k}^4 \quad (8.121)$$

Gradient Noise Model



almost surely, for nonnegative scalars $\{\bar{\beta}_k^4, \bar{\sigma}_{s,k}^4\}$. Using an argument similar to (3.56), we can similarly conclude from these conditions that

$$\mathbb{E} \left[\|s_{k,i}(\phi_{k,i-1})\|^4 \mid \mathcal{F}_{i-1} \right] \leq (\beta_{4,k}^4/h^4) \|\tilde{\phi}_{k,i-1}\|^4 + \sigma_{s4,k}^4 \quad (8.122)$$

for some non-negative parameters defined by:

$$\beta_{4,k}^4 \triangleq 8\bar{\beta}_k^4 \quad (8.123)$$

$$\sigma_{s4,k}^4 \triangleq 8(\bar{\beta}_k/h)^4 \|w^\star\|^4 + \bar{\sigma}_{s4,k}^4 \quad (8.124)$$



Gradient Noise Model

We will not need to introduce condition (8.121) in addition to the second-order moment condition (8.115). This is because, as explained earlier following (3.50), condition (8.121) implies that condition (8.115) also holds, namely, it follows from (8.121) that

$$\mathbb{E} \left[\|s_{k,i}(\phi)\|^2 \mid \mathcal{F}_{i-1} \right] \leq (\bar{\beta}_k/h)^2 \|\phi\|^2 + \bar{\sigma}_{s,k}^2 \quad (8.125)$$

Stability Results



Key Stability Results

We are now ready to examine the stability of the mean-error process, $\mathbb{E} \tilde{\mathbf{w}}_i$, the mean-square-error, $\mathbb{E} \|\tilde{\mathbf{w}}_i\|^2$, and the fourth-order moment, $\mathbb{E} \|\tilde{\mathbf{w}}_i\|^4$, by using the network error recursion (8.152). The key results proven in the current chapter are that for sufficiently small step-sizes, and for each agent k , it will hold that

$$\limsup_{i \rightarrow \infty} \|\mathbb{E} \tilde{\mathbf{w}}_{k,i}\| = O(\mu_{\max}) \quad (9.1)$$

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 = O(\mu_{\max}) \quad (9.2)$$

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^4 = O(\mu_{\max}^2) \quad (9.3)$$



Network Stability

60

Lecture #17: Stability of Multi-Agent Networks

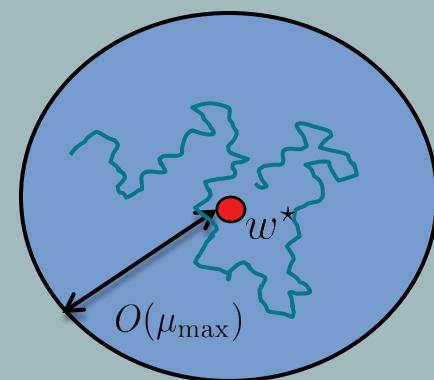
EE210B: Inference over Networks (A. H. Sayed)

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

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

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

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





Setting

61

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where μ_{\max} is an upper bound on the largest step-size parameter across the network since, from (8.52), we parameterized all step-sizes as scaled multiples of μ_{\max} , namely,

$$\mu_k \stackrel{\Delta}{=} \tau_k \mu_{\max}, \quad k = 1, 2, \dots, N \quad (9.4)$$

where $0 < \tau_k \leq 1$. The error vectors, $\{\tilde{w}_{k,i}\}$, in the above expressions are measured relative to the limit vector, w^* :

$$\tilde{w}_{k,i} = w^* - w_{k,i} \quad (9.5)$$



Setting

62

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where w^* was defined by (8.55) as the unique minimum of the weighted aggregate cost function, $J^{\text{glob},*}(w)$, from (8.53), namely,

$$J^{\text{glob},*}(w) \triangleq \sum_{k=1}^N q_k J_k(w) \quad (9.6)$$

and the $\{q_k\}$ are positive scalars corresponding to the entries of the vector:

$$q \triangleq \text{diag}\{\mu_1, \mu_2, \dots, \mu_N\} A_2 p \quad (9.7)$$



Setting

63

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Here, the vector p refers to the Perron eigenvector of the matrix product

$$P \triangleq A_1 A_o A_2 \quad (9.8)$$

and is defined through the relations:

$$Pp = p, \quad \mathbf{1}^T p = 1, \quad p_k > 0 \quad (9.9)$$

For ease of reference, we recall the definition of the original aggregate cost function (8.44), namely,

$$J^{\text{glob}}(w) \triangleq \sum_{k=1}^N J_k(w) \quad (9.10)$$

Second-Order Stability

Recall: Error Dynamics



$$\tilde{\mathbf{w}}_i^e = \mathcal{B}_{i-1} \tilde{\mathbf{w}}_{i-1}^e + \mathcal{A}_2^\top \mathcal{M} s_i^e(\mathbf{w}_{i-1}^e) - \mathcal{A}_2^\top \mathcal{M} b^e, \quad i \geq 0 \quad (8.152)$$

$$\mathcal{B}_{i-1} \triangleq \mathcal{A}_2^\top (\mathcal{A}_o^\top - \mathcal{M} \mathcal{H}_{i-1}) \mathcal{A}_1^\top \quad (8.153)$$

$$\mathcal{A}_o \triangleq A_o \otimes I_{2M}, \quad \mathcal{A}_1 \triangleq A_1 \otimes I_{2M}, \quad \mathcal{A}_2 \triangleq A_2 \otimes I_{2M} \quad (8.154)$$

$$\mathcal{M} \triangleq \text{diag}\{ \mu_1 I_{2M}, \mu_2 I_{2M}, \dots, \mu_N I_{2M} \} \quad (8.155)$$

$$\mathcal{H}_{i-1} \triangleq \text{diag}\{ \mathbf{H}_{1,i-1}, \mathbf{H}_{2,i-1}, \dots, \mathbf{H}_{N,i-1} \} \quad (8.156)$$

$$\mathbf{H}_{k,i-1} \triangleq \int_0^1 \nabla_w^2 J_k(w^* - t\tilde{\phi}_{k,i-1}) dt \quad (8.157)$$

Second-Order Error Moment



Theorem 9.1 (Network mean-square-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 further that the first and second-order moments of the gradient noise process satisfy the conditions in Assumption 8.1. Then, the network is mean-square stable for sufficiently small step-sizes, namely, it holds that

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{w}_{k,i}\|^2 = O(\mu_{\max}), \quad k = 1, 2, \dots, N \quad (9.11)$$

for any $\mu_{\max} < \mu_o$, for some small enough μ_o .

Proof



67

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Proof. The derivation is demanding. We follow arguments motivated by the analysis in [70, 277] and they involve, as an initial step, transforming the error recursion (9.12) shown below into a more convenient form shown later in (9.60). We establish the result for the general case of complex data and, therefore, $h = 2$ throughout this derivation.

We start from the network error recursion (8.152):

$$\tilde{\mathbf{w}}_i^e = \mathcal{B}_{i-1} \tilde{\mathbf{w}}_{i-1}^e + \mathcal{A}_2^\top \mathcal{M} s_i^e(\mathbf{w}_{i-1}^e) - \mathcal{A}_2^\top \mathcal{M} b^e, \quad i \geq 0 \quad (9.12)$$

where



Proof

68

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}\mathcal{B}_{i-1} &= \mathcal{A}_2^\top (\mathcal{A}_o^\top - \mathcal{M}\mathcal{H}_{i-1}) \mathcal{A}_1^\top \\ &= \mathcal{A}_2^\top \mathcal{A}_o^\top \mathcal{A}_1^\top - \mathcal{A}_2^\top \mathcal{M}\mathcal{H}_{i-1} \mathcal{A}_1^\top \\ &\triangleq \mathcal{P}^\top - \mathcal{A}_2^\top \mathcal{M}\mathcal{H}_{i-1} \mathcal{A}_1^\top\end{aligned}\tag{9.13}$$

in terms of the matrix

$$\begin{aligned}\mathcal{P}^\top &\triangleq \mathcal{A}_2^\top \mathcal{A}_o^\top \mathcal{A}_1^\top \\ &= (A_2^\top \otimes I_{2M})(A_o^\top \otimes I_{2M})(A_1^\top \otimes I_{2M}) \\ &= (A_2^\top A_o^\top A_1^\top \otimes I_{2M}) \\ &= P^\top \otimes I_{2M}\end{aligned}\tag{9.14}$$



Proof

69

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

The matrix $P = A_1 A_o A_2$ is left-stochastic and assumed primitive. It follows that it has a single eigenvalue at one while all other eigenvalues are strictly inside the unit circle. We let p denote its Perron eigenvector, which is already defined by (9.9). This vector determines the entries of q defined by (9.7). Note, for later reference, that the k -entry of q can be extracted by computing the inner product of q with the k -th basis vector, e_k , which has a unit entry at the k -th location and zeros elsewhere, i.e.,

$$\begin{aligned} q_k &= \mu_k (e_k^\top A_2 p) \\ &\stackrel{(9.4)}{=} \mu_{\max} \tau_k (e_k^\top A_2 p) \end{aligned} \tag{9.15}$$

Proof



70

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Obviously, it holds for the extended matrices $\{\mathcal{P}, \mathcal{A}_2\}$ that

$$\mathcal{P}(p \otimes I_{2M}) = (p \otimes I_{2M}) \quad (9.16)$$

$$\mathcal{MA}_2(p \otimes I_{2M}) = (q \otimes I_{2M}) \quad (9.17)$$

$$(\mathbb{1}^\top \otimes I_{2M})(p \otimes I_{2M}) = I_{2M} \quad (9.18)$$

Moreover, since A_1 and A_2 are left-stochastic, it holds that

$$\mathcal{A}_1^\top(\mathbb{1} \otimes I_{2M}) = (\mathbb{1} \otimes I_{2M}) \quad (9.19)$$

$$\mathcal{A}_2^\top(\mathbb{1} \otimes I_{2M}) = (\mathbb{1} \otimes I_{2M}) \quad (9.20)$$

Proof



71

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

The derivation that follows exploits the eigen-structure of P . We start by noting that the $N \times N$ matrix P admits a Jordan canonical decomposition of the form [113, p.128]:

$$P \triangleq V_\epsilon J V_\epsilon^{-1} \quad (9.21)$$

$$J = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & J_\epsilon \end{array} \right] \quad (9.22)$$

$$V_\epsilon = \left[\begin{array}{c|c} p & V_R \end{array} \right] \quad (9.23)$$

$$V_\epsilon^{-1} = \left[\begin{array}{c} \mathbf{1}^\top \\ \hline V_L^\top \end{array} \right] \quad (9.24)$$



Proof

72

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where the matrix J_ϵ consists of Jordan blocks, with each one of them having the generic form (say, for a Jordan block of size 4×4):

$$\begin{bmatrix} \lambda & & & \\ \epsilon & \lambda & & \\ & \epsilon & \lambda & \\ & & \epsilon & \lambda \end{bmatrix} \quad (9.25)$$

with $\epsilon > 0$ appearing on the lower¹ diagonal, and where the eigenvalue λ may be complex but has magnitude strictly less than one. The scalar ϵ is any small positive number that is independent of μ_{\max} . Obviously, since $V_\epsilon^{-1}V_\epsilon = I_N$, it holds that

Proof



¹For any $N \times N$ matrix A , the traditional Jordan decomposition $A = TJ'T^{-1}$ involves Jordan blocks in J' that have ones on the lower diagonal instead of ϵ . However, if we introduce the diagonal matrix $E = \text{diag}\{1, \epsilon, \epsilon^2, \dots, \epsilon^{N-1}\}$, then $A = TE^{-1}EJ'E^{-1}ET^{-1}$, which we rewrite as $A = V_\epsilon J V_\epsilon^{-1}$ with $V_\epsilon = TE^{-1}$ and $J = EJ'E^{-1}$. The matrix J now has ϵ values instead of ones on the lower diagonal.



Proof

74

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mathbf{1}^\top V_R = 0 \quad (9.26)$$

$$V_L^\top p = 0 \quad (9.27)$$

$$V_L^\top V_R = I_{N-1} \quad (9.28)$$

The matrices $\{V_\epsilon, J, V_\epsilon^{-1}\}$ have dimensions $N \times N$ while the matrices $\{V_L, J_\epsilon, V_R\}$ have dimensions $(N - 1) \times (N - 1)$. The Jordan decomposition of the extended matrix $\mathcal{P} = P \otimes I_{2M}$ is given by

$$\mathcal{P} = (V_\epsilon \otimes I_{2M})(J \otimes I_{2M})(V_\epsilon^{-1} \otimes I_{2M}) \quad (9.29)$$

so that substituting into (9.13) we obtain

$$\mathcal{B}_{i-1} = ((V_\epsilon^{-1})^\top \otimes I_{2M}) \left\{ (J^\top \otimes I_{2M}) - \mathcal{D}_{i-1}^\top \right\} (V_\epsilon^\top \otimes I_{2M}) \quad (9.30)$$



Proof

75

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where

$$\begin{aligned}\mathcal{D}_{i-1}^T &\triangleq \left(V_\epsilon^T \otimes I_{2M} \right) \mathcal{A}_2^T \mathcal{M} \mathcal{H}_{i-1} \mathcal{A}_1^T \left((V_\epsilon^{-1})^T \otimes I_{2M} \right) \\ &\equiv \left[\begin{array}{cc} \mathbf{D}_{11,i-1}^T & \mathbf{D}_{21,i-1}^T \\ \mathbf{D}_{12,i-1}^T & \mathbf{D}_{22,i-1}^T \end{array} \right] \end{aligned} \quad (9.31)$$

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

$$\mathcal{A}_1 = A_1 \otimes I_{2M}, \quad \mathcal{A}_2 = A_2 \otimes I_{2M} \quad (9.32)$$

we find that the block entries $\{\mathbf{D}_{mn,i-1}\}$ in (9.31) are given by



Proof

76

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mathbf{D}_{11,i-1} = \sum_{k=1}^N q_k \mathbf{H}_{k,i-1}^\top \quad (9.33)$$

$$\mathbf{D}_{12,i-1} = (\mathbf{1}^\top \otimes I_{2M}) \mathcal{H}_{i-1}^\top \mathcal{M} (A_2 V_R \otimes I_{2M}) \quad (9.34)$$

$$\mathbf{D}_{21,i-1} = (V_L^\top A_1 \otimes I_{2M}) \mathcal{H}_{i-1}^\top (q \otimes I_{2M}) \quad (9.35)$$

$$\mathbf{D}_{22,i-1} = (V_L^\top A_1 \otimes I_{2M}) \mathcal{H}_{i-1}^\top \mathcal{M} (A_2 V_R \otimes I_{2M}) \quad (9.36)$$

Let us now show that the entries in each of these matrices is in the order of $O(\mu_{\max})$, as well as verify that the matrix norm sequences of these matrices are uniformly bounded from above for all i . To begin with, recall from (8.157) that

$$\mathbf{H}_{k,i-1} \triangleq \int_0^1 \nabla_w^2 J_k(w^\star - t\tilde{\phi}_{k,i-1}) dt \quad (9.37)$$

Proof



77

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

and, moreover, by assumption, all individual costs $J_k(w)$ are convex functions with at least one of them, say, the cost function of index k_o , being ν_d -strongly-convex. This fact implies that, for any w ,

$$\nabla_w^2 J_{k_o}(w) \geq \frac{\nu_d}{h} I_{hM} > 0, \quad \nabla_w^2 J_k(w) \geq 0, \quad k \neq k_o \quad (9.38)$$

Consequently,

$$\mathbf{H}_{k_o,i-1} \geq \frac{\nu_d}{h} I_{hM} > 0, \quad \mathbf{H}_{k,i-1} \geq 0, \quad k \neq k_o \quad (9.39)$$

and, therefore, $\mathbf{D}_{11,i-1} > 0$. More specifically, the matrix sequence $\mathbf{D}_{11,i-1}$ is uniformly bounded from below as follows:

Proof



$$\begin{aligned}
 D_{11,i-1} &\geq q_{k_o} \frac{\nu_d}{h} I_{hM} \\
 &\stackrel{(9.15)}{=} \mu_{\max} \tau_{k_o} (e_{k_o}^\top A_2 p) \frac{\nu_d}{h} I_{hM} \\
 &= O(\mu_{\max})
 \end{aligned} \tag{9.40}$$

On the other hand, from the upper bound on the sum of the Hessian matrices in (6.13), and since each individual Hessian matrix is at least non-negative definite, we get

$$H_{k,i-1} \leq \frac{\delta_d}{h} I_{hM} \tag{9.41}$$

so that the matrix sequence $D_{11,i-1}$ is uniformly bounded from above as well:



Proof

79

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} \mathbf{D}_{11,i-1} &\leq q_{\max} N \frac{\delta_d}{h} I_{hM} \\ &\stackrel{(9.15)}{=} \mu_{\max} \tau_{k_{\max}} (e_{k_{\max}}^T A_2 p) N \frac{\delta_d}{h} I_{hM} \\ &= O(\mu_{\max}) \end{aligned} \tag{9.42}$$

where k_{\max} denotes the k -index of the largest q_k entry. Combining results (9.40)–(9.42) we conclude that

$$\mathbf{D}_{11,i-1} = O(\mu_{\max}) \tag{9.43}$$

Actually, since $\mathbf{D}_{11,i-1}$ is Hermitian positive-definite, we also conclude that its eigenvalues (which are positive and real) are $O(\mu_{\max})$. This is because from the relation



Proof

80

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mu_{\max} \tau_{k_o} (e_{k_o}^T A_2 p) \frac{\nu_d}{h} I_{hM} \leq D_{11,i-1} \leq \mu_{\max} \tau_{k_{\max}} (e_{k_{\max}}^T A_2 p) N \frac{\delta_d}{h} I_{hM} \quad (9.44)$$

we can write, more compactly,

$$c_1 \mu_{\max} I_{hM} \leq D_{11,i-1} \leq c_2 \mu_{\max} I_{hM} \quad (9.45)$$

for some positive constants c_1 and c_2 that are independent of μ_{\max} and i . Accordingly, for the eigenvalues of $D_{11,i-1}$, we can write

$$c_1 \mu_{\max} \leq \lambda(D_{11,i-1}) \leq c_2 \mu_{\max} \quad (9.46)$$

It follows that the eigenvalues of $I_{2M} - D_{11,i-1}^T$ are $1 - O(\mu_{\max})$ so that, in terms of the 2-induced norm and for sufficiently small μ_{\max} :



Proof

81

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}\|I_{2M} - \mathbf{D}_{11,i-1}^\top\| &= \rho(I_{2M} - \mathbf{D}_{11,i-1}^\top) \\ &\leq 1 - \sigma_{11}\mu_{\max} \\ &= 1 - O(\mu_{\max})\end{aligned}\tag{9.47}$$

for some positive constant σ_{11} that is independent of μ_{\max} and i .

Similarly, from (9.39) and (9.41), and since each $\mathbf{H}_{k,i-1}$ is bounded from below and from above, we can conclude that

$$\mathbf{D}_{12,i-1} = O(\mu_{\max}), \quad \mathbf{D}_{21,i-1} = O(\mu_{\max}), \quad \mathbf{D}_{22,i-1} = O(\mu_{\max})\tag{9.48}$$

and that the norms of these matrix sequences are also uniformly bounded from above. For example, using the 2-induced norm (i.e., maximum singular value):



Proof

$$\begin{aligned}\|D_{21,i-1}\| &\leq \|V_L^\top A_1 \otimes I_{2M}\| \|q \otimes I_{2M}\| \|\mathcal{H}_{i-1}^\top\| \\ &\leq \|V_L^\top A_1 \otimes I_{2M}\| \|q \otimes I_{2M}\| \left(\max_{1 \leq k \leq N} \|\mathbf{H}_{k,i-1}\| \right) \\ &\stackrel{(9.41)}{\leq} \|V_L^\top A_1 \otimes I_{2M}\| \|q \otimes I_{2M}\| \left(\frac{\delta_d}{h} \right) \\ &= \|V_L^\top A_1 \otimes I_{2M}\| \|q\| \left(\frac{\delta_d}{h} \right) \\ &\leq \|V_L^\top A_1 \otimes I_{2M}\| \sqrt{N q_{\max}^2} \left(\frac{\delta_d}{h} \right) \\ &= \|V_L^\top A_1 \otimes I_{2M}\| \sqrt{N} \mu_{\max} \tau_{k_{\max}} (e_{k_{\max}}^\top A_2 p) \left(\frac{\delta_d}{h} \right)\end{aligned}\tag{9.49}$$



Proof

83

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

so that

$$\|\mathbf{D}_{21,i-1}\| \leq \sigma_{21}\mu_{\max} = O(\mu_{\max}) \quad (9.50)$$

for some positive constant σ_{21} . In the above derivation we used the fact that $\|q \otimes I_{2M}\| = \|q\|$ since, from Table F.1 in the appendix, the singular values of a Kronecker product are given by all possible products of the singular values of the individual matrices. A similar argument applies to $\mathbf{D}_{12,i-1}$ and $\mathbf{D}_{22,i-1}$ for which we can verify that

$$\|\mathbf{D}_{12,i-1}\| \leq \sigma_{12}\mu_{\max} = O(\mu_{\max}), \quad \|\mathbf{D}_{22,i-1}\| \leq \sigma_{22}\mu_{\max} = O(\mu_{\max}) \quad (9.51)$$

for some positive constants σ_{21} and σ_{22} . Let

$$\mathcal{V}_\epsilon \stackrel{\Delta}{=} V_\epsilon \otimes I_{2M}, \quad \mathcal{J}_\epsilon \stackrel{\Delta}{=} J_\epsilon \otimes I_{2M} \quad (9.52)$$



Proof: Recall

84

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mathcal{B}_{i-1} = \left((V_\epsilon^{-1})^\top \otimes I_{2M} \right) \left\{ (J^\top \otimes I_{2M}) - \mathcal{D}_{i-1}^\top \right\} \left(V_\epsilon^\top \otimes I_{2M} \right) \quad (9.30)$$

where

$$\begin{aligned} \mathcal{D}_{i-1}^\top &\triangleq \left(V_\epsilon^\top \otimes I_{2M} \right) \mathcal{A}_2^\top \mathcal{M} \mathcal{H}_{i-1} \mathcal{A}_1^\top \left((V_\epsilon^{-1})^\top \otimes I_{2M} \right) \\ &\equiv \begin{bmatrix} D_{11,i-1}^\top & D_{21,i-1}^\top \\ D_{12,i-1}^\top & D_{22,i-1}^\top \end{bmatrix} \end{aligned} \quad (9.31)$$



Proof

Then, using (9.30), we can write

$$\mathcal{B}_{i-1} = (\mathcal{V}_\epsilon^{-1})^\top \begin{bmatrix} I_{2M} - \mathbf{D}_{11,i-1}^\top & -\mathbf{D}_{21,i-1}^\top \\ -\mathbf{D}_{12,i-1}^\top & \mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top \end{bmatrix} \mathcal{V}_\epsilon^\top \quad (9.53)$$

To simplify the notation, we drop the argument \mathbf{w}_{i-1}^e in (9.12) and write \mathbf{s}_i^e instead of $\mathbf{s}_i^e(\mathbf{w}_{i-1}^e)$ from this point onwards. We now multiply both sides of the error recursion (9.12) from the left by $\mathcal{V}_\epsilon^\top$:

$$\mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_i^e = \mathcal{V}_\epsilon^\top \mathcal{B}_{i-1} (\mathcal{V}_\epsilon^{-1})^\top \mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_{i-1}^e + \mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \mathcal{M} \mathbf{s}_i^e - \mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \mathcal{M} \mathbf{b}^e, \quad i \geq 0 \quad (9.54)$$

Proof



and let

$$\mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_i^e = \begin{bmatrix} (p^\top \otimes I_{2M}) \tilde{\mathbf{w}}_i^e \\ (V_R^\top \otimes I_{2M}) \tilde{\mathbf{w}}_i^e \end{bmatrix} \triangleq \begin{bmatrix} \bar{\mathbf{w}}_i^e \\ \check{\mathbf{w}}_i^e \end{bmatrix} \quad (9.55)$$

$$\mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \mathcal{M} \mathbf{s}_i^e = \begin{bmatrix} (p^\top \otimes I_{2M}) \mathcal{A}_2^\top \mathcal{M} \mathbf{s}_i^e \\ (V_R^\top \otimes I_{2M}) \mathcal{A}_2^\top \mathcal{M} \mathbf{s}_i^e \end{bmatrix} \triangleq \begin{bmatrix} \bar{\mathbf{s}}_i^e \\ \check{\mathbf{s}}_i^e \end{bmatrix} \quad (9.56)$$

$$\mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \mathcal{M} \mathbf{b}^e = \begin{bmatrix} (p^\top \otimes I_{2M}) \mathcal{A}_2^\top \mathcal{M} \mathbf{b}^e \\ (V_R^\top \otimes I_{2M}) \mathcal{A}_2^\top \mathcal{M} \mathbf{b}^e \end{bmatrix} \triangleq \begin{bmatrix} 0 \\ \check{\mathbf{b}}^e \end{bmatrix} \quad (9.57)$$

where the zero entry in the last equality is due to the fact that



Proof

87

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} (p^\top \otimes I_{2M}) \mathcal{A}_2^\top \mathcal{M} b^e &= (q^\top \otimes I_{2M}) b^e \\ &= \sum_{k=1}^N q_k b_k^e \\ &= - \sum_{k=1}^N q_k \begin{bmatrix} \nabla_{w^*} J_k(w^*) \\ \nabla_{w^\top} J_k(w^*) \end{bmatrix} \\ &= - \sum_{k=1}^N q_k \begin{bmatrix} [\nabla_w J_k(w^*)]^* \\ [\nabla_w J_k(w^*)]^\top \end{bmatrix} \\ &\stackrel{(8.55)}{=} 0 \tag{9.58} \end{aligned}$$



Proof

88

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Moreover, from the expression for \check{b}^e in (9.57), we note that it depends on \mathcal{M} and b^e . Recall from (8.110) and (8.144) that the entries of b^e are defined in terms of the gradient vectors $\nabla_{w^*} J_k(w^*)$.

It follows that b^e has bounded norm and we conclude that

$$\check{b}^e = O(\mu_{\max}) \quad (9.59)$$

Using the just introduced transformed variables, we can rewrite (9.54) in the form



Proof

89

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{bmatrix} \bar{\mathbf{w}}_i^e \\ \check{\mathbf{w}}_i^e \end{bmatrix} = \begin{bmatrix} I_{2M} - D_{11,i-1}^\top & -D_{21,i-1}^\top \\ -D_{12,i-1}^\top & \mathcal{J}_\epsilon^\top - D_{22,i-1}^\top \end{bmatrix} \begin{bmatrix} \bar{\mathbf{w}}_{i-1}^e \\ \check{\mathbf{w}}_{i-1}^e \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{s}}_i^e \\ \check{\mathbf{s}}_i^e \end{bmatrix} - \begin{bmatrix} 0 \\ \check{b}^e \end{bmatrix} \quad (9.60)$$

or, in expanded form,

$$\bar{\mathbf{w}}_i^e = (I_{2M} - D_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - D_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \bar{\mathbf{s}}_i^e \quad (9.61)$$

$$\check{\mathbf{w}}_i^e = (\mathcal{J}_\epsilon^\top - D_{22,i-1}^\top) \check{\mathbf{w}}_{i-1}^e - D_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{\mathbf{s}}_i^e - \check{b}^e \quad (9.62)$$

Conditioning both sides on \mathcal{F}_{i-1} , computing the conditional second-order moments, and using the conditions from Assumption 8.1 on the gradient noise process we get

Proof



90

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mathbb{E} \left[\|\bar{\mathbf{w}}_i^e\|^2 \mid \mathcal{F}_{i-1} \right] = \|(I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e\|^2 + \mathbb{E} \left[\|\bar{\mathbf{s}}_i^e\|^2 \mid \mathcal{F}_{i-1} \right] \quad (9.63)$$

and

$$\begin{aligned} \mathbb{E} \left[\|\check{\mathbf{w}}_i^e\|^2 \mid \mathcal{F}_{i-1} \right] &= \|(\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top) \check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e - \check{b}^e\|^2 + \\ &\quad \mathbb{E} \left[\|\check{\mathbf{s}}_i^e\|^2 \mid \mathcal{F}_{i-1} \right] \end{aligned} \quad (9.64)$$

Proof



91

Computing the expectations again we conclude that

$$\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^2 = \mathbb{E} \|(I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e\|^2 + \mathbb{E} \|\bar{\mathbf{s}}_i^e\|^2 \quad (9.65)$$

and

$$\mathbb{E} \|\check{\mathbf{w}}_i^e\|^2 = \mathbb{E} \|(\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top) \check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e - \check{\mathbf{b}}^e\|^2 + \mathbb{E} \|\check{\mathbf{s}}_i^e\|^2 \quad (9.66)$$

Continuing with the first variance (9.65), we can appeal to Jensen's inequality (F.26) from the appendix and apply it to the function $f(x) = \|x\|^2$ to bound the variance as follows:

Proof



92

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}
 & \mathbb{E} \|\bar{\mathbf{w}}_i^e\|^2 \\
 = & \mathbb{E} \left\| (1-t) \frac{1}{1-t} (I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - t \frac{1}{t} \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \right\|^2 + \mathbb{E} \|\bar{s}_i^e\|^2 \\
 \leq & (1-t) \mathbb{E} \left\| \frac{1}{1-t} (I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e \right\|^2 + t \mathbb{E} \left\| \frac{1}{t} \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \right\|^2 + \mathbb{E} \|\bar{s}_i^e\|^2 \\
 \leq & \frac{1}{1-t} \mathbb{E} [\|I_{2M} - \mathbf{D}_{11,i-1}^\top\|^2 \|\bar{\mathbf{w}}_{i-1}^e\|^2] + \frac{1}{t} \mathbb{E} [\|\mathbf{D}_{21,i-1}^\top\|^2 \|\check{\mathbf{w}}_{i-1}^e\|^2] + \mathbb{E} \|\bar{s}_i^e\|^2 \\
 \leq & \frac{(1 - \sigma_{11}\mu_{\max})^2}{1-t} \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \frac{\sigma_{21}^2 \mu_{\max}^2}{t} \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \mathbb{E} \|\bar{s}_i^e\|^2 \tag{9.67}
 \end{aligned}$$

for any arbitrary positive number $t \in (0, 1)$. We select

$$t = \sigma_{11}\mu_{\max} \tag{9.68}$$

Proof



93

Then, the last inequality can be written as

$$\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^2 \leq (1 - \sigma_{11}\mu_{\max})\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \left(\frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}} \right) \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \mathbb{E} \|\bar{\mathbf{s}}_i^e\|^2 \quad (9.69)$$

We now repeat a similar argument for the second variance relation (9.66). Thus, using Jensen's inequality again we have



Proof

94

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mathbb{E} \|\check{w}_i^e\|^2 = \quad (9.70)$$

$$\begin{aligned} &= \mathbb{E} \left\| \mathcal{J}_\epsilon^\top \check{w}_{i-1}^e - [\mathbf{D}_{22,i-1}^\top \check{w}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{w}_{i-1}^e + \check{b}^e] \right\|^2 + \mathbb{E} \|\check{s}_i^e\|^2 \\ &= \mathbb{E} \left\| t \frac{1}{t} \mathcal{J}_\epsilon^\top \check{w}_{i-1}^e - (1-t) \frac{1}{1-t} [\mathbf{D}_{22,i-1}^\top \check{w}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{w}_{i-1}^e + \check{b}^e] \right\|^2 + \mathbb{E} \|\check{s}_i^e\|^2 \\ &\leq \frac{1}{t} \mathbb{E} \left\| \mathcal{J}_\epsilon^\top \check{w}_{i-1}^e \right\|^2 + \frac{1}{1-t} \mathbb{E} \left\| \mathbf{D}_{22,i-1}^\top \check{w}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{w}_{i-1}^e + \check{b}^e \right\|^2 + \mathbb{E} \|\check{s}_i^e\|^2 \end{aligned}$$

for any arbitrary positive number $t \in (0, 1)$. Now note that



Proof

95

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}\|\mathcal{J}_\epsilon^\top \check{\mathbf{w}}_{i-1}^e\|^2 &= (\check{\mathbf{w}}_{i-1}^e)^* (\mathcal{J}_\epsilon^\top)^* \mathcal{J}_\epsilon^\top \check{\mathbf{w}}_{i-1}^e \\ &= (\check{\mathbf{w}}_{i-1}^e)^* (\mathcal{J}_\epsilon \mathcal{J}_\epsilon^*)^\top \check{\mathbf{w}}_{i-1}^e \\ &\leq \rho(\mathcal{J}_\epsilon \mathcal{J}_\epsilon^*) \|\check{\mathbf{w}}_{i-1}^e\|^2\end{aligned}\tag{9.71}$$

where we called upon the Rayleigh-Ritz characterization of the eigenvalues of Hermitian matrices [104, 113], namely,

$$\lambda_{\min}(C) \|x\|^2 \leq x^* C x \leq \lambda_{\max}(C) \|x\|^2\tag{9.72}$$

Proof



for any Hermitian matrix C . Applying this result to the Hermitian and non-negative definite matrix $C = (\mathcal{J}_\epsilon \mathcal{J}_\epsilon^*)^\top$, and noting that $\rho(C) = \rho(C^\top)$, we obtain (9.71). From definition (9.52) for \mathcal{J}_ϵ we further get

$$\begin{aligned}
 \rho(\mathcal{J}_\epsilon \mathcal{J}_\epsilon^*) &= \rho[(J_\epsilon \otimes I_{2M})(J_\epsilon^* \otimes I_{2M})] \\
 &= \rho[(J_\epsilon J_\epsilon^* \otimes I_{2M})] \\
 &= \rho(J_\epsilon J_\epsilon^*)
 \end{aligned} \tag{9.73}$$

The matrix J_ϵ is block diagonal and consists of Jordan blocks. Assume initially that it consists of a single Jordan block, say, of size 4×4 , for illustration purposes. Then, we can write:

Proof



97

$$\begin{aligned} J_\epsilon J_\epsilon^* &= \begin{bmatrix} \lambda & & & \\ \epsilon & \lambda & & \\ & \epsilon & \lambda & \\ & & \epsilon & \lambda \end{bmatrix} \begin{bmatrix} \lambda^* & \epsilon & & \\ & \lambda^* & \epsilon & \\ & & \lambda^* & \epsilon \\ & & & \lambda^* \end{bmatrix} \\ &= \begin{bmatrix} |\lambda|^2 & \epsilon\lambda & & \\ \epsilon\lambda^* & |\lambda|^2 + \epsilon^2 & \epsilon\lambda & \\ & \epsilon\lambda^* & |\lambda|^2 + \epsilon^2 & \epsilon\lambda \\ & & \epsilon\lambda^* & |\lambda|^2 + \epsilon^2 \end{bmatrix} \quad (9.74) \end{aligned}$$



Proof

98

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Using the property that the spectral radius of a matrix is bounded by any of its norms, and using the 1-norm (maximum absolute column sum), we get for the above example

$$\begin{aligned}\rho(J_\epsilon J_\epsilon^*) &\leq \|J_\epsilon J_\epsilon^*\|_1 \\ &= |\lambda|^2 + \epsilon^2 + \epsilon|\lambda^*| + \epsilon|\lambda| \\ &= (|\lambda| + \epsilon)^2\end{aligned}\tag{9.75}$$

If J_ϵ consists of multiple Jordan blocks, say, L of them with eigenvalue λ_ℓ each, then

$$\rho(J_\epsilon J_\epsilon^*) \leq \max_{1 \leq \ell \leq L} (|\lambda_\ell| + \epsilon)^2 = (\rho(J_\epsilon) + \epsilon)^2\tag{9.76}$$



Proof

99

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where $\rho(J_\epsilon)$ does not depend on ϵ and is equal to the second largest eigenvalue in magnitude in J , which we know is strictly less than one in magnitude. Substituting this conclusion into (9.70) gives

$$\begin{aligned}\mathbb{E} \|\check{\mathbf{w}}_i^e\|^2 &\leq \frac{1}{t}(\rho(J_\epsilon) + \epsilon)^2 \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \\ &\quad \frac{1}{1-t} \mathbb{E} \left\| \mathbf{D}_{22,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{b}^e \right\|^2 + \mathbb{E} \|\check{s}_i^e\|^2\end{aligned}\tag{9.77}$$

Since we know that $\rho(J_\epsilon) \in (0, 1)$, then we can select ϵ small enough to ensure $\rho(J_\epsilon) + \epsilon \in (0, 1)$. We then select

$$t = \rho(J_\epsilon) + \epsilon\tag{9.78}$$

Proof



100

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

and rewrite (9.77) as

$$\begin{aligned} \mathbb{E} \|\check{\mathbf{w}}_i^e\|^2 &\leq (\rho(J_\epsilon) + \epsilon) \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \mathbb{E} \|\check{\mathbf{s}}_i^e\|^2 + \\ &\quad \left(\frac{1}{1 - \rho(J_\epsilon) - \epsilon} \right) \mathbb{E} \left\| \mathbf{D}_{22,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{b}^e \right\|^2 \end{aligned} \tag{9.79}$$

We can bound the last term on the right-hand side of the above expression as follows:

Proof



101

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} \mathbb{E} \left\| \mathbf{D}_{22,i-1}^T \check{\mathbf{w}}_{i-1}^e + \mathbf{D}_{12,i-1}^T \bar{\mathbf{w}}_{i-1}^e + \check{b}^e \right\|^2 &= & (9.80) \\ &= \mathbb{E} \left\| \frac{1}{3} 3\mathbf{D}_{22,i-1}^T \check{\mathbf{w}}_{i-1}^e + \frac{1}{3} 3\mathbf{D}_{12,i-1}^T \bar{\mathbf{w}}_{i-1}^e + \frac{1}{3} 3\check{b}^e \right\|^2 \\ &\leq \frac{1}{3} \mathbb{E} \left\| 3\mathbf{D}_{22,i-1}^T \check{\mathbf{w}}_{i-1}^e \right\|^2 + \frac{1}{3} \mathbb{E} \left\| 3\mathbf{D}_{12,i-1}^T \bar{\mathbf{w}}_{i-1}^e \right\|^2 + \frac{1}{3} \|3\check{b}^e\|^2 \\ &\leq 3\mathbb{E} \left\| \mathbf{D}_{22,i-1}^T \check{\mathbf{w}}_{i-1}^e \right\|^2 + 3\mathbb{E} \left\| \mathbf{D}_{12,i-1}^T \bar{\mathbf{w}}_{i-1}^e \right\|^2 + 3\|\check{b}^e\|^2 \\ &\leq 3\sigma_{22}^2 \mu_{\max}^2 \mathbb{E} \left\| \check{\mathbf{w}}_{i-1}^e \right\|^2 + 3\sigma_{12}^2 \mu_{\max}^2 \mathbb{E} \left\| \bar{\mathbf{w}}_{i-1}^e \right\|^2 + 3\|\check{b}^e\|^2 \end{aligned}$$

Proof



102

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Substituting into (9.79) we obtain

$$\begin{aligned} \mathbb{E} \|\check{\mathbf{w}}_i^e\|^2 &\leq \left(\rho(J_\epsilon) + \epsilon + \frac{3\sigma_{22}^2 \mu_{\max}^2}{1 - \rho(J_\epsilon) - \epsilon} \right) \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \\ &\quad \left(\frac{3\sigma_{12}^2 \mu_{\max}^2}{1 - \rho(J_\epsilon) - \epsilon} \right) \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \\ &\quad \left(\frac{3}{1 - \rho(J_\epsilon) - \epsilon} \right) \|\check{\mathbf{b}}^e\|^2 + \mathbb{E} \|\check{\mathbf{s}}_i^e\|^2 \end{aligned} \tag{9.81}$$

Proof



We now bound the noise terms, $\mathbb{E} \|\bar{s}_i^e\|^2$ in (9.69) and $\mathbb{E} \|\check{s}_i^e\|^2$ in (9.81). For that purpose, we first note that

$$\begin{aligned}\mathbb{E} \|\bar{s}_i^e\|^2 + \mathbb{E} \|\check{s}_i^e\|^2 &= \mathbb{E} \left\| \begin{bmatrix} \bar{s}_i^e \\ \check{s}_i^e \end{bmatrix} \right\|^2 \\ &= \mathbb{E} \left\| \mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \mathcal{M} s_i^e \right\|^2 \\ &\leq \left\| \mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \right\|^2 \|\mathcal{M}\|^2 \mathbb{E} \|s_i^e\|^2 \\ &\leq v_1^2 \mu_{\max}^2 \mathbb{E} \|s_i^e\|^2\end{aligned}\tag{9.82}$$



Proof

104

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where the positive constant v_1 is independent of μ_{\max} and is equal to the following norm

$$v_1 \triangleq \|\mathcal{V}_\epsilon^\top \mathcal{A}_2^\top\| \quad (9.83)$$

On the other hand, using (8.113)–(8.114), we have

$$\mathbb{E} \|\mathbf{s}_i^e\|^2 = \sum_{k=1}^N \mathbb{E} \|\mathbf{s}_{k,i}^e\|^2 = 2 \left(\sum_{k=1}^N \mathbb{E} \|\mathbf{s}_{k,i}\|^2 \right) \quad (9.84)$$

Proof



105

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

in terms of the variances of the individual gradient noise processes, $\mathbb{E} \|\mathbf{s}_{k,i}\|^2$, and where we used the fact that

$$\|\mathbf{s}_{k,i}^e\|^2 = 2\|\mathbf{s}_{k,i}\|^2 \quad (9.85)$$

Now, for each term $\mathbf{s}_{k,i}$ we have

Proof



106

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}\mathbb{E} \|s_{k,i}\|^2 &\stackrel{(8.118)}{\leq} (\beta_k^2/h^2) \mathbb{E} \|\tilde{\phi}_{k,i-1}\|^2 + \sigma_{s,k}^2 \\ &= (\beta_k^2/h^2) \mathbb{E} \left\| \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \tilde{\mathbf{w}}_{\ell,i-1} \right\|^2 + \sigma_{s,k}^2 \\ &\stackrel{(F.26)}{\leq} (\beta_k^2/h^2) \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \mathbb{E} \|\tilde{\mathbf{w}}_{\ell,i-1}\|^2 + \sigma_{s,k}^2 \\ &\leq (\beta_k^2/h^2) \sum_{\ell=1}^N \mathbb{E} \|\tilde{\mathbf{w}}_{\ell,i-1}\|^2 + \sigma_{s,k}^2\end{aligned}$$

Proof



107

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} &= (\beta_k^2/2h^2) \mathbb{E} \left\| \tilde{\mathbf{w}}_{i-1}^e \right\|^2 + \sigma_{s,k}^2 \\ &= (\beta_k^2/2h^2) \mathbb{E} \left\| (\mathcal{V}_\epsilon^{-1})^\top \mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_{i-1}^e \right\|^2 + \sigma_{s,k}^2 \\ &\leq (\beta_k^2/2h^2) \left\| (\mathcal{V}_\epsilon^{-1})^\top \right\|^2 \mathbb{E} \left\| \mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_{i-1}^e \right\|^2 + \sigma_{s,k}^2 \\ &\stackrel{(9.55)}{=} (\beta_k^2/2h^2) v_2^2 \left[\mathbb{E} \left\| \bar{\mathbf{w}}_{i-1}^e \right\|^2 + \mathbb{E} \left\| \check{\mathbf{w}}_{i-1}^e \right\|^2 \right] + \sigma_{s,k}^2 \end{aligned} \tag{9.86}$$



Proof

108

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where $h = 2$ for complex data, while the positive constant v_2 is independent of μ_{\max} and denotes the norm

$$v_2 \triangleq \left\| (\mathcal{V}_\epsilon^{-1})^\top \right\| \quad (9.87)$$

In this way, we can bound the term $\mathbb{E} \|s_i^e\|^2$ as follows:

$$\begin{aligned} \mathbb{E} \|s_i^e\|^2 &= 2 \left(\sum_{k=1}^N \mathbb{E} \|s_{k,i}\|^2 \right) \\ &\leq v_2^2 \beta_d^2 [\mathbb{E} \|\bar{w}_{i-1}^e\|^2 + \mathbb{E} \|\check{w}_{i-1}^e\|^2] + \sigma_s^2 \end{aligned} \quad (9.88)$$

where we introduced the scalars:

Proof



109

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\beta_d^2 \triangleq \sum_{k=1}^N \beta_k^2 / h^2 \quad (9.89)$$

$$\sigma_s^2 \triangleq \sum_{k=1}^N 2 \sigma_{s,k}^2 \quad (9.90)$$

Substituting into (9.82) we get

$$\mathbb{E} \|\bar{s}_i^e\|^2 + \mathbb{E} \|\check{s}_i^e\|^2 \leq v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 [\mathbb{E} \|\bar{w}_{i-1}^e\|^2 + \mathbb{E} \|\check{w}_{i-1}^e\|^2] + v_1^2 \mu_{\max}^2 \sigma_s^2 \quad (9.91)$$

Proof



110

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Using this bound in (9.69) and (9.81) we find that

$$\begin{aligned}\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^2 &\leq (1 - \sigma_{11}\mu_{\max} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2) \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \\ &\quad \left(\frac{\sigma_{21}^2 \mu_{\max}}{\sigma_{11}} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 \right) \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + v_1^2 \mu_{\max}^2 \sigma_s^2\end{aligned}\tag{9.92}$$

and

Proof



111

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} \mathbb{E} \|\check{\mathbf{w}}_i^e\|^2 &\leq \left(\rho(J_\epsilon) + \epsilon + \frac{3\sigma_{22}^2 \mu_{\max}^2}{1 - \rho(J_\epsilon) - \epsilon} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 \right) \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \\ &\quad \left(\frac{3\sigma_{12}^2 \mu_{\max}^2}{1 - \rho(J_\epsilon) - \epsilon} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 \right) \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \\ &\quad \left(\frac{3}{1 - \rho(J_\epsilon) - \epsilon} \right) \|\check{\mathbf{b}}^e\|^2 + v_1^2 \mu_{\max}^2 \sigma_s^2 \end{aligned} \tag{9.93}$$

We introduce the scalar coefficients

Proof



112

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$a = 1 - \sigma_{11}\mu_{\max} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 = 1 - O(\mu_{\max}) \quad (9.94)$$

$$b = \frac{\sigma_{21}^2 \mu_{\max}}{\sigma_{11}} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 = O(\mu_{\max}) \quad (9.95)$$

$$c = \frac{3\sigma_{12}^2 \mu_{\max}^2}{1 - \rho(J_\epsilon) - \epsilon} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 = O(\mu_{\max}^2) \quad (9.96)$$

$$\begin{aligned} d &= \rho(J_\epsilon) + \epsilon + \frac{3\sigma_{22}^2 \mu_{\max}^2}{1 - \rho(J_\epsilon) - \epsilon} + v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 \\ &= \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \end{aligned} \quad (9.97)$$

$$e = v_1^2 \mu_{\max}^2 \sigma_s^2 = O(\mu_{\max}^2) \quad (9.98)$$

$$f = \left(\frac{3}{1 - \rho(J_\epsilon) - \epsilon} \right) \|\check{b}^e\|^2 = O(\mu_{\max}^2) \quad (9.99)$$



Proof

113

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

since $\|\check{b}^e\| = O(\mu_{\max})$. Using these parameters, we can combine (9.92) and (9.93) into a single compact inequality recursion as follows:

$$\begin{bmatrix} \mathbb{E}\|\bar{\mathbf{w}}_i^e\|^2 \\ \mathbb{E}\|\check{\mathbf{w}}_i^e\|^2 \end{bmatrix} \preceq \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\Gamma} \begin{bmatrix} \mathbb{E}\|\bar{\mathbf{w}}_{i-1}^e\|^2 \\ \mathbb{E}\|\check{\mathbf{w}}_{i-1}^e\|^2 \end{bmatrix} + \begin{bmatrix} e \\ e+f \end{bmatrix} \quad (9.100)$$

in terms of the 2×2 coefficient matrix Γ indicated above and whose entries are of the form

$$\Gamma = \begin{bmatrix} 1 - O(\mu_{\max}) & O(\mu_{\max}) \\ O(\mu_{\max}^2) & \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \end{bmatrix} \quad (9.101)$$

Proof



114

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Now, we invoke again the property that the spectral radius of a matrix is upper bounded by any of its norms, and use the 1–norm (maximum absolute column sum), to conclude that

$$\rho(\Gamma) \leq \max \left\{ 1 - O(\mu_{\max}) + O(\mu_{\max}^2), \rho(J_\epsilon) + \epsilon + O(\mu_{\max}) + O(\mu_{\max}^2) \right\} \quad (9.102)$$

Since $\rho(J_\epsilon) < 1$ is independent of μ_{\max} , and since ϵ and μ_{\max} are small positive numbers that can be chosen arbitrarily small and independently of each other, it is clear that the right-hand side of the above expression can be made strictly smaller than one for sufficiently small ϵ and μ_{\max} . In that case, $\rho(\Gamma) < 1$ so that Γ is stable. Moreover, it holds that

Proof



$$\begin{aligned}(I_2 - \Gamma)^{-1} &= \begin{bmatrix} 1-a & -b \\ -c & 1-d \end{bmatrix}^{-1} \\ &= \frac{1}{(1-a)(1-d) - bc} \begin{bmatrix} 1-d & b \\ c & 1-a \end{bmatrix} \\ &= \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(\mu_{\max}) & O(1) \end{bmatrix}\end{aligned}\tag{9.103}$$

If we now iterate (9.100), and since Γ is stable, we conclude that

Proof



116

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} \limsup_{i \rightarrow \infty} \begin{bmatrix} \mathbb{E} \|\bar{\mathbf{w}}_i^e\|^2 \\ \mathbb{E} \|\check{\mathbf{w}}_i^e\|^2 \end{bmatrix} &\preceq (I_2 - \Gamma)^{-1} \begin{bmatrix} e \\ e + f \end{bmatrix} \\ &= \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(\mu_{\max}) & O(1) \end{bmatrix} \begin{bmatrix} O(\mu_{\max}^2) \\ O(\mu_{\max}^2) \end{bmatrix} \\ &= \begin{bmatrix} O(\mu_{\max}) \\ O(\mu_{\max}^2) \end{bmatrix} \end{aligned} \tag{9.104}$$

from which we conclude that

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\bar{\mathbf{w}}_i^e\|^2 = O(\mu_{\max}), \quad \limsup_{i \rightarrow \infty} \mathbb{E} \|\check{\mathbf{w}}_i^e\|^2 = O(\mu_{\max}^2) \tag{9.105}$$

Proof



117

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

and, therefore,

$$\begin{aligned}\limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i^e\|^2 &= \limsup_{i \rightarrow \infty} \mathbb{E} \left\| (\mathcal{V}_\epsilon^{-1})^\top \begin{bmatrix} \bar{\mathbf{w}}_i^e \\ \check{\mathbf{w}}_i^e \end{bmatrix} \right\|^2 \\ &\leq \limsup_{i \rightarrow \infty} v_2^2 [\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^2 + \mathbb{E} \|\check{\mathbf{w}}_i^e\|^2] \\ &= O(\mu_{\max})\end{aligned}\tag{9.106}$$

which leads to the desired result (9.11).

□



Remark

We remark that the type of derivation used in the above proof, which starts from a stochastic recursion of the form (9.60) and transforms it into a deterministic recursion of the form (9.100), with the sizes of the parameters specified in terms of μ_{\max} and with a Γ matrix of the form (9.101), will be a recurring technique in our presentation. For example, we will encounter a similar derivation in two more locations in the current chapter while establishing Theorems 9.2 and 9.6 further ahead — see expressions (9.153) and (9.301); these theorems deal with the stability of the fourth and first-order moments of the error vector. We will also encounter a similar derivation in the next chapter — see expressions (10.48), (10.77), and (10.89).

Fourth-Order Stability

Fourth-Order Error Moment



120

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

In the next chapter we will derive a long-term model to approximate the behavior of the network in the long term, as $i \rightarrow \infty$, and for sufficiently small step-sizes. The long-term model will be more tractable for performance analysis in the steady-state regime. At that point, we will argue that performance results that are derived from analyzing the long-term model provide accurate expressions for the performance results of the original network model to first-order in the step-size parameters. This is

Fourth-Order Error Moment



121

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

a reassuring conclusion that will lead to useful closed-form performance expressions. These results will be established under the condition that the fourth-order moment of the error vector, $\mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^4$, is asymptotically stable. We therefore establish this fact here and call upon it later in the analysis. To do so, we will rely on condition (8.121) on the fourth-order moments of the individual gradient noise processes.



Fourth-Order Error Moment

Theorem 9.2 (Fourth-order moment 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 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 network error vectors are stable for sufficiently small step-sizes, namely, it holds that

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{w}_{k,i}\|^4 = O(\mu_{\max}^2), \quad k = 1, 2, \dots, N \quad (9.107)$$



Sketch of Proof

123

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Following similar arguments to the second-order case, it can be verified that:

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

where

$$\Gamma = \begin{bmatrix} 1 - O(\mu_{\max}) & O(\mu_{\max}) \\ O(\mu_{\max}^2) & \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \end{bmatrix}, \quad (I - \Gamma)^{-1} = \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(\mu_{\max}) & O(1) \end{bmatrix}$$

Sketch of Proof



124

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} \limsup_{i \rightarrow \infty} \begin{bmatrix} \mathbb{E}\|\bar{\mathbf{w}}_i^e\|^4 \\ \mathbb{E}\|\check{\mathbf{w}}_i^e\|^4 \end{bmatrix} &\prec \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(\mu_{\max}) & O(1) \end{bmatrix} \left(\begin{bmatrix} O(\mu_{\max}^3) \\ O(\mu_{\max}^4) \end{bmatrix} + \begin{bmatrix} O(\mu_{\max}^4) \\ O(\mu_{\max}^4) \end{bmatrix} \right) \\ &= \begin{bmatrix} O(\mu_{\max}^2) \\ O(\mu_{\max}^4) \end{bmatrix} \end{aligned}$$

Sketch of Proof



125

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} \limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i^e\|^4 &= \limsup_{i \rightarrow \infty} \mathbb{E} \left(\left\| (\mathcal{V}_\epsilon^{-1})^\top \begin{bmatrix} \bar{\mathbf{w}}_i^e \\ \check{\mathbf{w}}_i^e \end{bmatrix} \right\|^2 \right)^2 \\ &\leq \left\| (\mathcal{V}_\epsilon^{-1})^\top \right\|^4 \left(\limsup_{i \rightarrow \infty} \mathbb{E} (\|\bar{\mathbf{w}}_i^e\|^2 + \|\check{\mathbf{w}}_i^e\|^2)^2 \right) \\ &\leq \limsup_{i \rightarrow \infty} 2v_2^4 (\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^4 + \mathbb{E} \|\check{\mathbf{w}}_i^e\|^4) \\ &= O(\mu_{\max}^2) \end{aligned} \tag{9.157}$$

which leads to the desired result (9.107). □



Detailed Proof

Proof. We again establish the result for the general case of complex data and, therefore, $h = 2$ throughout this derivation. We recall relations (9.61)–(9.62), namely,

$$\bar{\mathbf{w}}_i^e = (I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \bar{\mathbf{s}}_i^e \quad (9.108)$$

$$\check{\mathbf{w}}_i^e = (\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top) \check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{\mathbf{s}}_i^e - \check{\mathbf{b}}^e \quad (9.109)$$

Now note that, for any (deterministic or random) column vectors \mathbf{a} and \mathbf{b} , it holds that

$$\begin{aligned} \|\mathbf{a} + \mathbf{b}\|^4 &= \|\mathbf{a}\|^4 + \|\mathbf{b}\|^4 + 2\|\mathbf{a}\|^2 \|\mathbf{b}\|^2 + \\ &\quad 4\text{Re}(\mathbf{a}^* \mathbf{b}) [\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 + \text{Re}(\mathbf{a}^* \mathbf{b})] \end{aligned} \quad (9.110)$$

Proof



127

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

so that using the vector inequalities

$$[\operatorname{Re}(\mathbf{a}^* \mathbf{b})]^2 \leq |\mathbf{a}^* \mathbf{b}|^2 \leq \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 \quad (9.111)$$

and

$$2\operatorname{Re}(\mathbf{a}^* \mathbf{b}) \leq \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 \quad (9.112)$$

we get

$$\|\mathbf{a} + \mathbf{b}\|^4 \leq \|\mathbf{a}\|^4 + 3\|\mathbf{b}\|^4 + 8\|\mathbf{a}\|^2 \|\mathbf{b}\|^2 + 4\|\mathbf{a}\|^2 \operatorname{Re}(\mathbf{a}^* \mathbf{b}) \quad (9.113)$$

Applying this inequality to (9.108) with the identifications



Proof

128

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\mathbf{a} \leftarrow (I_{2M} - \mathcal{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \quad (9.114)$$

$$\mathbf{b} \leftarrow \bar{\mathbf{s}}_i^e \quad (9.115)$$

we obtain

$$\begin{aligned} \|\bar{\mathbf{w}}_i^e\|^4 &\leq \| (I_{2M} - \mathcal{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \|^4 + 3\|\bar{\mathbf{s}}_i^e\|^4 + \\ &\quad 8\| (I_{2M} - \mathcal{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \|^2 \|\bar{\mathbf{s}}_i^e\|^2 + \\ &\quad 4\| (I_{2M} - \mathcal{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \|^2 \operatorname{Re}(\mathbf{a}^* \bar{\mathbf{s}}_i^e) \end{aligned} \quad (9.116)$$

Proof



129

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Conditioning on \mathcal{F}_{i-1} and computing the expectations of both sides, we will find that the expectation of the last term on the right-hand side of the above expression is zero due to the assumed properties on the gradient noise process. Taking expectations again we then conclude that

$$\begin{aligned} \mathbb{E} \|\bar{\mathbf{w}}_i^e\|^4 &\leq \mathbb{E} \|(I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e\|^4 + 3(\mathbb{E} \|\bar{\mathbf{s}}_i^e\|^4) + \\ &\quad 8 \left(\mathbb{E} \|(I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e\|^2 \right) (\mathbb{E} \|\bar{\mathbf{s}}_i^e\|^2) \end{aligned}$$

Proof



130

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}
 &= \mathbb{E} \left\| (1-t) \frac{1}{1-t} (I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - t \frac{1}{t} \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \right\|^4 + \\
 &\quad 3 (\mathbb{E} \|\bar{\mathbf{s}}_i^e\|^4) + 8 (\mathbb{E} \|\bar{\mathbf{s}}_i^e\|^2) \times \\
 &\quad \left(\mathbb{E} \left\| (1-t) \frac{1}{1-t} (I_{2M} - \mathbf{D}_{11,i-1}^\top) \bar{\mathbf{w}}_{i-1}^e - t \frac{1}{t} \mathbf{D}_{21,i-1}^\top \check{\mathbf{w}}_{i-1}^e \right\|^2 \right) \\
 &\leq \frac{(1 - \sigma_{11}\mu_{\max})^4}{(1-t)^3} \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + \frac{\sigma_{21}^4 \mu_{\max}^4}{t^3} \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + 3 \mathbb{E} \|\bar{\mathbf{s}}_i^e\|^4 + \\
 &\quad 8 (\mathbb{E} \|\bar{\mathbf{s}}_i^e\|^2) \left(\frac{(1 - \sigma_{11}\mu_{\max})^2}{1-t} \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \frac{\sigma_{21}^2 \mu_{\max}^2}{t} \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 \right)
 \end{aligned} \tag{9.117}$$

Proof



131

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

for any arbitrary positive number $t \in (0, 1)$. Similarly, using the identifications

$$\mathbf{a} \leftarrow (\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top) \check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e - \check{b}^e \quad (9.118)$$

$$\mathbf{b} \leftarrow \check{s}_i^e \quad (9.119)$$

for relation (9.109), we can establish the inequality

Proof



132

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} \|\check{\mathbf{w}}_i^e\|^4 &\leq \left\|(\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top)\check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e - \check{b}^e\right\|^4 + 3\|\check{\mathbf{s}}_i^e\|^4 + \\ &8 \left\|(\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top)\check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e - \check{b}^e\right\|^2 \|\check{\mathbf{s}}_i^e\|^2 + \\ &4\|\mathbf{a}\|^2 \operatorname{Re}(\mathbf{a}^* \mathbf{b}) \end{aligned} \tag{9.120}$$

from which we conclude that, again for any positive scalar $t \in (0, 1)$:

Proof



133

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}
 & \mathbb{E} \|\check{\mathbf{w}}_i^e\|^4 \\
 & \leq \mathbb{E} \left\| (\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top) \check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e - \check{b}^e \right\|^4 + 3\mathbb{E} \|\check{\mathbf{s}}_i^e\|^4 + \\
 & \quad 8(\mathbb{E} \|\check{\mathbf{s}}_i^e\|^2) \left(\mathbb{E} \left\| (\mathcal{J}_\epsilon^\top - \mathbf{D}_{22,i-1}^\top) \check{\mathbf{w}}_{i-1}^e - \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e - \check{b}^e \right\|^2 \right) \\
 & \leq \mathbb{E} \left\| t \frac{1}{t} \mathcal{J}_\epsilon^\top \check{\mathbf{w}}_{i-1}^e - (1-t) \frac{1}{1-t} \left[\mathbf{D}_{22,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{b}^e \right] \right\|^4 + \\
 & \quad 3(\mathbb{E} \|\check{\mathbf{s}}_i^e\|^4) + 8(\mathbb{E} \|\check{\mathbf{s}}_i^e\|^2) \times \\
 & \quad \left(\mathbb{E} \left\| t \frac{1}{t} \mathcal{J}_\epsilon^\top \check{\mathbf{w}}_{i-1}^e - (1-t) \frac{1}{1-t} \left[\mathbf{D}_{22,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{b}^e \right] \right\|^2 \right)
 \end{aligned}$$

Proof



134

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} &\leq \frac{1}{t^3} \|\mathcal{J}_\epsilon\|^4 \mathbb{E} \left\| \check{\mathbf{w}}_{i-1}^e \right\|^4 + \\ &\quad \frac{1}{(1-t)^3} \mathbb{E} \left\| \mathbf{D}_{22,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{b}^e \right\|^4 + \\ &\quad 3 \left(\mathbb{E} \|\check{s}_i^e\|^4 \right) + 8 \left(\mathbb{E} \|\check{s}_i^e\|^2 \right) \times \\ &\quad \left(\frac{1}{t} \|\mathcal{J}_\epsilon\|^2 \mathbb{E} \left\| \check{\mathbf{w}}_{i-1}^e \right\|^2 + \frac{1}{1-t} \mathbb{E} \left\| \mathbf{D}_{22,i-1}^\top \check{\mathbf{w}}_{i-1}^e + \mathbf{D}_{12,i-1}^\top \bar{\mathbf{w}}_{i-1}^e + \check{b}^e \right\|^2 \right) \end{aligned}$$

Proof



135

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} &\leq \frac{(\rho(J_\epsilon) + \epsilon)^4}{t^3} \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + \\ &\quad \frac{27\sigma_{22}^4\mu_{\max}^4}{(1-t)^3} \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + \frac{27\sigma_{12}^4\mu_{\max}^4}{(1-t)^3} \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + 27\|\check{\mathbf{b}}^e\|^4 + \\ &\quad 3(\mathbb{E} \|\check{\mathbf{s}}_i^e\|^4) + 8\frac{(\rho(J_\epsilon) + \epsilon)^2}{t} (\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2) \mathbb{E} \|\check{\mathbf{s}}_i^e\|^2 + \\ &\quad 8 \left(\frac{3\sigma_{22}^2\mu_{\max}^2}{1-t} \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \frac{3\sigma_{12}^2\mu_{\max}^2}{1-t} \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + 3\|\check{\mathbf{b}}^e\|^2 \right) \mathbb{E} \|\check{\mathbf{s}}_i^e\|^2 \end{aligned} \tag{9.121}$$

Proof



where in the last inequality we used the result

$$\begin{aligned}\|a + b + c\|^4 &= \left\| \frac{1}{3}3a + \frac{1}{3}3b + \frac{1}{3}3c \right\|^4 \\ &\leq \frac{1}{3}\|3a\|^4 + \frac{1}{3}\|3b\|^4 + \frac{1}{3}\|3c\|^4 \\ &= 27\|a\|^4 + 27\|b\|^4 + 27\|c\|^4\end{aligned}\tag{9.122}$$

Proof



137

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We now bound the fourth-order noise terms that appear in expressions (9.121) and (9.121) for $\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^4$ and $\mathbb{E} \|\check{\mathbf{w}}_i^e\|^4$, namely, $\mathbb{E} \|\bar{\mathbf{s}}_i^e\|^4$ and $\mathbb{E} \|\check{\mathbf{s}}_i^e\|^4$. Thus, note that

$$\begin{aligned}\mathbb{E} \|\bar{\mathbf{s}}_i^e\|^4 + \mathbb{E} \|\check{\mathbf{s}}_i^e\|^4 &\leq \mathbb{E} (\|\bar{\mathbf{s}}_i^e\|^2 + \|\check{\mathbf{s}}_i^e\|^2)^2 \\&= \mathbb{E} \left(\left\| \begin{bmatrix} \bar{\mathbf{s}}_i^e \\ \check{\mathbf{s}}_i^e \end{bmatrix} \right\|^2 \right)^2 \\&= \mathbb{E} \left\| \mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \mathcal{M} \mathbf{s}_i^e \right\|^4 \\&\leq \left\| \mathcal{V}_\epsilon^\top \mathcal{A}_2^\top \right\|^4 \|\mathcal{M}\|^4 \mathbb{E} \|\mathbf{s}_i^e\|^4 \\&\leq v_1^4 \mu_{\max}^4 \mathbb{E} \|\mathbf{s}_i^e\|^4\end{aligned}\tag{9.123}$$



Proof

138

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

On the other hand, using Jensen's inequality (F.26) and applying it to the convex function $f(x) = x^2$,

$$\begin{aligned}\mathbb{E} \|\boldsymbol{s}_i^e\|^4 &= \mathbb{E} (\|\boldsymbol{s}_i^e\|^2)^2 \\ &= 4 \mathbb{E} \left(\sum_{k=1}^N \|\boldsymbol{s}_{k,i}\|^2 \right)^2 \\ &= 4 \mathbb{E} \left(\frac{1}{N} N \|\boldsymbol{s}_{1,i}\|^2 + \frac{1}{N} N \|\boldsymbol{s}_{2,i}\|^2 + \dots + \frac{1}{N} N \|\boldsymbol{s}_{N,i}\|^2 \right)^2 \\ &\stackrel{(F.26)}{\leq} \frac{4}{N} \mathbb{E} ((N \|\boldsymbol{s}_{1,i}\|^2)^2 + (N \|\boldsymbol{s}_{2,i}\|^2)^2 + \dots + (N \|\boldsymbol{s}_{N,i}\|^2)^2) \\ &= 4N \sum_{k=1}^N \mathbb{E} \|\boldsymbol{s}_{k,i}\|^4\end{aligned}\tag{9.124}$$



Proof

139

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

in terms of the fourth-order moments of the individual gradient noise processes, $\mathbb{E} \|s_{k,i}\|^4$. Likewise, we have

$$\begin{aligned} \sum_{\ell=1}^N \|\tilde{\mathbf{w}}_{\ell,i-1}\|^4 &\leq \left(\|\tilde{\mathbf{w}}_{1,i-1}\|^2 + \|\tilde{\mathbf{w}}_{2,i-1}\|^2 + \dots + \|\tilde{\mathbf{w}}_{N,i-1}\|^2 \right)^2 \\ &= \left(\|\tilde{\mathbf{w}}_{i-1}\|^2 \right)^2 \\ &= \left(\frac{1}{2} \|\tilde{\mathbf{w}}_{i-1}^e\|^2 \right)^2 \\ &= \frac{1}{4} \|\tilde{\mathbf{w}}_{i-1}^e\|^4 \end{aligned} \tag{9.125}$$



Proof

140

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Therefore, for each term $s_{k,i}$ in (9.124) we can write

$$\begin{aligned}\mathbb{E} \|s_{k,i}\|^4 &\stackrel{(8.122)}{\leq} (\beta_{4,k}^4/h^4)\mathbb{E} \|\tilde{\phi}_{k,i-1}\|^4 + \sigma_{s4,k}^4 \\ &= (\beta_{4,k}^4/h^4)\mathbb{E} \left\| \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \tilde{\mathbf{w}}_{\ell,i-1} \right\|^4 + \sigma_{s4,k}^4 \\ &\stackrel{(F.26)}{\leq} (\beta_{4,k}^4/h^4) \sum_{\ell \in \mathcal{N}_k} a_{1,\ell k} \mathbb{E} \|\tilde{\mathbf{w}}_{\ell,i-1}\|^4 + \sigma_{s4,k}^4\end{aligned}$$

Proof



141

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned} &\leq (\beta_{4,k}^4/h^4) \sum_{\ell=1}^N \mathbb{E} \|\tilde{\mathbf{w}}_{\ell,i-1}\|^4 + \sigma_{s4,k}^4 \\ &\stackrel{(9.125)}{\leq} (\beta_{4,k}^4/4h^4) \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}^e\|^4 + \sigma_{s4,k}^4 \\ &= (\beta_{4,k}^4/4h^4) \mathbb{E} \left\| (\mathcal{V}_\epsilon^{-1})^\top \mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_{i-1}^e \right\|^4 + \sigma_{s4,k}^4 \\ &\leq (\beta_{4,k}^4/4h^4) \left\| (\mathcal{V}_\epsilon^{-1})^\top \right\|^4 \mathbb{E} \|\mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_{i-1}^e\|^4 + \sigma_{s4,k}^4 \\ &\stackrel{(a)}{=} 2(\beta_{4,k}^4/4h^4) v_2^4 [\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4] + \sigma_{s4,k}^4 \end{aligned} \tag{9.126}$$

Proof



142

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where in step (a) we used (9.55) to conclude that

$$\begin{aligned}\|\mathcal{V}_\epsilon^\top \tilde{\mathbf{w}}_{i-1}^e\|^4 &= \left(\left\| \begin{bmatrix} \bar{\mathbf{w}}_{i-1}^e \\ \check{\mathbf{w}}_{i-1}^e \end{bmatrix} \right\|^2 \right)^2 \\ &= \left(\|\bar{\mathbf{w}}_{i-1}^e\|^2 + \|\check{\mathbf{w}}_{i-1}^e\|^2 \right)^2 \\ &\leq 2 \|\bar{\mathbf{w}}_{i-1}^e\|^4 + 2 \|\check{\mathbf{w}}_{i-1}^e\|^4\end{aligned}\tag{9.127}$$

In this way, we can bound the term $\mathbb{E} \|\mathbf{s}_i^e\|^4$ as follows:

$$\mathbb{E} \|\mathbf{s}_i^e\|^4 \leq v_2^4 \beta_{d4}^4 [\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4] + \sigma_{s4}^4\tag{9.128}$$



Proof

143

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where we introduced the scalars:

$$\beta_{d4}^4 \triangleq 2N \left(\sum_{k=1}^N \beta_{4,k}^4 / h^4 \right) \quad (9.129)$$

$$\sigma_{s4}^4 \triangleq 4N \left(\sum_{k=1}^N \sigma_{s4,k}^4 \right) \quad (9.130)$$

Proof



144

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Substituting into (9.123) we get

$$\begin{aligned} \mathbb{E} \|\bar{s}_i^e\|^4 + \mathbb{E} \|\check{s}_i^e\|^4 &\leq v_1^4 v_2^4 \beta_{d4}^4 \mu_{\max}^4 [\mathbb{E} \|\bar{w}_{i-1}^e\|^4 + \mathbb{E} \|\check{w}_{i-1}^e\|^4] + \\ &\quad v_1^4 \mu_{\max}^4 \sigma_{s4}^4 \end{aligned} \tag{9.131}$$

Returning to (9.117), selecting $t = \sigma_{11} \mu_{\max}$, and using the bounds (9.91) and (9.131), we then find that

Proof



Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}
 \mathbb{E} \|\bar{\mathbf{w}}_i^e\|^4 &\leq (1 - \sigma_{11}\mu_{\max})\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + \frac{\sigma_{21}^4\mu_{\max}}{\sigma_{11}^3}\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + \\
 &\quad 3v_1^4v_2^4\beta_{d4}^4\mu_{\max}^4 [\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4] + 3v_1^4\mu_{\max}^4\sigma_{s4}^4 + \\
 &\quad 8v_1^2v_2^2\beta_d^2\mu_{\max}^2(1 - \sigma_{11}\mu_{\max}) (\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2)^2 + \\
 &\quad 8v_1^2v_2^2\beta_d^2\mu_{\max}^2(1 - \sigma_{11}\mu_{\max}) (\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2) (\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2) + \\
 &\quad 8v_1^2\mu_{\max}^2\sigma_s^2(1 - \sigma_{11}\mu_{\max})\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \\
 &\quad 8\frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}v_1^2v_2^2\beta_d^2\mu_{\max}^2 (\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2)^2 + \\
 &\quad 8\frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}v_1^2v_2^2\beta_d^2\mu_{\max}^2 (\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2) (\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2) + \\
 &\quad 8\frac{\sigma_{21}^2\mu_{\max}}{\sigma_{11}}v_1^2\mu_{\max}^2\sigma_s^2\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 \tag{9.132}
 \end{aligned}$$



Proof

146

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

Now, for any real random variables \mathbf{a} and \mathbf{b} it holds that

$$(\mathbb{E} \mathbf{a})^2 \leq \mathbb{E} \mathbf{a}^2 \quad (9.133)$$

and

$$2 (\mathbb{E} \mathbf{a}^2) \cdot (\mathbb{E} \mathbf{b}^2) \leq \mathbb{E} \mathbf{a}^4 + \mathbb{E} \mathbf{b}^4 \quad (9.134)$$

This latter property can be established as follows. Using $(\mathbb{E} \mathbf{a}^2 - \mathbb{E} \mathbf{b}^2)^2 \geq 0$, we get

$$\begin{aligned} 2 (\mathbb{E} \mathbf{a}^2) \cdot (\mathbb{E} \mathbf{b}^2) &\leq (\mathbb{E} \mathbf{a}^2)^2 + (\mathbb{E} \mathbf{b}^2)^2 \\ &\leq \mathbb{E} \mathbf{a}^4 + \mathbb{E} \mathbf{b}^4 \end{aligned} \quad (9.135)$$

Proof



147

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

These properties enable us to write

$$2 \left(\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 \right) \left(\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 \right) \leq \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 \quad (9.136)$$

$$\left(\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 \right)^2 \leq \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 \quad (9.137)$$

$$\left(\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 \right)^2 \leq \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 \quad (9.138)$$

so that

$$\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^4 \leq a \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + b \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + a' \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + b' \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + e \quad (9.139)$$



Proof

148

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

where the constant parameters $\{a, b, a', b', e\}$ have the following form

$$a = 1 - \sigma_{11}\mu_{\max} + O(\mu_{\max}^2) \quad (9.140)$$

$$b = O(\mu_{\max}) \quad (9.141)$$

$$a' = O(\mu_{\max}^2) \quad (9.142)$$

$$b' = O(\mu_{\max}^3) \quad (9.143)$$

$$e = O(\mu_{\max}^4) \quad (9.144)$$

In a similar manner, using (9.121) and selecting

$$t = \rho(J_\epsilon) + \epsilon < 1 \quad (9.145)$$

Proof



149

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

$$\begin{aligned}
 \mathbb{E} \|\check{\mathbf{w}}_i^e\|^4 &\leq (\rho(J_\epsilon) + \epsilon) \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + \\
 &\quad \frac{27\mu_{\max}^4}{(1 - \rho(J_\epsilon) - \epsilon)^3} [\sigma_{22}^4 \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + \sigma_{12}^4 \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4] + \\
 &\quad 3v_1^4 v_2^4 \beta_{d4}^4 \mu_{\max}^4 [\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4] + 3v_1^4 \mu_{\max}^4 \sigma_{s4}^4 \\
 &\quad 8(\rho(J_\epsilon) + \epsilon) v_1^2 \mu_{\max}^2 \sigma_s^2 \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + 27 \|\check{b}^e\|^4 + \\
 &\quad 4(\rho(J_\epsilon) + \epsilon) v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 [\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + 3\mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4] + \\
 &\quad \frac{24\mu_{\max}^4 v_1^2 \sigma_s^2}{1 - \rho(J_\epsilon) - \epsilon} [\sigma_{22}^2 \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + \sigma_{12}^2 \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2] + \\
 &\quad \frac{12\mu_{\max}^4 v_1^2 v_2^2 \beta_d^2}{1 - \rho(J_\epsilon) - \epsilon} \times \\
 &\quad [(\sigma_{22}^2 + 3\sigma_{12}^2) \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + (\sigma_{12}^2 + 3\sigma_{22}^2) \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2] + \\
 &\quad 24 \|\check{b}^e\|^2 v_1^2 v_2^2 \beta_d^2 \mu_{\max}^2 [\mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2] + \\
 &\quad 24 \|\check{b}^e\|^2 v_1^2 \mu_{\max}^2 \sigma_s^2
 \end{aligned} \tag{9.146}$$



Proof

150

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

so that

$$\mathbb{E} \|\check{\mathbf{w}}_i^e\|^4 \leq c \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^4 + d \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^4 + c' \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 + d' \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 + f \quad (9.147)$$

where the constant parameters $\{c, d, c', d', f\}$ have the following form

$$c = O(\mu_{\max}^2) \quad (9.148)$$

$$d = \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \quad (9.149)$$

$$c' = O(\mu_{\max}^4) \quad (9.150)$$

$$d' = O(\mu_{\max}^2) \quad (9.151)$$

$$f = O(\mu_{\max}^4) \quad (9.152)$$



Proof

151

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

In other words, we can write

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

in terms of the 2×2 coefficient matrix Γ indicated above and whose entries are of the form

$$\Gamma = \begin{bmatrix} 1 - O(\mu_{\max}) & O(\mu_{\max}) \\ O(\mu_{\max}^2) & \rho(J_\epsilon) + \epsilon + O(\mu_{\max}^2) \end{bmatrix}, \quad (I - \Gamma)^{-1} = \begin{bmatrix} O(1/\mu_{\max}) & O(1) \\ O(\mu_{\max}) & O(1) \end{bmatrix} \quad (9.154)$$

Proof



152

Lecture #17: Stability of Multi-Agent Networks

EE210B: Inference over Networks (A. H. Sayed)

We again find that Γ is a stable matrix for sufficiently small μ_{\max} and ϵ . Moreover, using (9.105) we have

$$\limsup_{i \rightarrow \infty} \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \begin{bmatrix} \mathbb{E} \|\bar{\mathbf{w}}_{i-1}^e\|^2 \\ \mathbb{E} \|\check{\mathbf{w}}_{i-1}^e\|^2 \end{bmatrix} = \begin{bmatrix} O(\mu_{\max}^3) \\ O(\mu_{\max}^4) \end{bmatrix} \quad (9.155)$$

In this case, we can iterate (9.153) and use (9.103) to conclude that

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\bar{\mathbf{w}}_i^e\|^4 = O(\mu_{\max}^2), \quad \limsup_{i \rightarrow \infty} \mathbb{E} \|\check{\mathbf{w}}_i^e\|^4 = O(\mu_{\max}^4) \quad (9.156)$$

and, therefore,

Proof



$$\begin{aligned} \limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i^e\|^4 &= \limsup_{i \rightarrow \infty} \mathbb{E} \left(\left\| (\mathcal{V}_\epsilon^{-1})^\top \begin{bmatrix} \bar{\mathbf{w}}_i^e \\ \check{\mathbf{w}}_i^e \end{bmatrix} \right\|^2 \right)^2 \\ &\leq \left\| (\mathcal{V}_\epsilon^{-1})^\top \right\|^4 \left(\limsup_{i \rightarrow \infty} \mathbb{E} (\|\bar{\mathbf{w}}_i^e\|^2 + \|\check{\mathbf{w}}_i^e\|^2)^2 \right) \\ &\leq \limsup_{i \rightarrow \infty} 2v_2^4 (\mathbb{E} \|\bar{\mathbf{w}}_i^e\|^4 + \mathbb{E} \|\check{\mathbf{w}}_i^e\|^4) \\ &= O(\mu_{\max}^2) \end{aligned} \tag{9.157}$$

which leads to the desired result (9.107). □

End of Lecture

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.