

# INFERENCE OVER NETWORKS

## LECTURE #23: Role of Informed Agents

**Professor Ali H. Sayed**  
**UCLA Electrical Engineering**





# Reference

2

Lecture #23: Role of Informed Agents

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

## Chapter 13 (Role of Informed Agents, pp. 646-661):

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



# Setting

3

Lecture #23: Role of Informed Agents

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

We assumed in our presentation so far that all agents in the network have *continuous* access to data measurements and are able to evaluate their gradient vector approximations. However, it is observed in nature that the behavior of biological networks is often driven more heavily by a small fraction of informed agents as happens, for example, with bees and fish [12, 22, 125, 219]. This phenomenon motivates us to examine in this chapter multi-agent networks where only a *fraction* of the agents are informed, while the remaining agents are uninformed.

# Informed & Uninformed Agents

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.

# Informed Agents



Informed agents are defined as those agents that are capable of evaluating their gradient vector approximation continuously from streaming data and of performing the two tasks of adapting their iterates and consulting with their neighbors. Uninformed agents, on the other hand, are incapable of performing adaptation but can still participate in the consultation process with their neighbors. In this way, uninformed agents continue to assist in the diffusion of information across the network and act primarily as relay agents. We illustrate these two definitions



# Informed Agents

6

Lecture #23: Role of Informed Agents

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

by considering a strongly-connected network running, for example, the ATC diffusion strategy (7.19). When an agent  $k$  is informed, it employs a strictly positive step-size and performs the two steps of adaptation and combination:

$$\text{(informed)} \quad \begin{cases} \boldsymbol{\psi}_{k,i} &= \mathbf{w}_{k,i-1} - \frac{2\mu}{h} \widehat{\nabla_{w^*} J}_k(\mathbf{w}_{k,i-1}) \\ \mathbf{w}_{k,i} &= \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \boldsymbol{\psi}_{\ell,i} \end{cases} \quad (13.1)$$



# Informed Agents

7

Lecture #23: Role of Informed Agents

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

where  $h = 1$  for real data and  $h = 2$  for complex data. When an agent is uninformed, we set its step-size parameter to zero,  $\mu_k = 0$ , so that they are unable to perform the adaptation step but continue to perform the aggregation step. Their update equations therefore reduce to

$$\text{(uninformed)} \quad \begin{cases} \boldsymbol{\psi}_{k,i} &= \mathbf{w}_{k,i-1} \\ \mathbf{w}_{k,i} &= \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \boldsymbol{\psi}_{\ell,i} \end{cases} \quad (13.2)$$

# Informed Agents



which collapse into the more compact form:

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k} \mathbf{w}_{\ell,i-1} \quad (13.3)$$

Although unnecessary for our treatment, we will assume for simplicity of presentation that the step-size parameter is uniform and equal to  $\mu$  across all informed agents:

$$\mu_k = \begin{cases} \mu, & \text{(informed agent)} \\ 0, & \text{(uninformed agent)} \end{cases} \quad (13.4)$$

# Informed Agents



9

Lecture #23: Role of Informed Agents

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

We will also focus on diffusion and consensus networks. Recall from (8.7)–(8.10) that the consensus and diffusion strategies correspond to the following choices for  $\{A_o, A_1, A_2\}$  in terms of a single combination matrix  $A$  in the general description (8.46):

$$\text{consensus: } A_o = A, \quad A_1 = I_N = A_2 \quad (13.5)$$

$$\text{CTA diffusion: } A_1 = A, \quad A_2 = I_N = A_o \quad (13.6)$$

$$\text{ATC diffusion: } A_2 = A, \quad A_1 = I_N = A_o \quad (13.7)$$

# Conditions on Costs

Course EE210B  
Spring Quarter 2015

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



# Conditions

11

Lecture #23: Role of Informed Agents

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

We recall the definition of the aggregate cost function for the case when all agents are informed:

$$J^{\text{glob}}(w) \triangleq \sum_{k=1}^N J_k(w) \quad (13.8)$$

Let  $\mathcal{N}_I$  denote the set of indices of informed agents in the network:

$$\mathcal{N}_I \triangleq \{k : \text{such that } \mu_k = \mu > 0\} \quad (13.9)$$

The number of elements in  $\mathcal{N}_I$  is denoted by

$$N_I = |\mathcal{N}_I| \quad (13.10)$$



# Conditions

12

Lecture #23: Role of Informed Agents

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

The remaining agents are uninformed. We assume the network has at least one informed agent so that  $N_I \geq 1$ .

Now, observe from the definitions of informed and uninformed agents that if some agent  $k_o$  happens to be uninformed, then information about its gradient vector and, hence, cost function  $J_{k_o}(w)$ , is excluded from the overall learning process. For this reason, the effective global cost that the network will be minimizing is redefined as

$$J^{\text{glob,eff}}(w) \triangleq \sum_{k \in \mathcal{N}_I} J_k(w) \quad (13.11)$$

# Conditions



where the sum is over the individual costs of the informed agents. Clearly, if the individual costs share a common minimizer (which is the situation of most interest to us in this chapter), then the global minimizers of (13.8) and (13.11) will coincide. In general, though, the minimizers of these global costs may be different, and the minimizer of (13.11) will change with the set  $\mathcal{N}_I$ . For this reason, whenever necessary, we shall write  $w^o(\mathcal{N}_I)$  to highlight the dependency of the minimizer of (13.11) on the set of informed agents.



# Conditions

In this chapter, whenever we refer to the global cost, we will be referring to the *effective* global cost (13.11) since entries from uninformed agents are excluded. It is this global cost, along with the individual costs of the informed agents, that we now need to assume to satisfy the conditions in Assumption 6.1. Specifically, the individual cost functions,  $J_k(w)$  for  $k \in \mathcal{N}_I$ , are each twice-differentiable and convex, with at least one of them being  $\nu_d$ -strongly convex. Moreover, the effective aggregate cost function,  $J^{\text{glob,eff}}(w)$ , is also twice-differentiable and satisfies

$$0 < \frac{\nu_d}{h} I_{hM} \leq \nabla_w^2 J^{\text{glob,eff}}(w) \leq \frac{\delta_d}{h} I_{hM} \quad (13.12)$$



# Conditions

for some positive parameters  $\nu_d \leq \delta_d$ . In other words, conditions that we introduced in the earlier chapters on the cost functions  $\{J^{\text{glob}}(w), J_k(w), k = 1, 2, \dots, N\}$  will now need to be satisfied by the informed agents and by the effective global cost,  $\{J^{\text{glob,eff}}(w), J_k(w), k \in \mathcal{N}_I\}$ . For example, the smoothness condition (10.1) on the individual cost functions will now be required to be satisfied by the informed agents. Likewise, the gradient noise processes at the informed agents will need to satisfy the conditions in Assumption 8.1 or the fourth-order moment condition (8.121), as well as the smoothness condition (11.10) on their covariance matrices.



# Conditions

The limit point of the network will continue to be denoted by  $w^*$  and it is now defined as unique minimum of the following weighted aggregate cost function,  $J^{\text{glob,eff},*}(w)$ , from (8.53), namely,

$$J^{\text{glob,eff}*}(w) \triangleq \sum_{k \in \mathcal{N}_I} \mu_k p_k J_k(w) \quad (13.13)$$

where the sum is again defined over the set of informed agents, and where the  $\{p_k\}$  are the entries of the Perron eigenvector of the primitive combination matrix  $A$ :

$$Ap = p, \quad \mathbf{1}^\top p = 1, \quad p_k > 0 \quad (13.14)$$

# Conditions



The limit vector,  $w^*$ , that results from (13.13) is again dependent on the set of informed agents. For this reason, whenever necessary, we shall also write  $w^*(\mathcal{N}_I)$  to highlight the dependency of the minimizer of (13.13) on  $\mathcal{N}_I$ .

Under these adjustments, with requirements now imposed on the informed agents and with the network still assumed to be strongly-connected, it can be verified that the multi-agent network continues to be stable in the mean-square sense and in the mean sense, namely, for



# Conditions

all agents  $k = 1, 2, \dots, N$  (informed and uninformed alike):

$$\limsup_{i \rightarrow \infty} \|\mathbb{E} \tilde{\mathbf{w}}_{k,i}\| = O(\mu) \quad (13.15)$$

$$\limsup_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 = O(\mu) \quad (13.16)$$

These facts are justified as follows. With regards to mean-square-error stability, we refer to the general proof in step (c) of [Theorem 9.1](#). The two main differences that will occur if we repeat the argument relate to expressions [\(9.33\)](#) and [\(9.58\)](#), which now become



# Conditions

19

Lecture #23: Role of Informed Agents

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

$$\mathbf{D}_{11,i-1} = \sum_{k \in \mathcal{N}_I} \mu p_k \mathbf{H}_{k,i-1}^\top \quad (13.17)$$

$$0 = \sum_{k \in \mathcal{N}_I} \mu p_k b_k^e \quad (13.18)$$

with the sums evaluated over the set of informed agents. It will continue to hold that  $\mathbf{D}_{11,i-1} > 0$  in view of condition (13.12). Likewise, result (13.18) will hold in view of (13.13) from which we conclude that  $w^*$  now satisfies

$$\sum_{k \in \mathcal{N}_I} \mu p_k \nabla_w J_k(w^*) = 0 \quad (13.19)$$

# Conditions



20

Lecture #23: Role of Informed Agents

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

With regards to mean stability, if we refer to the proof of Theorem 9.3, we will again conclude that the matrix  $\mathcal{B}$  remains stable since the matrix  $D_{11}$  defined by (9.195) will now become

$$D_{11} = \sum_{k \in \mathcal{N}_I} \mu p_k H_k^T \quad (13.20)$$

and it remains positive-definite.

# MSE Performance

Course EE210B  
Spring Quarter 2015

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

# MSE Performance



The results in the sequel reveal some interesting facts about adaptation and learning in the presence of informed and uninformed agents [213, 247, 250]. For example, it will be seen that when the set of informed agents is enlarged, the convergence rate of the network will become faster albeit at the expense of possible deterioration in mean-square-error performance. In other words, the MSD and ER performance metrics do not necessarily improve with a larger proportion of informed agents. The arguments in this chapter extend the presentation from [213] to the case of complex-valued arguments.



# MSE Performance

23

Lecture #23: Role of Informed Agents

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

Thus, consider strongly-connected networks running the consensus or diffusion strategies (7.9), (7.18), or (7.19). We recall from expression (11.118) that, when all agents are informed, the MSD performance of these distributed solutions is given by:

$$\text{MSD}_{\text{dist,av}} = \frac{\mu}{2h} \text{Tr} \left[ \left( \sum_{k=1}^N p_k H_k \right)^{-1} \left( \sum_{k=1}^N p_k^2 G_k \right) \right] \quad (13.21)$$



# MSE Performance

24

Lecture #23: Role of Informed Agents

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

We also recall from (11.139) that the convergence rate of the error variances,  $\mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2$ , towards this MSD value is given by

$$\alpha_{\text{dist}} = 1 - 2\mu \lambda_{\min} \left\{ \sum_{k=1}^N p_k H_k \right\} + o(\mu) \quad (13.22)$$

in terms of the smallest eigenvalue of the sum of weighted Hessian matrices. In the above expression, the parameter  $\alpha_{\text{dist}} \in (0, 1)$  and the smaller the value of  $\alpha_{\text{dist}}$  is, the faster the convergence behavior becomes.

# MSE Performance



If we now consider the case where some agents are uninformed, and repeat the derivation that led to (11.47) and (11.118), we will find that the same result still hold if we set  $\mu_k = 0$  for the uninformed agents [68, 213, 247, 250], namely,

$$\alpha_{\text{dist}} = 1 - 2\mu \lambda_{\min} \left\{ \sum_{k \in \mathcal{N}_I} p_k H_k \right\} + o(\mu)$$

$$\text{MSD}_{\text{dist},k} = \text{MSD}_{\text{dist,av}} = \frac{\mu}{2h} \text{Tr} \left[ \left( \sum_{k \in \mathcal{N}_I} p_k H_k \right)^{-1} \left( \sum_{k \in \mathcal{N}_I} p_k^2 G_k \right) \right]$$



# MSE Performance

26

Lecture #23: Role of Informed Agents

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

Observe now that since the entries of  $p$  are positive for primitive left-stochastic matrices  $A$ , it is clear from (13.23) that, for small step-sizes, if the set of informed agents is enlarged from  $\mathcal{N}_I$  to

$$\mathcal{N}'_I \supset \mathcal{N}_I \quad (13.25)$$

then the convergence rate improves (i.e., faster convergence with  $\alpha_{\text{dist}}$  becoming smaller). However, from (13.24), the network MSD may decrease, remain unchanged, or increase depending on the values of  $\{H_k, G_k\}$ . This situation is illustrated in Figure 13.1.

# MSE Performance



27

Lecture #23: Role of Informed Agents

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

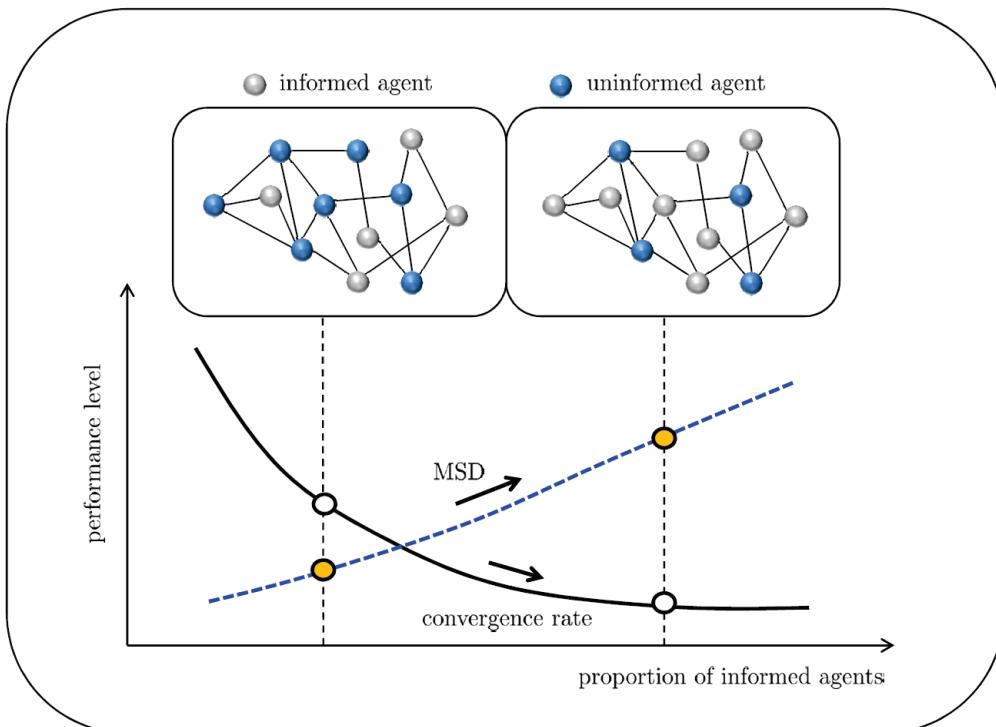


Figure 13.1



# MSE Performance

Note that the previous statements compare the convergence rates and MSD levels relative to the minimizers  $w^*(\mathcal{N}_I)$  and  $w^*(\mathcal{N}'_I)$  of the weighted effective costs (13.13) that would correspond to the sets  $\mathcal{N}_I$  and  $\mathcal{N}'_I$ . These minimizers are generally different and, therefore, these comparisons amount to determining how well and how fast the network configuration,  $\mathcal{N}_I$  or  $\mathcal{N}'_I$ , converge towards their respective limit points. The next example describes the useful scenario when the two minimizers,  $w^*(\mathcal{N}_I)$  and  $w^*(\mathcal{N}'_I)$ , coincide since the corresponding individual costs will share a common minimizer.

# Example #13.1



**Example 13.1** (Role of informed agents over MSE networks). For the MSE network of Example 6.3 with uniform step-sizes and uniform covariance matrices, i.e.,  $\mu_k \equiv \mu$  and  $R_{u,k} \equiv R_u > 0$ , we have

$$H_k = \begin{bmatrix} R_u & 0 \\ 0 & R_u^\top \end{bmatrix} \equiv H, \quad G_k = \sigma_{v,k}^2 \begin{bmatrix} R_u & \times \\ \times & R_u^\top \end{bmatrix} \quad (13.26)$$

Moreover, all costs  $J_k(w)$  share the same minimizer so that  $w^* = w^o$  for any set of informed agents. Using  $h = 2$  for complex data, it follows that expressions (13.23) and (13.24) reduce to



# Example #13.1

$$\alpha_{\text{dist}} \approx 1 - 2\mu \lambda_{\min}(R_u) \left( \sum_{k \in \mathcal{N}_I} p_k \right) \quad (13.27)$$

$$\text{MSD}_{\text{dist,av}} = \frac{\mu M}{h} \left( \sum_{k \in \mathcal{N}_I} p_k \right)^{-1} \left( \sum_{k \in \mathcal{N}_I} p_k^2 \sigma_{v,k}^2 \right) \quad (13.28)$$

where the symbol  $\approx$  in the expression for  $\alpha_{\text{dist}}$  signifies that we are ignoring the higher-order term  $o(\mu)$  for sufficiently small step-sizes. It is now clear that if the set of informed agents is enlarged to  $\mathcal{N}'_I \supset \mathcal{N}_I$ , then the convergence rate improves (i.e., faster convergence with  $\alpha_{\text{dist}}$  becoming smaller). However, from (13.28), the network MSD may decrease, remain unchanged, or increase depending on the values of the noise variances  $\{\sigma_{v,k}^2\}$  at the new informed agents. We illustrate this behavior by considering two cases of interest.

# Example #13.1



Assume first that  $A$  is doubly-stochastic. Then,  $p_k = 1/N$  and the above expressions reduce to:

$$\alpha_{\text{dist}} \approx 1 - 2\mu \left( \frac{N_I}{N} \right) \lambda_{\min}(R_u) \quad (13.29)$$

$$\text{MSD}_{\text{dist,av}} = \frac{\mu M}{h} \frac{1}{N} \left( \frac{1}{N_I} \sum_{k \in \mathcal{N}_I} \sigma_{v,k}^2 \right) \quad (13.30)$$

It is seen that if we add a new informed agent of index  $k' \notin \mathcal{N}_I$ , then the convergence rate improves because  $N_I$  increases but the MSD performance of the network will get worse if

# Example #13.1



$$\left( \frac{1}{N_I + 1} \sum_{k \in \mathcal{N}_{I+1}} \sigma_{v,k}^2 \right) > \left( \frac{1}{N_I} \sum_{k \in \mathcal{N}_I} \sigma_{v,k}^2 \right) \quad (13.31)$$

where  $\mathcal{N}_{I+1} = \mathcal{N}_I \cup \{k'\}$  or, equivalently, if

$$\sigma_{v,k'}^2 > \frac{1}{N_I} \sum_{k \in \mathcal{N}_I} \sigma_{v,k}^2 \quad (13.32)$$

That is, the MSD performance gets worse if the incoming noise power at the newly added agent is worse than the average noise power of the existing informed agents.

# Example #13.1



Let us consider next the case in which the combination weights  $\{a_{\ell k}\}$  are selected according to the averaging rule (which is left-stochastic):

$$a_{\ell k} = \begin{cases} 1/n_k, & \ell \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases} \quad (13.33)$$

in terms of the degrees of the various agents. Recall that  $n_k$  is equal to the number of neighbors that agent  $k$  has. It can be verified that the Perron eigenvector  $p$  is given by:

# Example #13.1



$$p = \left( \sum_{k=1}^N n_k \right)^{-1} \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{bmatrix} \quad (13.34)$$

In this case, expressions (13.27) and (13.28) reduce to

# Example #13.1



$$\alpha_{\text{dist}} \approx 1 - 2\mu \lambda_{\min}(R_u) \left( \frac{\sum_{k \in \mathcal{N}_I} n_k}{\sum_{k=1}^N n_k} \right) \quad (13.35)$$

$$\text{MSD}_{\text{dist,av}} = \frac{\mu M}{h} \left( \frac{1}{\sum_{k=1}^N n_k} \right) \left( \frac{1}{\sum_{k \in \mathcal{N}_I} n_k} \right) \left( \sum_{k \in \mathcal{N}_I} n_k^2 \sigma_{v,k}^2 \right) \quad (13.36)$$

It is again seen that if we add a new informed agent  $k' \notin \mathcal{N}_I$ , then the convergence rate improves. However, the MSD performance of the network will get worse if

# Example #13.1



$$\left( \frac{1}{\sum_{k \in \mathcal{N}_{I+1}} n_k} \right) \left( \sum_{k \in \mathcal{N}_{I+1}} n_k^2 \sigma_{v,k}^2 \right) > \left( \frac{1}{\sum_{k \in \mathcal{N}_I} n_k} \right) \left( \sum_{k \in \mathcal{N}_I} n_k^2 \sigma_{v,k}^2 \right) \quad (13.37)$$

or, equivalently, if

$$n_{k'} \sigma_{v,k'}^2 > \left( \sum_{k \in \mathcal{N}_I} n_k \right)^{-1} \left( \sum_{k \in \mathcal{N}_I} n_k^2 \sigma_{v,k}^2 \right) \quad (13.38)$$



# Example #13.1

37

Lecture #23: Role of Informed Agents

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

where the degrees of the agents are now involved in the inequality *in addition* to the noise variances. The above condition can be expressed in terms of a *weighted* harmonic mean as follows. Introduce the inverse variables

$$x_k \triangleq \frac{1}{n_k \sigma_{v,k}^2}, \quad k \in \mathcal{N}_I \quad (13.39)$$

which consist of the inverses of the noise variances scaled by  $n_k$ . Let  $x_H$  denote the weighted harmonic mean of these variables, with weights  $\{n_k\}$ , which is defined as

# Example #13.1



$$x_H \triangleq \left( \sum_{k \in \mathcal{N}_I} n_k \right) \left( \sum_{k \in \mathcal{N}_I} \frac{n_k}{x_k} \right)^{-1} \quad (13.40)$$

Then, condition (13.38) is equivalent to stating that

$$x'_{k'} \triangleq \frac{1}{n_{k'} \sigma_{v,k'}^2} < x_H \quad (13.41)$$

That is, the MSD performance will get worse if the new inverse variable,  $x'_{k'}$ , is smaller than the weighted harmonic mean of the inverse variables  $\{x_k\}$  associated with the existing informed agents.



# Example #13.1

We illustrate these results numerically for the case of the averaging rule (13.33) with uniform step-sizes across the agents set at  $\mu_k \equiv \mu = 0.002$ . Figure 13.2 shows two versions of the connected network topology with  $N = 20$  agents used in the simulations. In one version, the topology has 14 informed agents and 6 uninformed agents. In the second version, two of the previously uninformed agents are transformed back to the informed state so that the topology now ends up with 16 informed agents. The measurement noise variances,  $\{\sigma_{v,k}^2\}$ , and the power of the regression data, assumed uniform and of the form  $R_{u,k} = \sigma_u^2 I_M$ , are shown in the right and left plots of Figure 13.3, respectively.

# Example #13.1



● informed agent

● un-informed agent

● transitioned back to informed state

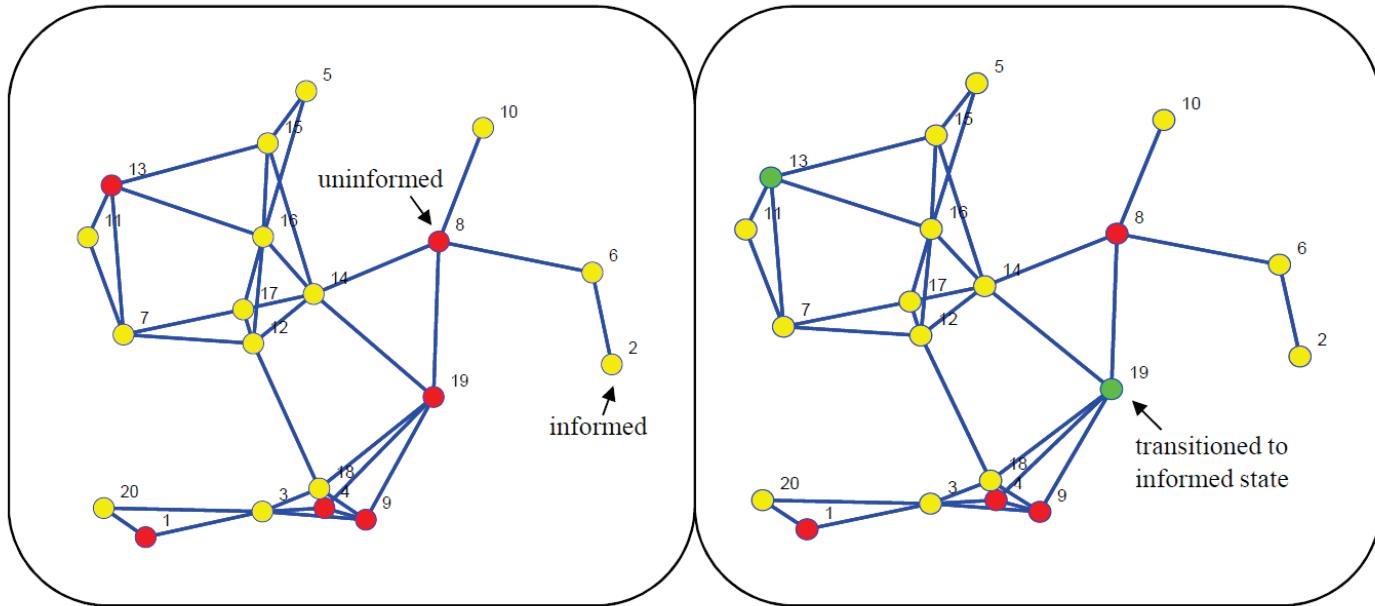


Figure 13.2

# Example #13.1

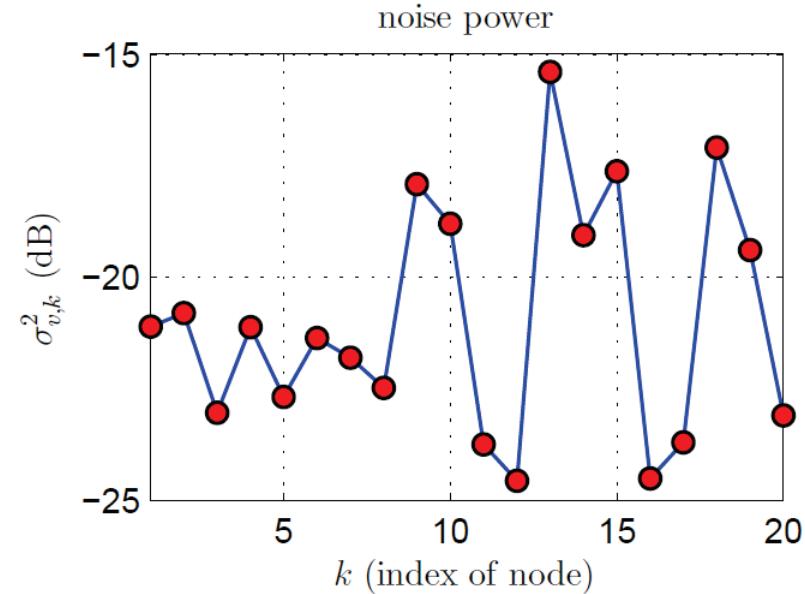
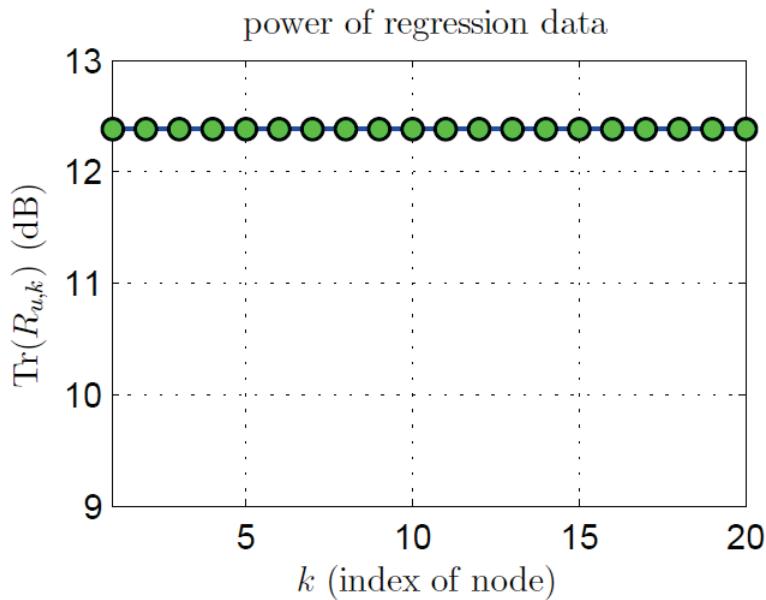


Figure 13.3

# Example #13.1



## Lecture #23: Role of Informed Agents

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

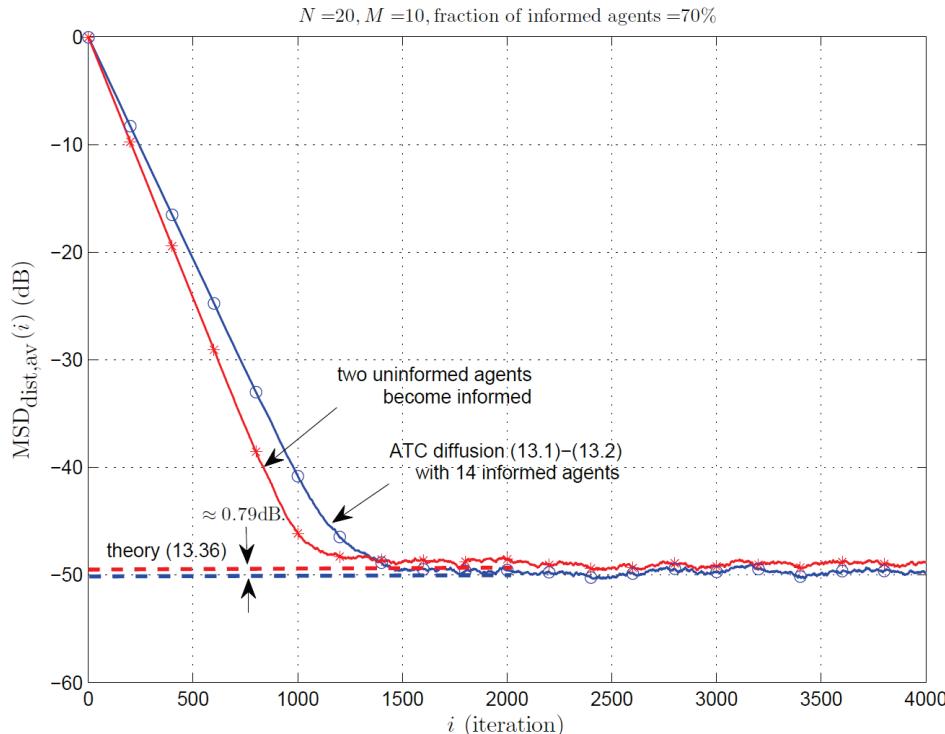


Figure 13.4

# Example #13.1



Figure 13.4 plots the evolution of the ensemble-average learning curves,  $\frac{1}{N}\mathbb{E}\|\tilde{\mathbf{w}}_i\|^2$ , for the ATC diffusion strategy (13.1)–(13.2). The curves are obtained by averaging the trajectories  $\{\frac{1}{N}\|\tilde{\mathbf{w}}_i\|^2\}$  over 200 repeated experiments. The label on the vertical axis in the figure refers to the learning curve  $\frac{1}{N}\mathbb{E}\|\tilde{\mathbf{w}}_i\|^2$  by writing  $\text{MSD}_{\text{dist,av}}(i)$ , with an iteration index  $i$ . Each experiment involves running the ATC diffusion strategy (13.1)–(13.2) with  $h = 2$  on complex-valued data  $\{\mathbf{d}_k(i), \mathbf{u}_{k,i}\}$  generated according to the model  $\mathbf{d}_k(i) = \mathbf{u}_{k,i}w^o + \mathbf{v}_k(i)$ , with  $M = 10$ . The unknown vector  $w^o$  is generated randomly and its norm is normalized to one. The solid horizontal lines in the figure represent the theoretical MSD values obtained from (13.36) for the two scenarios shown in Figure 13.2, namely,

# Example #13.1



$$\text{MSD}(\mathcal{N}_I) \approx -50.19 \text{ dB}, \quad \text{MSD}(\mathcal{N}'_I) \approx -49.40 \text{ dB} \quad (13.42)$$

where  $\mathcal{N}'_I$  denotes the enlarged set of informed agents shown on the right-hand side of Figure 13.2. It is observed in this simulation that when the set of informed agents is enlarged by adding agents #13 and #19, the convergence rate is improved while the MSD value is degraded by about 0.79dB.



# Example #13.2



**Example 13.2** (Performance degradation under fixed convergence rate). We continue with Example 13.1 and the case of the averaging rule (13.33). The current example is based on the discussion from [250] and its purpose is to show that even if we adjust the convergence rate of the network to remain fixed and invariant to the proportion of informed agents, the MSD performance of the network can still deteriorate if the set of informed agents is enlarged. To see this, we set the step-size to the following normalized value:

$$\mu = \mu_o \left( \sum_{k \in \mathcal{N}_I} n_k \right)^{-1} \quad (13.43)$$

# Example #13.2



for some small  $\mu_o > 0$ , and where the normalization is over the sum of the degrees of the informed agents. Note that this selection of  $\mu$  depends on  $\mathcal{N}_I$ . For this choice of  $\mu$ , the convergence rate given by (13.35) becomes

$$\alpha_{\text{dist}} \approx 1 - 2\mu_o \lambda_{\min}(R_u) \left( \sum_{k=1}^N n_k \right)^{-1} \quad (13.44)$$

which is independent of  $N_I$ . Therefore, no matter how the set  $\mathcal{N}_I$  is adjusted, the convergence rate of the network remains fixed. At the same time, the MSD level (13.36) becomes

# Example #13.2



$$\text{MSD}_{\text{dist,av}} = \frac{\mu_o M}{2} \left( \frac{1}{\sum_{k=1}^N n_k} \right) \left( \frac{1}{\sum_{k \in \mathcal{N}_I} n_k} \right)^2 \left( \sum_{k \in \mathcal{N}_I} n_k^2 \sigma_{v,k}^2 \right) \quad (13.45)$$

Some straightforward algebra will show that if we add a new informed agent  $k' \notin \mathcal{N}_I$ , then the MSD performance of the network will get worse if the parameters  $\{n'_k, \sigma_{v,k'}^2\}$  satisfy the inequality:

$$n_{k'} > 2 \left( \sum_{k \in \mathcal{N}_I} n_k \right) \left[ \frac{\left( \sum_{k \in \mathcal{N}_I} n_k \right)^2 \sigma_{v,k'}^2}{\sum_{k \in \mathcal{N}_I} n_k^2 \sigma_{v,k}^2} - 1 \right]^{-1} \quad (13.46)$$

# Example #13.2



We now verify that there exist situations under which the above requirement is satisfied so that the network MSD will end up increasing (an undesirable effect) even though the convergence rate has been set to a constant value.

Consider first the case in which all agents have the same degree, say,  $n_k \equiv n$  for all  $k$ . Then, condition (13.46) becomes

$$\sigma_{v,k'}^2 > \left(2 + \frac{1}{N_I}\right) \left(\frac{1}{N_I} \sum_{k \in \mathcal{N}_I} \sigma_{v,k}^2\right) \quad (13.47)$$

That is, if the new added noise variance is sufficiently larger than the average noise variance at the informed agents, then deterioration in performance will occur.

# Example #13.2



Our second example assumes the noise variances are uniform across all agents, say,  $\sigma_{v,k}^2 \equiv \sigma_v^2$  for all  $k$ . Then, condition (13.46) becomes

$$n'_k > 2 \left( \sum_{k \in \mathcal{N}_I} n_k \right) \left[ \frac{\left( \sum_{k \in \mathcal{N}_I} n_k \right)^2}{\left( \sum_{k \in \mathcal{N}_I} n_k^2 \right)} - 1 \right]^{-1} \quad (13.48)$$

so that if the degree of the new added agent is sufficiently large, then deterioration in performance will occur. The results in these two cases suggest that it is beneficial to keep few highly noisy or highly connected agents uninformed and for them to participate only in the aggregation task (13.2) and to act as relays.



# Controlling Performance Degradation

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.



# Controlling Performance

The previous arguments indicate that the MSD performance need not improve with the addition of informed agents. The deterioration in network performance can be controlled through proper selection of the combination weights, for example, when the matrix  $A$  is selected according to the Hastings rule (12.20). Recall that, under the condition of uniform step-sizes and uniform Hessian matrices, and assuming *all* agents are informed, i.e.,

$$\mu_k \equiv \mu > 0, \quad H_k \equiv H, \quad k = 1, 2, \dots, N \quad (13.49)$$

# Controlling Performance



we derived earlier in (12.21) the following expression for the entries of the optimized Perron eigenvector:

$$p_k^o = \frac{1}{\theta_k^2} \left( \sum_{\ell=1}^N \frac{1}{\theta_\ell^2} \right)^{-1}, \quad k = 1, 2, \dots, N \quad (13.50)$$

Now, assume the gradient noise factors,  $\{\theta_k^2\}$ , that result from assuming all agents are informed are known. Assume further that the partially informed network under study in this chapter (with both informed and uninformed agents) employs the Hastings rule (12.20) that would result



# Controlling Performance

from using the above Perron vector entries. Substituting these entries into (13.23) and (13.24) we find that the convergence rate and the MSD level of the partially informed network are now given by

$$\alpha_{\text{dist}} \approx 1 - 2\mu \lambda_{\min}(H) \left( \sum_{k \in \mathcal{N}_I} \frac{1}{\theta_k^2} \right) \left( \sum_{k=1}^N \frac{1}{\theta_k^2} \right)^{-1} \quad (13.51)$$

$$\text{MSD}_{\text{dist,av}} = \frac{\mu}{2h} \left( \sum_{k=1}^N \frac{1}{\theta_k^2} \right)^{-1} \quad (13.52)$$



# Controlling Performance

54

Lecture #23: Role of Informed Agents

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

We observe that when the agents employ the Hastings rule, the network MSD level becomes independent of  $N_I$  (and, hence, does not change with the addition of informed agents), while the convergence rate decreases (becomes faster) as the set of informed agents is enlarged (since the expression for  $\alpha_{\text{dist}}$  depends on  $N_I$ ).

# Excess-Risk Performance

Course EE210B  
Spring Quarter 2015

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



# ER Performance

We can repeat the analysis of the previous sections and examine how the excess-risk (ER) performance of distributed solutions varies as a function of the fraction of informed agents in the network. The treatment is similar and so we shall be brief. In a manner similar to the study of the MSD metric, the ER performance of distributed solutions with  $N_I$  informed agents can be deduced from (11.186) and is given by:

$$\text{ER}_{\text{dist},k} = \text{ER}_{\text{dist,av}} = \frac{\mu h}{4} \left( \sum_{k \in \mathcal{N}_I} p_k \right)^{-1} \text{Tr} \left( \sum_{k \in \mathcal{N}_I} p_k^2 R_{s,k} \right) \quad (13.53)$$

# ER Performance



where the sum of the  $\{p_k\}$  does not evaluate to one anymore because this sum runs over  $k \in \mathcal{N}_I$  only and not over the entire set of agents. It is again seen from (13.53) that the ER level of the network may increase, remain unchanged, or decrease with the addition of informed agents.

# Example #13.3



**Example 13.3** (Role of informed agents in online learning). We revisit Example 11.9, which deals with a collection of  $N$  learners. Using  $h = 1$  for real data, the ER performance level for the distributed solution, using  $N_I$  informed agents with step-size  $\mu_k \equiv \mu$ , can be deduced from (13.53) as

$$\text{ER}_{\text{dist,av}} = \frac{\mu}{4} \left( \sum_{k \in \mathcal{N}_I} p_k \right)^{-1} \left( \sum_{k \in \mathcal{N}_I} p_k^2 \right) \text{Tr}(R_s) \quad (13.54)$$

# Example #13.3



In particular, it is seen that if we add a new informed agent of index  $k' \notin \mathcal{N}_I$ , then the ER performance levels will get worse if

$$p_{k'} > \left( \sum_{k \in \mathcal{N}_I} p_k \right)^{-1} \left( \sum_{k \in \mathcal{N}_I} p_k^2 \right) \quad (13.55)$$

This condition is in terms of the entries  $\{p_k\}$ , which are determined by the combination policy,  $A$ . We again consider two choices for the combination matrices.



# Example #13.3

60

Lecture #23: Role of Informed Agents

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

Assume first that  $A$  is doubly-stochastic (such as the Metropolis rule (12.43)) so that  $p_k = 1/N$ . Then, condition (13.55) cannot be satisfied and we conclude that, for this case, the addition of informed agents cannot degrade network performance. Indeed, in this scenario, it can be readily seen that the ER expression (13.54) reduces to

$$\text{ER}_{\text{dist,av}} = \frac{\mu}{4} \left( \frac{1}{N} \right) \text{Tr}(R_s) \quad (13.56)$$

Both of these expressions are independent of  $N_I$ ; it is worth noting that in the current problem, the Hastings rule (12.20) reduces to the doubly-stochastic Metropolis rule (12.43), which explains why the ER result (13.56) is independent of  $N_I$ .



# Example #13.3

61

Lecture #23: Role of Informed Agents

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

Let us consider next the case in which the combination weights  $\{a_{\ell k}\}$  are selected according to the averaging rule (13.33). Using (13.34), condition (13.55) would then indicate that the network ER level will degrade if the degree of the newly added informed agent satisfies:

$$n_{k'} > \left( \sum_{k \in \mathcal{N}_I} n_k \right)^{-1} \left( \sum_{k \in \mathcal{N}_I} n_k^2 \right) \quad (13.57)$$



# 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.