Distributed Adaptive Filtering

Lei GUO, Siyu XIE

Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing

NecSys'16, September 8, 2016, Tokyo, Japan

Outline

Outline

1. Background
   - Motivation
   - Centralized vs. distributed
   - Existing literature
   - Basic problems
2. Main results
   - Review of single sensor case
   - New results for sensor network
   - Simulation results
3. Concluding remarks

With the development network technology, distributed adaptive filtering has attracted more and more attention:

- Collaborative spectral sensing in cognitive radio systems.
- Distributed noise cancelation.
- Field monitoring.
- Target localization in biological networks.
- Fish schooling, bee swarming, and bird flight in mobile adaptive networks.
- $\cdots\cdots$

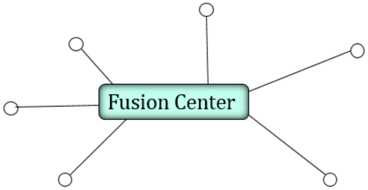
- A fundamental problem in distributed adaptive filtering :

  How to estimate or track an unknown signal process from distributed noisy measurements in a cooperative manner?

- There are basically two approaches: Centralized and Distributed

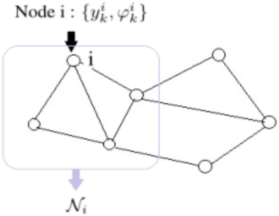
Outline

Centralized processing



Drawbacks: Communication capability, energy consumption, vulnerability in the fusion center.

Distributed processing



- Improve resilience to failure.
- Privacy and secrecy considerations.
- While each sensor is not capable of tracking the desired signal process, the information interaction among the sensors may lead to the desired behavior.

Main strategies

There are three main strategies for distributed processing, namely,

- Incremental strategies
- Diffusion strategies
- Consensus strategies

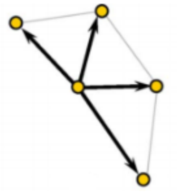
Incremental strategies

Start from a given network topology and a cyclic trajectory that covers all agents in the network, update the estimate one by one along the cyclic trajectory.

$$\widehat{\theta}_k = \widehat{\theta}_k^1 \to \widehat{\theta}_k^2 \to \cdots \to \widehat{\theta}_k^n = \widehat{\theta}_{k+1}$$

Diffusion strategies

There are two types of diffusion strategies: Combine-then-Adapt (CTA) and Adapt-then-Combine (ATC).



Combine: the weighted average of estimates generated by the neighbors of a given sensor.
Adapt: Adaptation using the innovation at a given sensor and other local information.

Consensus strategies

There is no need to select beforehand a cyclic trajectory. All the sensors reach a common value through local communications.

At every iteration k, all agents in the network can run their consensus update simultaneously by using iterates that are available from the iteration $k-1$ of their neighbors.

Outline

Overview

Distributed adaptive filtering has been studied widely in recent years:

- C. G. Lopes and A. H. Sayed, in IEEE ICASSP, 2006.
- A. H. Sayed and C. G. Lopes, IEICE Trans. FECCS, 2007.
- F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, in IEEE SPAWC, 2007.
- I. D. Schizas, G. Mateos and G. B. Giannakis, IEEE on ASSP, 2008.
- I. D. Schizas, G. Mateos and G. B. Giannakis, IEEE TSP, 2009.
- F. S. Cattivelli and A. H. Sayed, IEEE TSP, 2010.
- M.A. Tinati, A. Rastegarnia, A. Khalili, 3rd Conference on WMMN, 2010.
- A. Rastegarnia, M.A. Tinati, A. Khalili, IEEE ICCS, 2010.
- A. H. Sayed, Proceedings of the IEEE, 2014.
- H. Nosrati, M. Shamsi, etc., IEEE TSP, 2015.
- . . .

Almost all require independency and/or stationarity conditions in the theoretical analyses, which exclude applications to feedback systems.

Outline

Basic theoretical problems

- How to establish a theory on distributed adaptive filtering without independency and stationarity assumptions?
- What is the weakest possible information condition, under which the distributed adaptive filtering algorithm can fulfil the estimation task, in the natural case where any individual sensor cannot?
- How far can we extend the existing results for single sensor to distributed sensor networks?

Outline

Single sensor case

For a single sensor whose signals are generated by a stochastic regression model:

$$y_k = \boldsymbol{\theta}_k^T \boldsymbol{\varphi}_k + v_k, \quad k \geq 0.$$

where

$y_k \in \mathbb{R}$: scalar observation at time k

$v_k \in \mathbb{R}$: scalar noise at time k

$\boldsymbol{\varphi}_k \in \mathbb{R}^m$: regressor

$\boldsymbol{\theta}_k \in \mathbb{R}^m$: an unknown signal process to be estimated

Remark: y_k can also be regarded as being approximated or predicted by a linear combination $\boldsymbol{\theta}_k^T \boldsymbol{\varphi}_k$, with the unknown $\boldsymbol{\theta}_k$ to be estimated adaptively.

LMS

The most commonly used least mean squares (LMS) is a type of steepest descent algorithm that aims at minimizing the following mean square error (MSE) recursively:

$$e_k(\boldsymbol{\theta}) = \mathbb{E}(y_k - \boldsymbol{\varphi}_k^T \boldsymbol{\theta})^2, \qquad k \geq 0,$$

with the following standard form:

$$\widehat{\boldsymbol{\theta}}_{k+1} = \widehat{\boldsymbol{\theta}}_k + \mu \boldsymbol{\varphi}_k [y_k - (\boldsymbol{\varphi}_k)^T \widehat{\boldsymbol{\theta}}_k], \qquad k \geq 0,$$

where $\mu \in (0, 1)$ is the adaptation gain.

NLMS

For convenience of discussions, we consider the following normalized LMS:

$$\widehat{\boldsymbol{\theta}}_{k+1} = \widehat{\boldsymbol{\theta}}_k + \mu \frac{\boldsymbol{\varphi}_k}{1+ \parallel \boldsymbol{\varphi}_k \parallel^2} [y_k - (\boldsymbol{\varphi}_k)^T \widehat{\boldsymbol{\theta}}_k], \qquad k \geq 0,$$

where $\mu \in (0, 1)$ is the adaptation gain.

Remark: Similar treatments apply to the unnormalized LMS, save that a general stochastic averaging theorem is established to deal with possible unbounded regression signals.

A brief overview

Most literature require independency and stationarity conditions, e.g.,

- B. Widrow et al. (1976): independency
- S. Haykin (1996): independency
- O. Macchi (1995): M-dependency.
- ······

One exception is the work based on weak convergence where ϕ-mixing condition is used but needs vanishing adaptation gains, see e.g.,

- H. J. Kushner (1984)

In fact, how to relax these restrictions has been a long standing problem in adaptive filtering theory.

In the 1990s, a general theory was established by introducing a "conditional excitation condition" which requires neither independency/stationarity nor vanishing adaptation gains, and is applicable to feedback systems:

- L. Guo (*SICON*, 1994): Stability of NLMS, RLS and KF.
- L. Guo and L. Ljung (IEEE TAC, 1995): Performance of NLMS, RLS, KF and beyond.
- L. Guo,L. Ljung and G.J.Wang (IEEE TAC, 1997): Necessary and sufficient condition for stability of LMS.

These are the basis to our investigation of distributed adaptive filtering.

Error equation

Let us denote

$$\widetilde{\boldsymbol{\theta}}_k = \widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_k,$$
$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \gamma \boldsymbol{\omega}_k,$$

then the estimation error equation can be written as

$$\widetilde{\boldsymbol{\theta}}_{k+1} = \left(I_m - \mu \frac{\boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T}{1 + \|\boldsymbol{\varphi}_k\|^2} \right) \widetilde{\boldsymbol{\theta}}_k + \mu \frac{\boldsymbol{\varphi}_k v_k}{1 + \|\boldsymbol{\varphi}_k\|^2} - \gamma \boldsymbol{\omega}_k, \quad k \geq 0, \mu \in (0, 1).$$

Remark:
The product of random matrices $\prod \left(I_m - \mu \frac{\boldsymbol{\varphi}_k \boldsymbol{\varphi}_k^T}{1+\|\boldsymbol{\varphi}_k\|^2} \right)$ plays a key role.

Definitions

- A sequence $\{I - A_k, k \geq 0\}$ is called L_p-exponentially stable if $A = \{A_k, k \geq 0\}$ belongs to the following family

$$S_p(\lambda) = \left\{ A : \left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_p} \leq N\lambda^{k-i}, \forall k \geq i \geq 0, \exists N > 0 \right\},$$

where $\| \cdot \|_{L_p} = \{\mathbb{E} \| \cdot \|^p\}^{\frac{1}{p}}$ and $\lambda \in [0, 1)$.

- For a scalar sequence $a = \{a_k, k \geq 0\}$ with $a_k \in [0, 1]$ we denote

$$S^0(\lambda) = \left\{ a : \mathbb{E} \prod_{j=i+1}^{k} (1 - a_j) \leq N\lambda^{k-i}, \forall k \geq i \geq 0, \exists N > 0 \right\}.$$

Remark: If a scalar random sequence in $[0, 1]$ is uniformly bounded from below by a positive constant, then obviously it belongs to the family $S^0(\lambda)$.

Conditional excitation condition

There exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$, where λ_k is defined by

$$\lambda_k \triangleq \lambda_{min}\left\{\mathbb{E}\left[\frac{1}{h+1}\sum_{j=k+1}^{k+h}\frac{\varphi_j(\varphi_j)^T}{1+\|\varphi_j}\|^2\Big|\mathcal{F}_k\right]\right\}.$$

and where $\mathcal{F}_k = \sigma\{\varphi_j, \omega_j, v_{j-1}, j \leq k\}$.

Remark: This condition allows more interesting stochastic cases where no lower bound to λ_k exists. It can be verified for the following typical situations:

- ϕ-mixing processes.
- Signals generated by linear and non-linear state space stochastic models.
- Time varying linear stochastic models.

Stability for single LMS

Theorem.
Assume that the conditional excitation condition is satisfied. Then for any $\mu \in (0,1)$ and any $p \geq 1$,

$$\left\{ I_m - \mu \frac{\varphi_k \varphi_k^T}{1 + \|\varphi_k\|^2} \right\} \text{ is } L_p\text{-exponentially stable.}$$

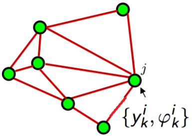
Question : Can we generalize this result to sensor network case?

Outline

Sensor network

For a network with n sensors, consider the signal model at each sensor i:

$$y_k^i = (\varphi_k^i)^T \boldsymbol{\theta}_k + v_k^i, \quad k \geq 0.$$



where

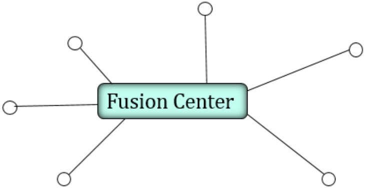
$y_k^i \in \mathbb{R}$: scalar observation of sensor i at time k

$v_k^i \in \mathbb{R}$: scalar noise of sensor i at time k

$\varphi_k^i \in \mathbb{R}^m$: regressor of sensor i

$\boldsymbol{\theta}_k \in \mathbb{R}^m$: an unknown signal process to be estimated

Centralized adaptive filtering



A natural centralized algorithm may be deduced by minimizing the following MSE recursively

$$e_k^{cen}(\boldsymbol{\theta}) = \mathbb{E}\left\{ \frac{1}{n} \sum_{i=1}^{n} [y_k^i - (\varphi_k^i)^T \boldsymbol{\theta}]^2 \right\}, \qquad k \geq 0.$$

Centralized adaptive filtering

Centralized NLMS algorithm

$$\widehat{\boldsymbol{\theta}}_{k+1}^{cen} = \widehat{\boldsymbol{\theta}}_{k}^{cen} + \mu\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{\varphi}_k^i}{1+\parallel \boldsymbol{\varphi}_k^i \parallel^2}[y_k^i - (\boldsymbol{\varphi}_k^i)^T\widehat{\boldsymbol{\theta}}_k^{cen}]\right\}$$

where $\mu \in (0,1)$ is the adaptation gain.

Denote $\widetilde{\boldsymbol{\theta}}_k^{cen} = \widehat{\boldsymbol{\theta}}_k^{cen} - \boldsymbol{\theta}_k$, we have the error equation

$$\widetilde{\theta}_{k+1}^{cen} = \left\{ I_m - \frac{\mu}{n}\sum_{i=1}^{n}\frac{\varphi_k^i(\varphi_k^i)^\tau}{1+\parallel \varphi_k^i \parallel^2}\right\}\widetilde{\theta}_k^{cen} + \frac{\mu}{n}\sum_{i=1}^{n}\frac{\varphi_k^i v_k^i}{1+\parallel \varphi_k^i \parallel^2} - \gamma\omega_{k+1},$$

Cooperative information condition

There exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$, where λ_k is defined by

$$\lambda_k \triangleq \lambda_{min}\left\{\mathbb{E}\left[\frac{1}{n(h+1)}\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\varphi_j^i(\varphi_j^i)^T}{1+\|\varphi_j^i\|^2}\Big|\mathcal{F}_k\right]\right\}.$$

and where $\mathcal{F}_k = \sigma\{\varphi_j^i, \omega_j, v_{j-1}^i, j \leq k, i = 1, \ldots, n\}$.

Remark: This condition is a natural extension of the single sensor case

$$\lambda_k \triangleq \lambda_{min}\left\{\mathbb{E}\left[\frac{1}{h+1}\sum_{j=k+1}^{k+h}\frac{\varphi_j(\varphi_j)^T}{1+\|\varphi_j\|^2}\Big|\mathcal{F}_k\right]\right\}.$$

where n is taken to be 1.

Stability of centralized LMS

Theorem:
Consider the centralized NLMS algorithm. If the cooperative information condition holds, then for any $\mu \in (0,1)$ and any $p \geq 1$,

$$\left\{ I_m - \frac{\mu}{n} \sum_{i=1}^{n} \frac{\varphi_k^i (\varphi_k^i)^\tau}{1+ \| \varphi_k^i \|^2} \right\} \text{ is } L_p - \text{exponentially stable.}$$

Question : Can we establish the same result for distributed adaptive filtering under the same cooperative information condition as in the centralized case?

Distributed adaptive filtering

The network connections are modeled as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$. The adjacency matrix $\mathcal{A} = \{a_{ij}\}$ reflects the interaction strength among neighboring nodes and the set of neighbors of each sensor k is denoted as

$$\mathcal{N}_k = \{l \in \mathcal{V} | (k, l) \in \mathcal{E}\}.$$

Distributed adaptive filtering

Each sensor i tries to minimize the following performance index composed of local information:

$$e_k^i(\boldsymbol{\theta}) = \mathbb{E}\left\{ \left[\frac{y_k^i - (\varphi_k^i)^T \boldsymbol{\theta}}{\sqrt{1 + \| \varphi_k^i \|^2}} \right]^2 + \nu \sum_{l \in \mathcal{N}_i} a_{li} (\widehat{\boldsymbol{\theta}}_k^l - \boldsymbol{\theta})^2 \right\},$$

$$k \geq 0, i = 1, \dots, n.$$

Remark: The first part corresponds to the usual NLMS, while the second part tries to minimize the weighted distance between the estimate of the agent i and its neighboring estimates.

Distributed adaptive filtering

The consensus type distributed NLMS algorithm :

$$\widehat{\theta}^i_{k+1} = \underbrace{\widehat{\theta}^i_k + \mu\Big\{ \frac{\varphi^i_k}{1+\parallel \varphi^i_k \parallel^2}[y^i_k - (\varphi^i_k)^T \widehat{\theta}^i_k]}_{error\ \ correction\ \ term}$$

$$\underbrace{-\nu \sum_{l \in \mathcal{N}_i} a_{li}(\widehat{\theta}^i_k - \widehat{\theta}^l_k)}_{consensus\ \ term}\Big\}, \qquad k \geq 0, i = 1, \ldots, n.$$

where $\mu \in (0,1), \nu \in (0,1)$ are adaptation gains.

Error equation: vector form

$$\begin{aligned}\widetilde{\boldsymbol{\Theta}}_{k+1} =&\{I_{mn} - \mu[\mathbf{F}_k + \nu(\mathcal{L} \otimes I_m)]\}\widetilde{\boldsymbol{\Theta}}_k + \mu \mathbf{L}_k \mathbf{V}_k - \gamma \boldsymbol{\Omega}_{k+1} \\ =&(I_{mn} - \mu \mathbf{G}_k)\widetilde{\boldsymbol{\Theta}}_k + \mu \mathbf{L}_k \mathbf{V}_k - \gamma \boldsymbol{\Omega}_{k+1}.\end{aligned}$$

where

$$\widetilde{\boldsymbol{\Theta}}_k \triangleq \mathrm{col}\{\widetilde{\boldsymbol{\theta}}_k^1, \ldots, \widetilde{\boldsymbol{\theta}}_k^n\}, \text{ where } \widetilde{\boldsymbol{\theta}}_k^i = \widehat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k,$$

$$\mathbf{G}_k \triangleq \mathbf{F}_k + \nu(\mathcal{L} \otimes I_m), \quad \mathbf{F}_k \triangleq \mathbf{L}_k \boldsymbol{\Phi}_k^T,$$

$$\mathbf{L}_k \triangleq \mathrm{diag}\left\{\frac{\varphi_k^1}{1+ \| \varphi_k^1 \|^2}, \ldots, \frac{\varphi_k^n}{1+ \| \varphi_k^n \|^2}\right\},$$

$$\boldsymbol{\Phi}_k \triangleq \mathrm{diag}\{\varphi_k^1, \ldots, \varphi_k^n\},$$

$$\mathbf{V}_k \triangleq \mathrm{col}\{v_k^1, \ldots, v_k^n\}, \quad \boldsymbol{\Omega}_{k+1} \triangleq \mathrm{col}\{\underbrace{\boldsymbol{\omega}_{k+1}, \ldots, \boldsymbol{\omega}_{k+1}}_{n}\}.$$

Conditions on topology and information

- *Condition* 1: The graph $\mathcal{G}$ is connected.

- *Condition* 2 (Cooperative Information Condition): There exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$, where λ_k is defined by

$$\lambda_k \triangleq \lambda_{min}\left\{\mathbb{E}\left[\frac{1}{n(h+1)}\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\varphi_j^i(\varphi_j^i)^T}{1+\|\varphi_j^i\|^2}\Big|\mathcal{F}_k\right]\right\}.$$

and where $\mathcal{F}_k = \sigma\{\varphi_j^i, \omega_j, v_{j-1}^i, j \leq k, i = 1, \ldots, n\}$.

A key lemma

Lemma 1: For any $\mu \in (0, \frac{1}{3}), \nu \in (0, 1)$, suppose that the graph $\mathcal{G}$ is connected and the cooperative information condition is satisfied, then $\rho_k \in S^0(\rho)$, where

$$\rho_k \triangleq \lambda_{min}\left\{\mathbb{E}\left[\frac{1}{1+h}\sum_{j=k+1}^{k+h} \mu \mathbf{G}_j \middle| \mathcal{F}_k\right]\right\},$$

and $\rho = \lambda^{\epsilon}, \epsilon = \frac{h}{h^2+2h+1} \cdot \delta_{m+1}\mu\nu, \delta_{m+1}$ is the $m+1$-th eigenvalue of $\mathcal{L} \otimes I_m$.

Remark: Transform the stochastic property of "summation" to that of "product" of the random matrices under cooperation.

Stability result

Theorem 1: Suppose that the graph $\mathcal{G}$ is connected and the cooperative information condition is satisfied. Then for any $\mu \in (0, \frac{1}{3}), \nu \in (0, 1), p \geq 1$, we have

$$\{I_{mn} - \mu \mathbf{G}_k, k \geq 1\} \text{ is } L_p - \text{exponentially stable.}$$

Furthermore, if for some $p \geq 1$ and $\beta > 1$,

$$\sigma_p \stackrel{\triangle}{=} \sup_k \|\xi_k log^{\beta}(e + \xi_k)\|_{L_p} < \infty, \|\widetilde{\mathbf{\Theta}}_0\|_{L_p} < \infty$$

hold where $\xi_k = \|\mathbf{V}_k\| + \|\mathbf{\Omega}_{k+1}\|$, then we have

$$\limsup_{k \to \infty} \|\widetilde{\mathbf{\Theta}}_k\|_{L_p} \leq c[\sigma_p log(e + \sigma_p^{-1})],$$

where c is a positive constant.

Necessity result

Theorem 2: Let $\{\varphi_k^i\}$ be ϕ-mixing processes and suppose that the graph $\mathcal{G}$ is connected. Then for any $\mu \in (0, \frac{1}{3}), \nu \in (0, 1)$,

$\{I_{mn} - \mu \mathbf{G}_k, k \geq 1\}$ is L_p-exponentially stable $(p \geq 1)$ **if and only if** the cooperative information condition is satisfied.

Further investigation

A random sequence $x = \{x_k\} \in \mathcal{M}_p (p \geq 1)$, if there exists a constant C_p^x depending only on p and the distribution of $\{x_k\}$ such that for any $k \geq 0$,

$$\left\| \sum_{i=k+1}^{k+h} x_i \right\|_{L_p} \leq C_p^x h^{\frac{1}{2}}, \quad \forall h \geq 1.$$

- *Condition 3*: For some $p \geq 1$, the initial estimation error is bounded, i.e. $\|\widetilde{\mathbf{\Theta}}_0\|_{L_{2p}} < \infty$, . Furthermore, let $\{\mathbf{L}_k \mathbf{V}_k\} \in \mathcal{M}_{2p}$ and $\{\mathbf{\Omega}_k\} \in \mathcal{M}_{2p}$.

Remark:
This condition simply implies that both the noises process and parameter variations are weakly dependent in a certain sense.

Further result

Theorem . Assume that Conditions 1-3 are satisfied, then for any $k \geq 0$ and $\mu \in (0, \frac{1}{3}), \nu \in (0, 1)$, we have

$$\|\widetilde{\Theta}_{k+1}\|_{L_p} = O\left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} + (1 - \alpha\mu)^{k+1}\right),$$

where $\alpha \in (0, 1)$ is a constant.

Remark:
The upper bound roughly indicates the tradeoff between noise sensitivity and tracking ability. More accurate results will be given below.

Performance approximation

Theorem 3: Under some further mild conditions on the observation noise and parameter variation, we have for any $k \geq 1$, $\mu \in (0, \frac{1}{3}), \nu \in (0,1)$

$$\|\mathbb{E}[\widetilde{\mathbf{\Theta}}_{k+1}\widetilde{\mathbf{\Theta}}_{k+1}^T] - \widehat{\mathbf{\Pi}}_{k+1}\| \leq c\bar{\delta}(\mu)\left[\mu + \frac{\gamma^2}{\mu} + (1-\alpha\mu)^{k+1}\right],$$

where $c > 0, \alpha \in (0,1)$ are constants and $\bar{\delta}(\mu)$ tends to zero as $\mu \to 0$ and $\widehat{\mathbf{\Pi}}_{k+1}$ is the main term of the estimation error covariance, which can be calculated recursively by

$$\begin{aligned}\widehat{\mathbf{\Pi}}_{k+1} =& (I_{mn} - \mu\mathbb{E}[\mathbf{G}_k])\widehat{\mathbf{\Pi}}_k(I_{mn} - \mu\mathbb{E}[\mathbf{G}_k])^T \\ &+ \mu^2\mathbb{E}[\mathbf{L}_k\mathbf{V}_k\mathbf{V}_k^T\mathbf{L}_k^T] + \gamma^2\mathbb{E}[\mathbf{\Omega}_{k+1}\mathbf{\Omega}_{k+1}^T].\end{aligned}$$

Simplifications

Let the regressors be (wide-sense) stationary, and let us denote

$$\mathbf{F} = \mathbb{E}[\mathbf{F}_k] = \mathrm{diag}\{\mathbf{F}^1, \cdots, \mathbf{F}^n\}, \quad \mathbf{G} = \mathbf{F} + \nu(\mathcal{L} \otimes I_m),$$
$$\mathbf{T} = \mathbb{E}[\mathbf{T}_k] = \mathbb{E}[\mathbf{L}_k \mathbf{V}_k \mathbf{V}_k^T \mathbf{L}_k^T], \quad \mathbf{Q}_\omega = \mathbf{Q}_\omega(k+1) = \mathbb{E}[\mathbf{\Omega}_{k+1}\mathbf{\Omega}_{k+1}^T].$$

Then the "main term" can be simplified as

$$\mathbf{\Pi}_k = \mu \bar{\mathbf{R}}_v + \frac{\gamma^2}{\mu}\bar{\mathbf{R}}_\omega + O\left(\bar{\delta}(\mu)\left[\mu + \frac{\gamma^2}{\mu}\right]\right) + o(1),$$

where "$o(1)$" tends to zero with exponential rate as $k \to \infty$, and

$$\bar{\mathbf{R}}_v = \int_0^\infty e^{-\mathbf{G}t}\mathbf{T}e^{-\mathbf{G}t}dt, \quad \bar{\mathbf{R}}_\omega = \int_0^\infty e^{-\mathbf{G}t}\mathbf{Q}_\omega e^{-\mathbf{G}t}dt.$$

Performance "optimization"

Note that $lim_{\mu \to 0} \bar{\delta}(\mu) = 0$. As a result, we have for all small μ and large k

$$\mathbf{\Pi}_k \sim \mu \bar{\mathbf{R}}_v + \frac{\gamma^2}{\mu} \bar{\mathbf{R}}_\omega,$$

which indicates that μ should be proportional to γ, and by minimizing the right-hand-side, we get the "optimal" choice $\mu^* = \gamma \sqrt{tr \bar{\mathbf{R}}_\omega / tr \bar{\mathbf{R}}_v}$ with the corresponding minimum value:

$$\sum_{i=1}^{n} \mathbb{E} \| \widetilde{\boldsymbol{\theta}}_k^i \|^2 \sim 2\gamma \sqrt{tr \bar{\mathbf{R}}_\omega \cdot tr \bar{\mathbf{R}}_v}.$$

Outline

Simulation examples

Example 1. We construct an example to illustrate the cooperative property of the distributed adaptive filtering :

- No individual sensor can estimate the parameters by itself, but the whole distributed sensor network can.

Let us take $n = 3$ with a connected graph. We will track an unknown 3-dimensional signal $\boldsymbol{\theta}_k$. Let the variation be $\boldsymbol{\omega}_k \sim N(0, 0.1, 3, 1)$ (Gaussian distribution), the observation model $y_k^i = (\varphi_k^i)^T \boldsymbol{\theta}_k + v_k^i (i = 1, 2, 3)$ with noises $v_k^i \sim N(0, 0.1)$.

Simulation results

Let $\varphi_k^i (i = 1, 2, 3)$ be generated by

$$\begin{cases} \mathbf{x}_k^i = & A_i \mathbf{x}_{k-1}^i + B_i \xi_k^i, \\ \varphi_k^i = & C_i \mathbf{x}_k^i, \end{cases}$$

where $\xi_k^i \sim U(-1, 1)$ (uniform distribution), and

$$A_1 = A_3 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/5 \end{pmatrix}, A_2 = \begin{pmatrix} 4/5 & 0 & 0 \\ 4/5 & 0 & 0 \\ 4/5 & 0 & 0 \end{pmatrix},$$

$$B_1 = (1, 0, 0)^T, B_2 = (1, 0, 0)^T, B_3 = (1, 0, 0)^T,$$

$$C_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Simulation results

Let $\mathbf{x}_0^1 = \mathbf{x}_0^2 = \mathbf{x}_0^3 = (1,1,1)^T, \boldsymbol{\theta}_0 = (1,1,1)^T, \hat{\boldsymbol{\theta}}_0^i = (0,0,0)^T, \mu = 0.3, \nu = 0.8$, plot the tracking error covariances.

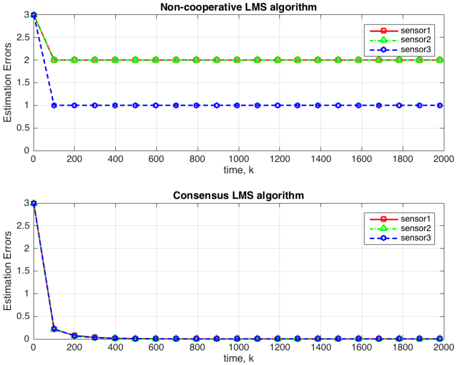


Figure: Tracking error covariances with $\gamma = 0$

Simulation results

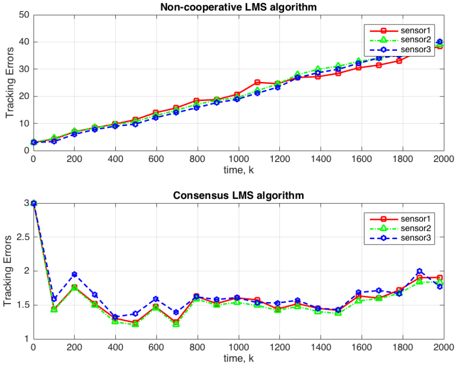


Figure: Tracking error covariances with $\gamma = 1$

Simulation results

Example 2. We construct another example to show that the full rank property 3 of the matrix is necessary. We assume that $A_i, B_i (i = 1, 2, 3)$ remain the same but with

$$C_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It is not difficult to verify that the related matrix in the cooperative information condition has rank 2 (not 3).

Simulation results

We select the same initial states as in the first example.

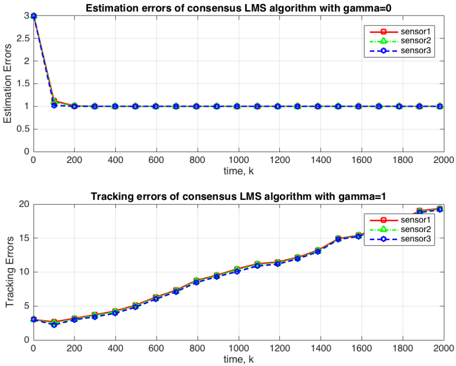


Figure: Tracking error covariances of the consensus LMS algorithm

Conclusions

- We have presented a weakest possible cooperative information condition, under which the distributed adaptive filtering algorithm can fulfil the estimation task, even when any individual sensor cannot. This gives a rigorous justification for the cooperation property of distributed filtering.
- This general cooperative information condition does not exclude applications of the distributed filtering theory to stochastic feedback systems, a desirable property that has rarely been achieved before.

Conclusions

- The cooperative information condition is also necessary for the stability of distributed adaptive filtering algorithm, for the commonly used ϕ-mixing processes.
- We have also shown that the actual tracking error covariance matrix can be well approximated by a simple linear and deterministic difference matrix equation which can be easily evaluated, analyzed, and even "optimized".

Further problems

- To extend the stability theorems to other distributed strategies and other adaptive filtering algorithms (e.g., RLS, KF based filters).
- To combine distributed adaptive filtering with distributed adaptive control problems.
- To expand the scope of applications to more complex dynamical systems.

THANK YOU!