Network Science
IST-CMU PhD course
Fall 2016
Instructor: `jxavier@isr.ist.utl.pt`
TA: João Martins, `joaoa@andrew.cmu.edu`

Important: the homework is due the December 12. Send a scanned pdf file with your answers (or typed in LaTeX, if you prefer) to the TA's email.

Homework 4

1. *Distributed detection.* A set of n agents observes a data stream $y(1), y(2), y(3), \ldots$, where $y(t) = (y_1(t), y_2(t), \ldots, y_n(t))$ is the network observation at time t; $y_i(t)$ is the private observation of agent i at time t. The data stream $(y(t))_{t\geq 1}$ is an independent and identically distributed sequence (i.i.d.) of random vectors generated by one of two stochastic sources:

$$\begin{aligned} H_0 &: y(t) \sim \mathcal{N}\left(0, \sigma^2 I\right) \\ H_1 &: y(t) \sim \mathcal{N}\left(\mu, \sigma^2 I\right) \end{aligned}$$

where $\mu = (\mu_1, \ldots, \mu_n) \in \mathbf{R}^n - \{0\}$ and $\sigma > 0$. In other words, all the network observations are either generated from a gaussian distribution with zero mean (hypothesis H_0) or from a gaussian distribution with mean $\mu \neq 0$ (hypothesis H_1). Note that, regardless of the active hypothesis, the $n \times n$ covariance matrix of the observations is $\sigma^2 I$.

Assume that the two hypotheses are equally probable. The optimal central detector, that sees the network observations until time t, decides as

$$\underbrace{\frac{1}{nt} \sum_{s=1}^{t} \log \frac{P_1\left(y(s)\right)}{P_0\left(y(s)\right)}}_{\ell(t)} \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{1}$$

where P_0 and P_1 correspond to the gaussian probabilities density functions (pdf) of hypotheses H_0 and H_1:

$$P_0(y) = \frac{1}{(2\pi)^{n/2}\, \sigma^n} e^{-\frac{\|y\|^2}{2\sigma^2}}, \qquad P_1(y) = \frac{1}{(2\pi)^{n/2}\, \sigma^n} e^{-\frac{\|y-\mu\|^2}{2\sigma^2}}.$$

Let $P_{\mathrm{e}}(t)$ be the probability of error of the centralized detector, at time t. We know that $P_{\mathrm{e}}(t)$ decays exponentially fast to zero. More precisely, we have

$$\lim_t -\frac{\log P_{\mathrm{e}}(t)}{t} = C$$

where C is the Chernoff distance between the two pdfs P_0 and P_1:

$$C = \frac{1}{8} \frac{\|\mu\|^2}{\sigma^2}.$$

The decision statistic $\ell(t)$ in (1) evolves as

$$\ell(t+1) = \frac{t}{t+1} \ell(t) + \frac{1}{n(t+1)} \log \frac{P_1\left(y(t+1)\right)}{P_0\left(y(t+1)\right)}$$

and motivates the distributed "tracker"

$$L(t+1) = \frac{t}{t+1} W L(t) + \frac{1}{t+1} l(t+1) \tag{2}$$

where $l(t) = (l_1(t), \ldots, l_n(t))$,

$$l_i(t) = \log \frac{P_{1;i}\left(y_i(t)\right)}{P_{0;i}\left(y_i(t)\right)}$$

is the log-likelihood of the data observed at agent i and time t, and $P_{0;i}$, $P_{1;i}$ are the marginals (at agent i) of the network pdfs P_0 and P_1:

$$P_{0;i}(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y_i^2}{2\sigma^2}}, \qquad P_{1;i}(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu_i)^2}{2\sigma^2}}.$$

The entry $L_i(t)$ of the vector $L(t) = (L_1(t), \ldots, L_n(t))$ represents the guess of agent i, at time t, of the central statistic $\ell(t)$. Thus, agent i, at time t, decides as

$$L_i(t) \underset{H_0}{\overset{H_1}{\gtrless}} 0.$$

Assume that the weight matrix W in (2) is symmetric and leads to consensus:

$$W^t \to J = \frac{1}{n}\mathbf{1}\mathbf{1}^T, \qquad \text{as } t \to \infty.$$

In this problem, you will show that the asymptotic performance of the distributed detector matches the asymptotic performance of the centralized detector. That is, you will show that the probability of error of agent i at time t, denoted $P_{\mathrm{e};i}(t)$, decays at the same rate of the probability of error of the centralized detector:

$$\lim_t -\frac{\log P_{\mathrm{e};i}(t)}{t} = C \tag{3}$$

for any agent $i = 1, \ldots, n$. Note that, in class, we have considered gaussian distributions with common mean across the agents: under H_0 the common mean was $-A/2$; under H_1 the common mean was $A/2$. Here, we allow different means under H_1, that is, we are not assuming that $\mu = (\mu_1, \ldots, \mu_n)$ is of the form $\mu = c\mathbf{1}$ for some $c \in \mathbf{R} - \{0\}$. We can have $\mu_i \neq \mu_j$ for some i and j. The proof that we saw in class doesn't work for this extension.

(a) Show that

$$L(t) = \frac{1}{t} W^{t-1} l(1) + \frac{1}{t} W^{t-2} l(2) + \cdots + \frac{1}{t} W l(t-1) + \frac{1}{t} l(t) \tag{4}$$

for all t.

(b) The probability of error of agent i at time t is given by

$$P_{\mathrm{e};i}(t) = \frac{1}{2}\mathbf{P}_0\left(L_i(t) > 0\right) + \frac{1}{2}\mathbf{P}_1\left(L_i(t) \leq 0\right)$$

where $\mathbf{P}_0$ and $\mathbf{P}_1$ are the probabilities under hypothesis H_0 and H_1. You are going to show that

$$\liminf_t -\frac{\log \mathbf{P}_0\left(L_i(t) > 0\right)}{t} \geq C. \tag{5}$$

The proof of

$$\liminf_t -\frac{\log \mathbf{P}_1\left(L_i(t) \leq 0\right)}{t} \geq C \tag{6}$$

is similar and will not be done. Note that (5) and (6) together imply (3) because, due to the general theory of detection, we also know that

$$\limsup_t -\frac{\log \mathbf{P}_0\left(L_i(t) > 0\right)}{t} \leq C, \qquad \limsup_t -\frac{\log \mathbf{P}_1\left(L_i(t) \leq 0\right)}{t} \leq C.$$

So, from now on, suppose that H_0 is the active hypothesis, that is, all $y(t)$ are drawn from the gaussian distribution $\mathcal{N}\left(0, \sigma^2 I\right)$. Show that $l(t) = (l_1(t), \ldots, l_n(t))$ is distributed as $\mathcal{N}\left(-\frac{1}{2\sigma^2}u, \frac{1}{\sigma^2}D(u)\right)$ where $u = (u_1, \ldots, u_n)$, $u_i := \mu_i^2$, and $D(u)$ is the diagonal matrix

$$D(u) = \begin{bmatrix} u_1 & & & \\ & u_2 & & \\ & & \ddots & \\ & & & u_n \end{bmatrix}.$$

(c) Show that $L(t)$ is distributed as $\mathcal{N}\left(\mu(t), \Sigma(t)\right)$ where

$$\mu(t) = -\frac{1}{2\sigma^2 t}\sum_{s=0}^{t-1} W^s u, \qquad \Sigma(t) = \frac{1}{\sigma^2 t^2}\sum_{s=0}^{t-1} W^s D(u) W^s.$$

(d) Note that, for any $\alpha > 0$,

$$\begin{aligned} \mathbf{P}_0\left(L_i(t) > 0\right) &= \mathbf{P}_0\left(e^{\alpha t L_i(t)} > 1\right) \\ &\leq \mathbf{E}_0\left(e^{\alpha t L_i(t)}\right) \end{aligned}$$

where we used Markov's inequality in the last step. Thus,

$$\log \mathbf{P}_0\left(L_i(t) > 0\right) \leq \log \mathbf{E}_0\left(e^{\alpha t L_i(t)}\right).$$

Here, the symbol $\mathbf{E}_0$ stands for the expectation operator under hypothesis H_0, that is, we assume the active hypothesis is H_0.
Show that

$$\log \mathbf{P}_0\left(L_i(t) > 0\right) \leq \alpha t e_i^T \mu(t) + \frac{1}{2}\alpha^2 t^2 e_i^T \Sigma(t) e_i \tag{7}$$

where $e_i \in \mathbf{R}^n$ is the ith column of the $n \times n$ identity: $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, with entry 1 in the ith coordinate.
Hint: note that $L_i(t) = e_i^T L(t)$ and use the fact that, for a scalar gaussian random variable $X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right)$, there holds

$$\log \mathbf{E}\left(e^X\right) = \mu_X + \frac{1}{2}\sigma_X^2.$$

(e) Minimize the right-hand side of (7) with respect to $\alpha > 0$ to show that

$$\log \mathbf{P}_0\left(L_i(t) > 0\right) \le -\frac{\left(e_i^T \mu(t)\right)^2}{2 e_i^T \Sigma(t) e_i}.$$

(f) Show that

$$\liminf_t -\frac{\log \mathbf{P}_0\left(L_i(t) > 0\right)}{t} \ge C.$$

2. *Consensus with noisy communication channels (or yet another nice application of the Robbins-Siegmund's supermartingale convergence lemma).* Consider the classical consensus iterations

$$x(t+1) = W x(t), \tag{8}$$

initialized with $x(0) = \theta$. Under the proper conditions on the weight matrix W (say, W consists of Metropolis weights), we have $x(t) \to \overline{\theta}\mathbf{1}$, where $\overline{\theta} = \mathbf{1}^T\theta/n$.

In practice, iterations (8) are implemented by transmitting the agents' states over physical communication channels (represented by the edges of the underlying communication graph). Model (8) assumes the channels are noiseless: for example, at time t, agent j transmits his state $x_j(t)$ over a channel to a neighbor agent i; agent i receives the state $x_j(t)$ without noise and proceeds to mix it with the states he received from the remaining neighbors. What if the channels add noise? Then, the state $x_j(t)$ is received at agent i as $\widehat{x}_{ij}(t) = x_j(t) + v_{ij}(t+1)$, where $v_{ij}(t+1)$ is the noise added by the channel from agent j to agent i.

So, when the communication channels add noise, the algorithm that is actually executed is no longer (8) but

$$x(t+1) = W x(t) + v(t+1) \tag{9}$$

where $v(t) = (v_1(t), \ldots, v_n(t))$,

$$v_i(t) = \sum_{j=1}^{n} w_{ij} v_{ij}(t).$$

It turns out that the recursions (9) do *not* lead to consensus. To illustrate, consider the network in figure 1. Figure 2 shows how the network states evolve. We see that the states wander about; they do not converge to $\overline{\theta} = 0$.

Consider now the recursion

$$x(t+1) = x(t) + \alpha(t+1)\left(y(t) - x(t)\right) + \beta(t+1)\left(\theta - x(t)\right) \tag{10}$$

where

$$y(t) := W x(t) + v(t+1) \tag{11}$$

models, as in (9), the result of mixing states communicated over noisy channels. Note that (10) can be implemented in a distributed manner. In (10),

$$\alpha(t) = \frac{1}{t^a}, \qquad \beta(t) = \frac{1}{t},$$

are given step sizes, with

$$0.5 < a < 1. \tag{12}$$

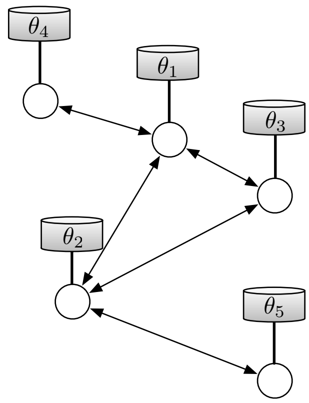

Figure 1: A set of five agents linked by undirected communication channels. Agent i holds θ_i. We consider $\theta_1 = -5$, $\theta_2 = -3$, $\theta_3 = 0$, $\theta_4 = 3$, $\theta_5 = 5$. So, $\overline{\theta} = 0$.

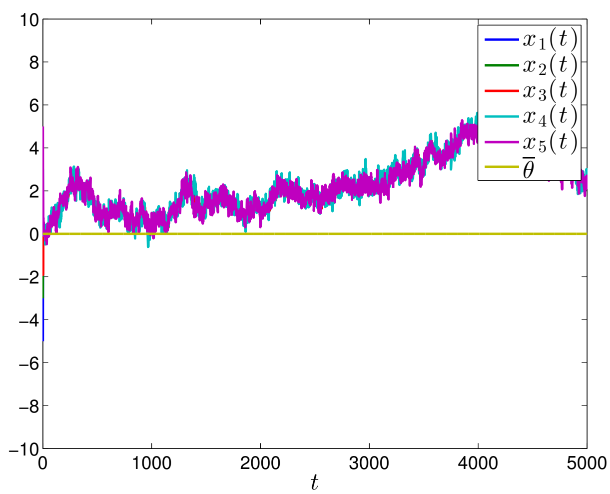

Figure 2: Trajectories of the recursions (9) with i.i.d. gaussian channel noise $v_{ij}(t) \sim \mathcal{N}(0, 0.1)$ and a Metropolis weight matrix W.

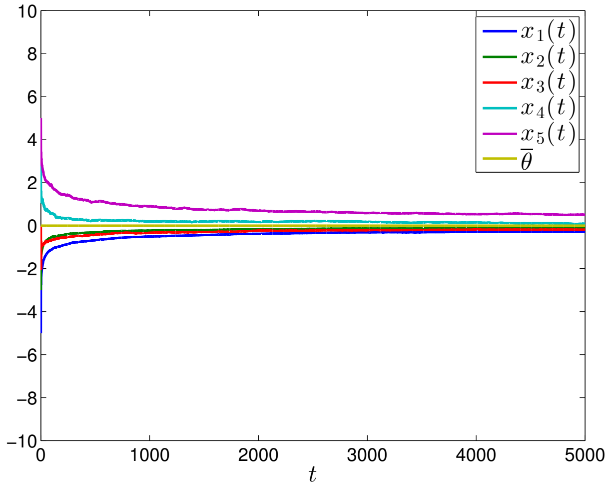

Figure 3: Trajectories of the recursions (10) with i.i.d. gaussian channel noise $v_{ij}(t) \sim \mathcal{N}(0, 0.1)$ and a Metropolis weight matrix W.

(Spoiler alert: in the exam, we will discuss the motivation behind (10).)

In this problem, you will show that (10) is immune to the channels' noise. To illustrate, figure 3 shows the network states, evolving as (10); we see that the all states converge to $\overline{\theta} = 0$ (given enough time). Note that, due to the noise term $v(t)$, the sequence of networks states $(x(t))_{t \geq 0}$ is random.

From now on, we consider $v_{ij}(t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.

Let $e(t) = x(t) - \overline{\theta}\mathbf{1}$ be the error at time t. You will show that $e(t) \to 0$ by showing that $\overline{e}(t) \to 0$ and $\widehat{e}(t) \to 0$, where $\overline{e}(t) = \mathbf{1}^T e(t)/n$ and $\widehat{e}(t) = U^T e(t)$. Recall that $U \in \mathbf{R}^{n \times (n-1)}$ is the matrix involved in the eigenvalue decomposition of the weight matrix W:

$$W = \begin{bmatrix} \frac{1}{\sqrt{n}}\mathbf{1} & U \end{bmatrix} \begin{bmatrix} 1 & \\ & \Lambda \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{n}}\mathbf{1}^T \\ U^T \end{bmatrix},$$

with $\|\Lambda\| < 1$.

(a) We start by plugging (11) into (10):

$$x(t+1) = x(t) + \alpha(t+1)\left(Wx(t) + v(t+1) - x(t)\right) + \beta(t+1)\left(\theta - x(t)\right). \tag{13}$$

Show that (13) implies

$$\overline{e}(t+1) = \left(1 - \beta(t+1)\right)\overline{e}(t) + \alpha(t+1)\overline{v}(t+1).$$

(b) Let $\mathcal{F}(t) = \{x(0), \ldots, x(t), v(0), \ldots, v(t)\}$. Show that

$$\mathbf{E}\left(\overline{e}(t+1)^2 \,|\, \mathcal{F}(t)\right) = \left(1 - \beta(t+1)\right)^2 \overline{e}(t)^2 + \alpha(t+1)^2 \sigma_{\overline{v}}^2 \tag{14}$$

where $\sigma_{\overline{v}}^2$ is the variance of $\overline{v}(t)$. (Note that the variance of $\overline{v}(t)$ is constant over time because $v(1), v(2), v(3), \ldots$ are identically distributed.)

(c) Show that (14) implies

$$\mathbf{E}\left(\overline{e}(t+1)^2 \,|\, \mathcal{F}(t)\right) \le \overline{e}(t)^2 - \beta(t+1)\overline{e}(t)^2 + \alpha(t+1)^2\sigma_{\overline{v}}^2. \tag{15}$$

(d) Show that the Robbins-Siegmund's lemma is applicable to (15) and conclude that: (i) the sequence $\left(\overline{e}(t)^2\right)_{t\ge 0}$ converges, and (ii) $\sum_{t\ge 0}\beta(t+1)\overline{e}(t)^2 < \infty$.

(e) Show that (i) and (ii) above imply $\overline{e}(t) \to 0$.

(f) We now analyze the error component $\widehat{e}(t)$. Show that (13) implies

$$\widehat{e}(t+1) = \left(I - \Gamma(t+1)\right)\widehat{e}(t) + \alpha(t+1)\widehat{v}(t+1) + \beta(t+1)\widehat{\theta} \tag{16}$$

where

$$\Gamma(t+1) := \begin{bmatrix} \gamma_1(t+1) & & \\ & \ddots & \\ & & \gamma_{n-1}(t+1) \end{bmatrix}$$

and $\gamma_i(t) = (1-\lambda_i)\alpha(t) + \beta(t)$.

(g) Looking at the coordinate i of equation (16), we have

$$\widehat{e}_i(t+1) = (1-\gamma_i(t+1))\widehat{e}_i(t) + \alpha(t+1)\widehat{v}_i(t+1) + \beta(t+1)\widehat{\theta}_i.$$

Show that

$$\begin{aligned}&\mathbf{E}\left(\widehat{e}_i(t+1)^2 \,|\, \mathcal{F}(t)\right) = \\ &\quad \left(1-\gamma_i(t+1)\right)^2\widehat{e}_i(t)^2 + \alpha(t+1)^2\sigma_i^2 + \beta(t+1)^2\widehat{\theta}_i^2 + 2(1-\gamma_i(t+1))\beta(t+1)\widehat{e}_i(t)\widehat{\theta}_i\end{aligned} \tag{17}$$

where σ_i^2 is the variance of $\widehat{v}_i(t)$.

(h) Show that (17) implies

$$\begin{aligned}&\mathbf{E}\left(\widehat{e}_i(t+1)^2 \,|\, \mathcal{F}(t)\right) \le \\ &\quad \left(1-\gamma_i(t+1)\right)\widehat{e}_i(t)^2 + \alpha(t+1)^2\sigma_i^2 + \beta(t+1)^2\widehat{\theta}_i^2 + 2\beta(t+1)|\widehat{e}_i(t)||\widehat{\theta}_i|.\end{aligned} \tag{18}$$

for t sufficiently large. That is, show that there exists a T such that $t \ge T$ implies (18).

(i) Choose q in the interval $\left(\frac{a}{2}, 0.5\right)$ (the interval is non-empty due to (12)) and let $p = 1-q$. Note that $p > 0.5$ and

$$\beta(t) = \frac{1}{t^p}\frac{1}{t^q}.$$

Show that (18) implies

$$\begin{aligned}&\mathbf{E}\left(\widehat{e}_i(t+1)^2 \,|\, \mathcal{F}(t)\right) \le \\ &\quad \left(1-\gamma_i(t+1)\right)\widehat{e}_i(t)^2 + \alpha(t+1)^2\sigma_i^2 + \beta(t+1)^2\widehat{\theta}_i^2 + \frac{1}{(t+1)^{2q}}\widehat{e}_i(t)^2 + \frac{1}{(t+1)^{2p}}\widehat{\theta}_i^2,\end{aligned} \tag{19}$$

for $t \ge T$.

(j) Show that (19) implies $\widehat{e}_i(t) \to 0$.