Decentralized Kalman Filtering

Aditya Mahajan
McGill University

Joint work with Mohammad Afshari

Fields Workshop on Stochastic Processes and their applications
10 August 2017

One-shot decentralized estimation

Model

State of the world : $x \sim \mathcal{N}(0, \text{var}(x))$

Observation of agent i: $y^i = C^i x + w^i_t, \quad w^i \sim \mathcal{N}(0, \text{var}(w^i))$

Estimate of agent i : $\hat{x}^i = g^i(y^i)$. Let $\hat{x} = \text{vec}(\hat{x}^1, \dots, \hat{x}^n)$

Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, \hat{x})]$ where $\dots$

One-shot decentralized estimation

Model

State of the world : $x \sim \mathcal{N}(0, \mathrm{var}(x))$

Observation of agent i: $y^i = C^i x + w^i_t, \quad w^i \sim \mathcal{N}(0, \mathrm{var}(w^i))$

Estimate of agent i : $\hat{x}^i = g^i(y^i)$. Let $\hat{x} = \mathrm{vec}(\hat{x}^1, \dots, \hat{x}^n)$

Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, \hat{x})]$ where ...

$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + q(\hat{x}^1 - \hat{x}^2)^2$$

One-shot decentralized estimation

Model

State of the world : $x \sim \mathcal{N}(0, \mathrm{var}(x))$

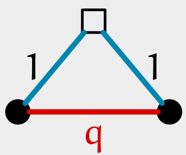
Observation of agent i: $y^i = C^i x + w^i_t, \quad w^i \sim \mathcal{N}(0, \mathrm{var}(w^i))$

Estimate of agent i : $\hat{x}^i = g^i(y^i)$. Let $\hat{x} = \mathrm{vec}(\hat{x}^1, \dots, \hat{x}^n)$

Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, \hat{x})]$ where ...

$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + q(\hat{x}^1 - \hat{x}^2)^2$$



One-shot decentralized estimation

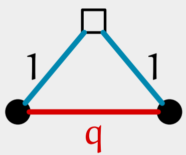
Model

- State of the world : $x \sim \mathcal{N}(0, \mathrm{var}(x))$
- Observation of agent i: $y^i = C^i x + w^i_t, \quad w^i \sim \mathcal{N}(0, \mathrm{var}(w^i))$
- Estimate of agent i : $\hat{x}^i = g^i(y^i)$. Let $\hat{x} = \mathrm{vec}(\hat{x}^1, \dots, \hat{x}^n)$

Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, \hat{x})]$ where ...

$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + q(\hat{x}^1 - \hat{x}^2)^2$$



$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + (x - \hat{x}^3)^2 + (x - \hat{x}^4)^2 + q(\hat{x}^1 - \hat{x}^2)^2 + q(\hat{x}^2 - \hat{x}^3)^2 + q(\hat{x}^3 - \hat{x}^4)^4 + q(\hat{x}^4 - \hat{x}^1)$$

One-shot decentralized estimation

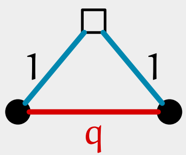
Model

- State of the world : $x \sim \mathcal{N}(0, \mathrm{var}(x))$
- Observation of agent i: $y^i = C^i x + w^i_t, \quad w^i \sim \mathcal{N}(0, \mathrm{var}(w^i))$
- Estimate of agent i : $\hat{x}^i = g^i(y^i)$. Let $\hat{x} = \mathrm{vec}(\hat{x}^1, \dots, \hat{x}^n)$

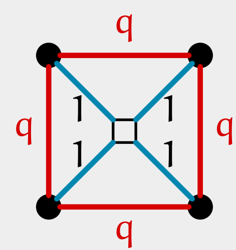
Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, \hat{x})]$ where ...

$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + q(\hat{x}^1 - \hat{x}^2)^2$$



$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + (x - \hat{x}^3)^2 + (x - \hat{x}^4)^2 + q(\hat{x}^1 - \hat{x}^2)^2 + q(\hat{x}^2 - \hat{x}^3)^2 + q(\hat{x}^3 - \hat{x}^4)^4 + q(\hat{x}^4 - \hat{x}^1)$$



One-shot decentralized estimation

Model

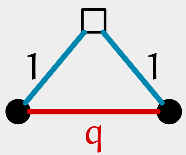
- State of the world : $x \sim \mathcal{N}(0, \mathrm{var}(x))$
- Observation of agent i: $y^i = C^i x + w^i_t, \quad w^i \sim \mathcal{N}(0, \mathrm{var}(w^i))$
- Estimate of agent i : $\hat{x}^i = g^i(y^i)$. Let $\hat{x} = \mathrm{vec}(\hat{x}^1, \dots, \hat{x}^n)$

Objective

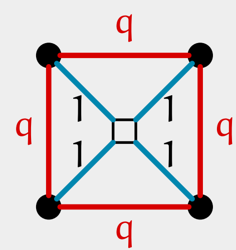
Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, \hat{x})]$ where . . .

$$c(x, \hat{x}) = \sum_{i=1}^{n} (x - \hat{x}^i)^\top M^{ii} (x - \hat{x}^i) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\hat{x}^i - \hat{x}^j)^\top M^{ij} (\hat{x}^i - \hat{x}^j)$$

$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + q(\hat{x}^1 - \hat{x}^2)^2$$



$$(x - \hat{x}^1)^2 + (x - \hat{x}^2)^2 + (x - \hat{x}^3)^2 + (x - \hat{x}^4)^2 + q(\hat{x}^1 - \hat{x}^2)^2 + q(\hat{x}^2 - \hat{x}^3)^2 + q(\hat{x}^3 - \hat{x}^4)^4 + q(\hat{x}^4 - \hat{x}^1)$$



Multi-step decentralized estimation (basic version)

Model

State of the world : $x_{t+1} = Ax_t + w^0_t, \quad w^0_t \sim \mathcal{N}(0, \text{var}(w^0))$

Observation of agent i: $y^i_t = C^i x_t + w^i_t, \quad w^i_t \sim \mathcal{N}(0, \text{var}(w^i))$

Estimate of agent i : $\hat{x}^i_t = g^i(y^i_{1:t})$. Let $\hat{x}_t = \text{vec}(\hat{x}^1_t, \dots, \hat{x}^n_t)$

Multi-step decentralized estimation (basic version)

Model

State of the world : $x_{t+1} = Ax_t + w^0_t, \quad w^0_t \sim \mathcal{N}(0, \mathrm{var}(w^0))$

Observation of agent i: $y^i_t = C^i x_t + w^i_t, \quad w^i_t \sim \mathcal{N}(0, \mathrm{var}(w^i))$

Estimate of agent i : $\hat{x}^i_t = g^i(y^i_{1:t})$. Let $\hat{x}_t = \mathrm{vec}(\hat{x}^1_t, \dots, \hat{x}^n_t)$

Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}\left[\sum_{t=1}^{T} c(x_t, \hat{x}_t)\right]$ where

$$c(x_t, \hat{x}_t) = \sum_{i=1}^{n} (x_t - \hat{x}^i_t)^\top M^{ii} (x_t - \hat{x}^i_t) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\hat{x}^i_t - \hat{x}^j_t)^\top M^{ij} (\hat{x}^i_t - \hat{x}^j_t)$$

Multi-step decentralized estimation (basic version)

Model

State of the world : $x_{t+1} = Ax_t + w_t^0, \quad w_t^0 \sim \mathcal{N}(0, \text{var}(w^0))$

Observation of agent i: $y_t^i = C^i x_t + w_t^i, \quad w_t^i \sim \mathcal{N}(0, \text{var}(w^i))$

Estimate of agent i : $\hat{x}_t^i = g^i(y_{1:t}^i)$. Let $\hat{x}_t = \text{vec}(\hat{x}_t^1, \dots, \hat{x}_t^n)$

Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}\left[\sum_{t=1}^{T} c(x_t, \hat{x}_t)\right]$ where

$$c(x_t, \hat{x}_t) = \sum_{i=1}^{n} (x_t - \hat{x}_t^i)^\top M^{ii} (x_t - \hat{x}_t^i) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\hat{x}_t^i - \hat{x}_t^j)^\top M^{ij} (\hat{x}_t^i - \hat{x}_t^j)$$

General version

Neighbors can communicate to one another over a communication graph.

$\hat{x}_t^i = g^i(I_t^i)$, where $I_1^i = y_1^i$ and for $t > 1$, $I_t^i = \text{vec}\left(y_t^i, I_{t-1}^i, \{I_{t-1}^j\}_{j \in N^i}\right)$.

Motivation

The model is interesting
and it ought to be useful!

Previous work on decentralized Kalman filtering

A very similar model was considered in [Barta 1978] and [Andersland and Teneketzis 1996].

- Barta, "On linear control of decentralized stochastic systems," PhD Thesis, MIT 1978.
- Andersland and Teneketzis, "Measurement Scheduling for Recursive Team Estimation," JOTA 1996.

Previous work on decentralized Kalman filtering

A very similar model was considered in [Barta 1978] and [Andersland and Teneketzis 1996].

Model Same as the basic multi-step model (i.e., no inter-agent communication).

Objective Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}\left[\sum_{t=1}^{T} c(x_t, \hat{x}_t)\right]$ where

$$c(x_t, \hat{x}_t) = \begin{bmatrix} x_t - \hat{x}_t^1 \\ \vdots \\ x_t - \hat{x}_t^n \end{bmatrix}^\top Q \begin{bmatrix} x_t - \hat{x}_t^1 \\ \vdots \\ x_t - \hat{x}_t^n \end{bmatrix}.$$

▷ Barta, "On linear control of decentralized stochastic systems," PhD Thesis, MIT 1978.

▷ Andersland and Teneketzis, "Measurement Scheduling for Recursive Team Estimation," JOTA 1996.

Barta's (or rather Andersland and Teneketzis's) change of variables

State model

Suppose $x \in \mathbb{R}^m$. Let $\mathbb{I} = \underbrace{\begin{bmatrix} I_m & & \\ & \ddots & \\ & & I_m \end{bmatrix}}_{n-\text{times}}$.

Define $X_t = \mathbb{I} * \underbrace{\begin{bmatrix} x_t & & \\ & \ddots & \\ & & x_t \end{bmatrix}}_{n-\text{times}}$, $\mathcal{A} = \mathbb{I} * \underbrace{\begin{bmatrix} A & & \\ & \ddots & \\ & & A \end{bmatrix}}_{n-\text{times}}$, $W^0_t = \mathbb{I} * \underbrace{\begin{bmatrix} w^0_t & & \\ & \ddots & \\ & & w^0_t \end{bmatrix}}_{n-\text{times}}$.

X_t and W^0_t are $nm^2 \times nm$ matrices. $\mathcal{A}$ is $nm^2 \times nm^2$

$$X_{t+1} = \mathcal{A} X_t + W^0_t$$

Barta's (or rather Andersland and Teneketzis's) change of variables

Observation model

$$Y_t = \mathbb{I} * \begin{bmatrix} y_t^1 & & \\ & \ddots & \\ & & y_t^n \end{bmatrix}, \mathcal{C} = \mathbb{I} * \begin{bmatrix} C^1 & & \\ & \ddots & \\ & & C^n \end{bmatrix}, W_t = \mathbb{I} * \begin{bmatrix} w_t^1 & & \\ & \ddots & \\ & & w_t^n \end{bmatrix}.$$

Then,

$$Y_t = \mathcal{C} X_t + W_t$$

Why?

Why?

Hilbert space

Let $\mathcal{X}$ denote the space of $(nm^2 \times nm)$-dimensional square integrable random variables. For $X, Z \in \mathcal{X}$, define

$$\langle X, Z \rangle = \operatorname{Tr} \mathbb{E}[X Q Z^\top], \qquad \|X\|^2_{\mathcal{H}} = \langle X, X \rangle$$

Why?

Hilbert space Let $\mathcal{X}$ denote the space of $(nm^2 \times nm)$-dimensional square integrable random variables. For $X, Z \in \mathcal{X}$, define

$$\langle X, Z \rangle = \operatorname{Tr} \mathbb{E}[XQZ^\top], \qquad \|X\|^2_{\mathcal{H}} = \langle X, X \rangle$$

Key Lemma Let X^*_t denote the minimizer of

$$\min_{\hat{X}_t \in \mathcal{X}} \|X_t - \hat{X}_t\|^2_{\mathcal{H}}$$

There exists a binary matrix S such that

$$\inf_{g^1_t, \dots, g^n_t} \mathbb{E}[(x_t - \hat{x}_t)^\top Q (x_t - \hat{x}_t)] = S\, \mathbb{E}[(X_t - \hat{X}^*_t) Q (X_t - \hat{X}^*_t)^\top] S^\top$$

Moreover, $\hat{x}^*_t = S\hat{X}^*_t$ achieves the minimum of the left hand side.

Why?

Hilbert space

Let $\mathcal{X}$ denote the space of $(nm^2 \times nm)$-dimensional square integrable random variables. For $X, Z \in \mathcal{X}$, define

$$\langle X, Z \rangle = \operatorname{Tr} \mathbb{E}[XQZ^\top], \qquad \|X\|^2_{\mathcal{H}} = \langle X, X \rangle$$

Key Lemma

Let X^*_t denote the minimizer of

$$\min_{\hat{X}_t \in \mathcal{X}} \|X_t - \hat{X}_t\|^2_{\mathcal{H}}$$

There exists a binary matrix S such that

$$\inf_{g^1_t, \dots, g^n_t} \mathbb{E}[(x_t - \hat{x}_t)^\top Q (x_t - \hat{x}_t)] = S\, \mathbb{E}[(X_t - \hat{X}^*_t) Q (X_t - \hat{X}^*_t)^\top] S^\top$$

Moreover, $\hat{x}^*_t = S\hat{X}^*_t$ achieves the minimum of the left hand side.

$\hat{X}^*_t$ is given by the orthogonal projection theorem. We can write down Kalman filtering equation!

This is too complicated (for us).

Our solution is much simpler.

A very brief introduction to static teams

Static teams (simplified version of Radner's model)

Model

- Decentralized system with n agents.
- $(x, y^1, \dots, y^n)$ jointly Gaussian. $\text{cov}(x, y^i) = \Theta^i$, $\text{cov}(y^i, y^j) = \Sigma^{ij}$.
- Agent i observes y^i and chooses $u^i = g^i(y^i)$.

- Radner, "Team Decision Problems," Annals of Math. Stats, 1962.

Static teams (simplified version of Radner's model)

Model

- Decentralized system with n agents.
- $(x, y^1, \dots, y^n)$ jointly Gaussian. $\text{cov}(x, y^i) = \Theta^i$, $\text{cov}(y^i, y^j) = \Sigma^{ij}$.
- Agent i observes y^i and chooses $u^i = g^i(y^i)$.

Objective

Choose $g = (g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, u)]$ where

$$c(x, u) = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(u^i)^\top R^{ij} u^j + 2\sum_{i=1}^{n}(u^i)^\top P_i x\right]$$

- Radner, "Team Decision Problems," Annals of Math. Stats, 1962.

The idea of Radner's solution

Necessary condition for optimality

A strategy $g = (g^1, \dots, g^n)$ is optimal only if for any other strategy $\tilde{g} = (\tilde{g}^1, \dots, \tilde{g}^n)$

$$J(\tilde{g}^i, g^{-i}) - J(g) \geqslant 0$$

This also implies that the strategy g is person by person optimal.

Sufficient condition for optimality

A strategy $g = (g^1, \dots, g^n)$ is optimal if for any other strategy $\tilde{g} = (\tilde{g}^1, \dots, \tilde{g}^n)$

$$J(\tilde{g}) - J(g) \geqslant 0$$

The idea of Radner's solution

Necessary condition for optimality

A strategy $g = (g^1, \ldots, g^n)$ is optimal only if for any other strategy $\tilde{g} = (\tilde{g}^1, \ldots, \tilde{g}^n)$

$$J(\tilde{g}^i, g^{-i}) - J(g) \geqslant 0$$

This also implies that the strategy g is person by person optimal.

Sufficient condition for optimality

A strategy $g = (g^1, \ldots, g^n)$ is optimal if for any other strategy $\tilde{g} = (\tilde{g}^1, \ldots, \tilde{g}^n)$

$$J(\tilde{g}) - J(g) \geqslant 0$$

Radner's key result was to show that PBPO implies team optimality.

The idea of Radner's solution

Necessary condition for optimality

A strategy $g = (g^1, \dots, g^n)$ is optimal only if for any other strategy $\tilde{g} = (\tilde{g}^1, \dots, \tilde{g}^n)$

$$J(\tilde{g}^i, g^{-i}) - J(g) \geqslant 0$$

This also implies that the strategy g is person by person optimal.

Sufficient condition for optimality

A strategy $g = (g^1, \dots, g^n)$ is optimal if for any other strategy $\tilde{g} = (\tilde{g}^1, \dots, \tilde{g}^n)$

$$J(\tilde{g}) - J(g) \geqslant 0$$

> Radner's key result was to show that PBPO implies team optimality.

Necessary and sufficient condition

$$g^i(y^i) = u_i \text{ such that } \frac{\partial}{\partial u^i} \mathbb{E}[c(x, g^{-i}(y^{-i}), u^i)) \,|\, y^i] = 0$$

Radner's solution (cont.)

Main result

Optimal control law is linear and is given by

$$u^i = F^i(y^i - \mathbb{E}[y^i]) + H^i\,\mathbb{E}[x],$$

$$F = -\Gamma^{-1}\eta, \qquad H = -R^{-1}P,$$

where

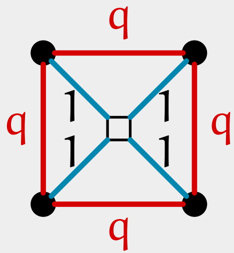
- $F = \text{vec}(F^1, F^2, \cdots, F^n)$
- $H = \text{rows}(H^1, H^2, \cdots, H^n)$.
- $\Gamma = [\Gamma^{ij}]$, where $\Gamma^{ij} = \Sigma^{ij} \otimes R^{ij}$. We can write $\Gamma = \Sigma * R$ (Khatri Rao product)
- $\eta = \text{vec}(P^1\Theta^1, P^2\Theta^2, \ldots, P^n\Theta^n)$.

- Khatri and Rao, "Solutions to some functional equations and their applications to characterization of probability distributions", Sankhya, 1968.

Key idea

The one-shot decentralized estimation problem is a static team

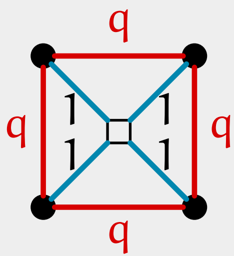
One-step decentralized estimation as a static team



In the decentralized estimation problem, we have

$$c(x, \hat{x}) = \sum_{i=1}^{n} (x - \hat{x}^i)^\top M^{ii} (x - \hat{x}^i) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\hat{x}^i - \hat{x}^j)^\top M^{ij} (\hat{x}^i - \hat{x}^j)$$

One-step decentralized estimation as a static team



In the decentralized estimation problem, we have

$$c(x,\hat{x}) = \sum_{i=1}^{n} (x-\hat{x}^i)^\top M^{ii}(x-\hat{x}^i) + \sum_{i=1}^{n}\sum_{j=i+1}^{n} (\hat{x}^i-\hat{x}^j)^\top M^{ij}(\hat{x}^i-\hat{x}^j)$$

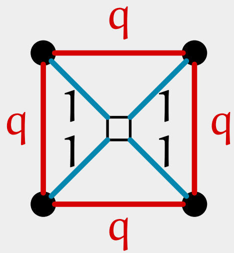
This can be written as $x^\top Q x + \hat{x}^\top R\hat{x} + 2\hat{x}^\top P x$, where

- $Q = \sum_{i=1}^{n} M^{ii}$,
- $P = \text{rows}(-M^{ii},\dots,-M^{nn})$
- $R = [R^{ij}]$, where

$$R^{ij} = \begin{cases} M^{ii} + \sum_{j\in N_j} M^{ij}, & \text{if } i=j \\ -M^{ij}, & \text{if } j\in N_i \\ 0, & \text{otherwise} \end{cases}$$

- $\Sigma^{ii} = C^i\Sigma_x(C^i)^\top + \text{var}(w^i)$
- $\Sigma^{ij} = C^i\Sigma_x(C^j)^\top$
- $\Theta^i = \Sigma_x(C^i)^\top$.

One-step decentralized estimation as a static team



In the decentralized estimation problem, we have

$$c(x, \hat{x}) = \sum_{i=1}^{n} (x - \hat{x}^i)^\top M^{ii} (x - \hat{x}^i) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\hat{x}^i - \hat{x}^j)^\top M^{ij} (\hat{x}^i - \hat{x}^j)$$

This can be written as $x^\top Q x + \hat{x}^\top R \hat{x} + 2\hat{x}^\top P x$, where

- $Q = \sum_{i=1}^{n} M^{ii}$,
- $P = \text{rows}(-M^{ii}, \dots, -M^{nn})$
- $R = [R^{ij}]$, where

$$R^{ij} = \begin{cases} M^{ii} + \sum_{j \in N_j} M^{ij}, & \text{if } i = j \\ -M^{ij}, & \text{if } j \in N_i \\ 0, & \text{otherwise} \end{cases}$$

- $\Sigma^{ii} = C^i \Sigma_x (C^i)^\top + \text{var}(w^i)$
- $\Sigma^{ij} = C^i \Sigma_x (C^j)^\top$
- $\Theta^i = \Sigma_x (C^i)^\top$.

Relation to graphs

If we think of M^{ij} as weights of a cost graph, then R is the graph Laplacian.

Optimal solution for one-shot decentralized estimation

Translating Radner's result

Since the model is a static team, from Radner's result we can say that the optimal estimates are

$$\hat{x}^i = F^i y^i$$

However, this form of the solution does not work well for the multi-step case.

Optimal solution for one-shot decentralized estimation

Translating Radner's result

Since the model is a static team, from Radner's result we can say that the optimal estimates are

$$\hat{x}^i = F^i y^i$$

However, this form of the solution does not work well for the multi-step case.

An alternative form of the solution

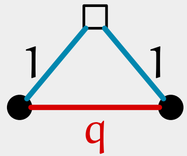
Let $\hat{x}^i_{\text{local}} = \mathbb{E}[x \,|\, y^i]$. Then, the optimal estimates are given by

$$\hat{x}^i = L^i \, \hat{x}^i_{\text{local}}, \qquad L = -\Gamma^{-1}\eta$$

where

- $L = \text{vec}(L^1, \dots, L^n)$
- $\hat{\Sigma}^{ij} = \text{cov}(\hat{x}^i, \hat{x}^j) = \Theta^i (\Sigma^{ii})^{-1} \Sigma^{ij} (\Sigma^{jj})^{-1} (\Theta^j)^\top$
- $\Gamma = [\Gamma^{ij}]$, where $\Gamma^{ij} = \hat{\Sigma}^{ij} \otimes R^{ij}$
- $\eta = \text{vec}(P^1 \hat{\Sigma}^{11}, \dots, P^n \hat{\Sigma}^{nn})$

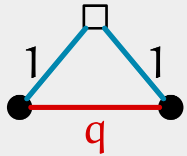
Examples of one-shot estimation



Suppose $x \sim \mathcal{N}(0, \sigma_0^2)$ and $y^i = x + w^i$ where $w^i \sim \mathcal{N}(0, \sigma^2)$. Then,

$$\Gamma = \begin{bmatrix} 1+q & -\alpha q \\ -\alpha q & 1+q \end{bmatrix}, \quad \text{where } \alpha = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}.$$

Examples of one-shot estimation

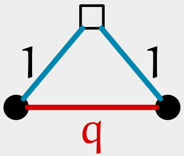


Suppose $x \sim \mathcal{N}(0, \sigma_0^2)$ and $y^i = x + w^i$ where $w^i \sim \mathcal{N}(0, \sigma^2)$. Then,

$$\Gamma = \begin{bmatrix} 1+q & -\alpha q \\ -\alpha q & 1+q \end{bmatrix}, \quad \text{where } \alpha = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}.$$

$$\Gamma^{-1} = \frac{1}{(1+q)^2 - (\alpha q)^2} \begin{bmatrix} 1+q & \alpha q \\ \alpha q & 1+q \end{bmatrix}.$$

Examples of one-shot estimation



Suppose $x \sim \mathcal{N}(0, \sigma_0^2)$ and $y^i = x + w^i$ where $w^i \sim \mathcal{N}(0, \sigma^2)$. Then,

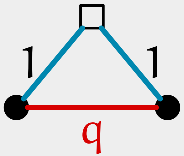
$$\Gamma = \begin{bmatrix} 1+q & -\alpha q \\ -\alpha q & 1+q \end{bmatrix}, \quad \text{where } \alpha = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}.$$

$$\Gamma^{-1} = \frac{1}{(1+q)^2 - (\alpha q)^2} \begin{bmatrix} 1+q & \alpha q \\ \alpha q & 1+q \end{bmatrix}.$$

Thus, $L = -\Gamma^{-1}\eta = -\dfrac{1}{(1+q)^2 - (\alpha q)^2} \begin{bmatrix} 1+q & \alpha q \\ \alpha q & 1+q \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$$= \frac{1}{1 + (1-\alpha)q} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Examples of one-shot estimation



Suppose $x \sim \mathcal{N}(0, \sigma_0^2)$ and $y^i = x + w^i$ where $w^i \sim \mathcal{N}(0, \sigma^2)$. Then,

$$\Gamma = \begin{bmatrix} 1+q & -\alpha q \\ -\alpha q & 1+q \end{bmatrix}, \quad \text{where } \alpha = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}.$$

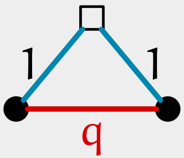
$$\Gamma^{-1} = \frac{1}{(1+q)^2 - (\alpha q)^2} \begin{bmatrix} 1+q & \alpha q \\ \alpha q & 1+q \end{bmatrix}.$$

$$\text{Thus, } L = -\Gamma^{-1}\eta = -\frac{1}{(1+q)^2 - (\alpha q)^2} \begin{bmatrix} 1+q & \alpha q \\ \alpha q & 1+q \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$= \frac{1}{1 + (1-\alpha)q} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

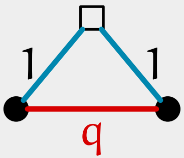
Hence, $\hat{x}^i = \dfrac{1}{1 + \bar{\alpha} q} \hat{x}^i_{\text{local}}$, where $\bar{\alpha} = \dfrac{\sigma^2}{\sigma_0^2 + \sigma^2}$. (Recall, $\hat{x}^i_{\text{local}} = \alpha y^i$.)

Examples of one-shot estimation

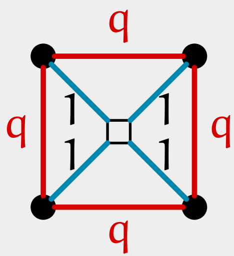


$$\hat{x}^i = \frac{1}{1 + \bar{\alpha} q} \hat{x}^i_{\text{local}}, \quad \text{where } \bar{\alpha} = \frac{\sigma^2}{\sigma_0^2 + \sigma^2}.$$

Examples of one-shot estimation

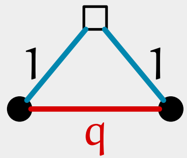


$$\hat{x}^i = \frac{1}{1 + \bar{\alpha} q} \hat{x}^i_{\text{local}}, \quad \text{where } \bar{\alpha} = \frac{\sigma^2}{\sigma_0^2 + \sigma^2}.$$

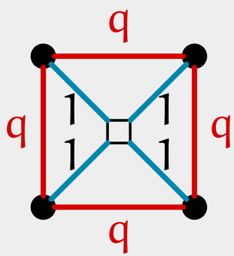


$$\hat{x}^i = \frac{1}{1 + 2\bar{\alpha} q} \hat{x}^i_{\text{local}}.$$

Examples of one-shot estimation



$$\hat{x}^i = \frac{1}{1 + \bar{\alpha} q} \hat{x}^i_{\text{local}}, \quad \text{where } \bar{\alpha} = \frac{\sigma^2}{\sigma_0^2 + \sigma^2}.$$



$$\hat{x}^i = \frac{1}{1 + 2\bar{\alpha} q} \hat{x}^i_{\text{local}}.$$

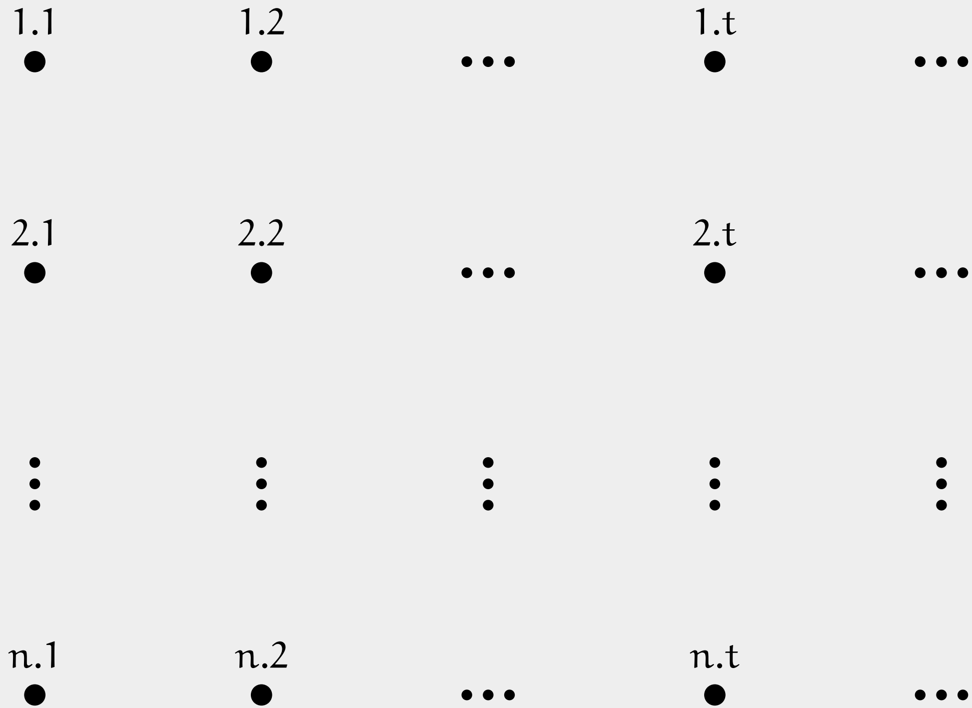
d-regular graph

$$\hat{x}^i = \frac{1}{1 + d\bar{\alpha} q} \hat{x}^i_{\text{local}}.$$

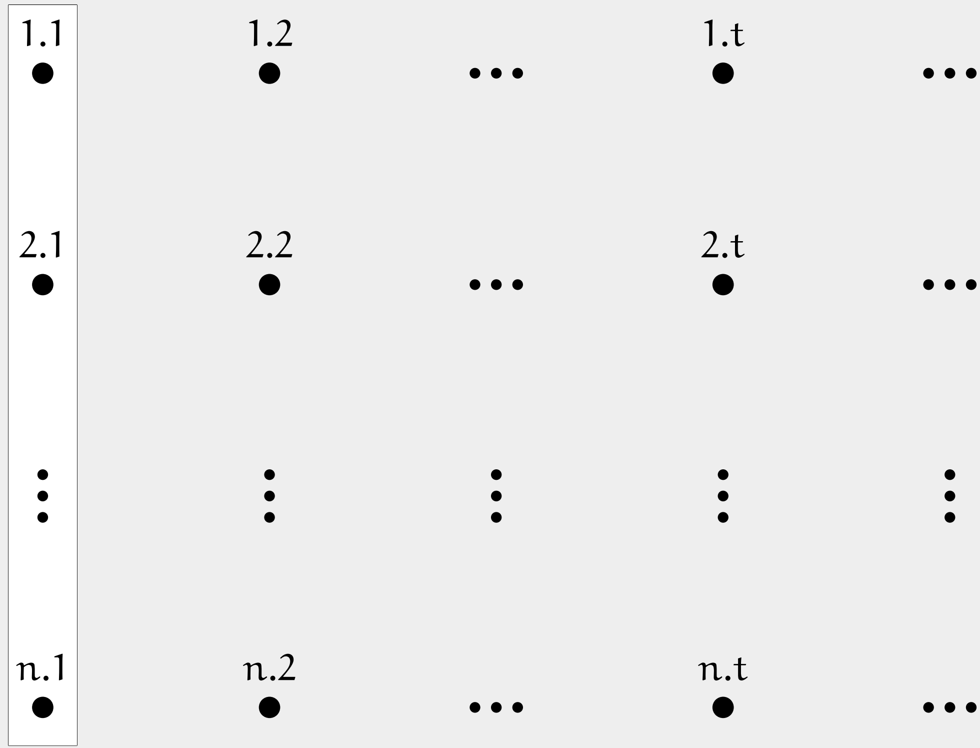
Proof: Show that $\Gamma L = -\eta$

Multi-step decentralized estimation

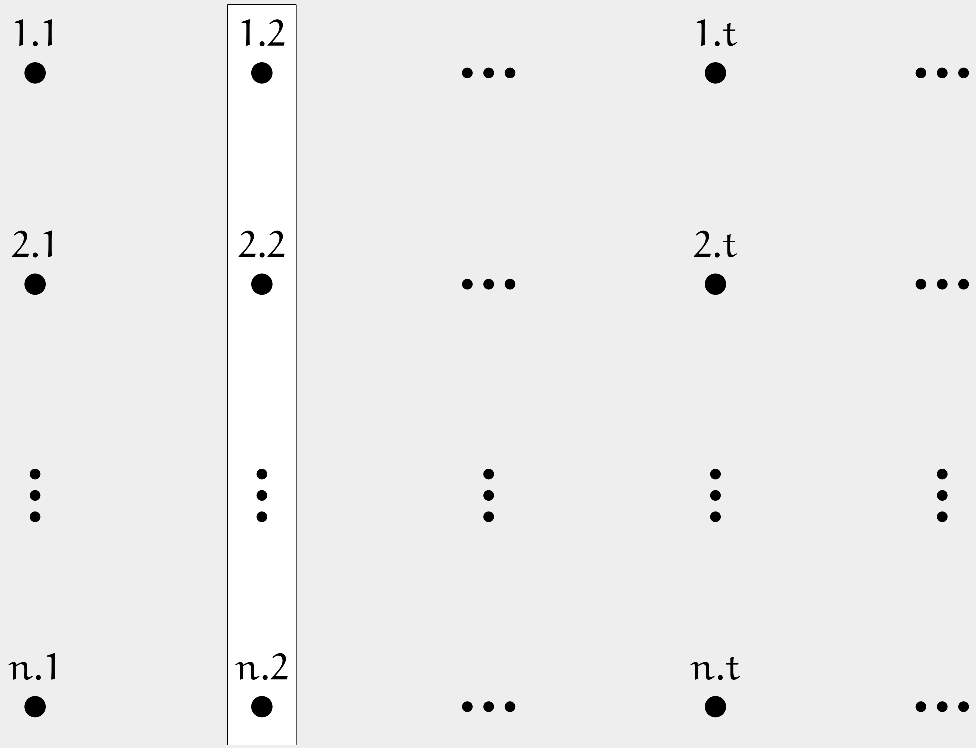
Multi-step decentralized estimation



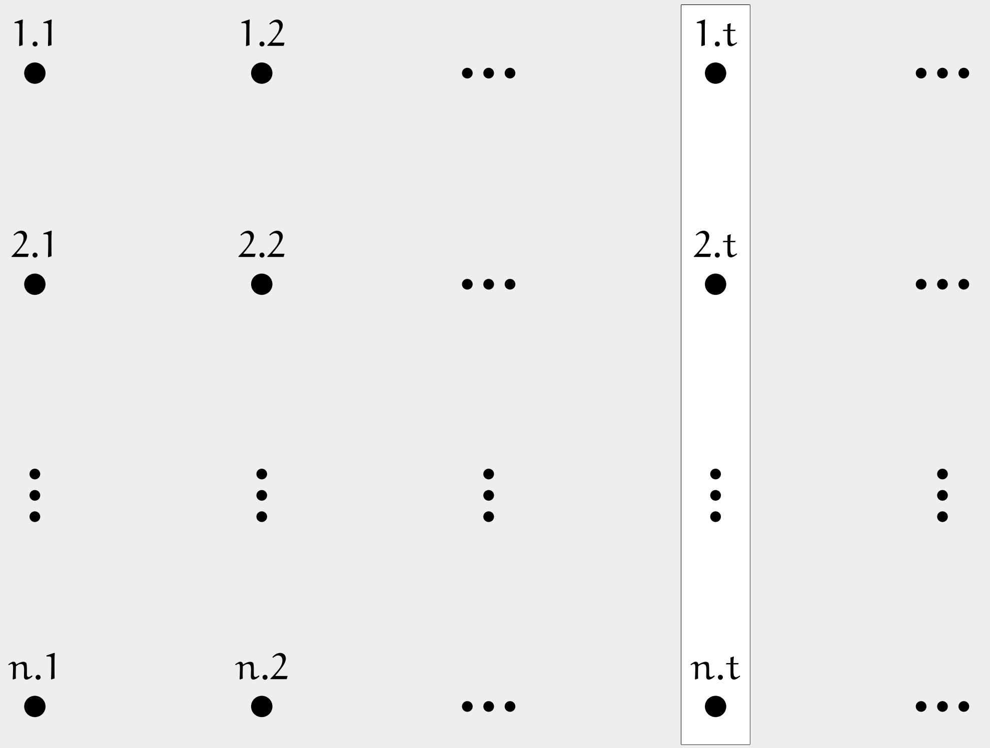
Multi-step decentralized estimation



Multi-step decentralized estimation



Multi-step decentralized estimation



Multi-step decentralized estimation

1.1 1.2 … 1.t …

2.

⋮

Instead of solving $\min \mathbb{E}\left[\sum_{t=1}^{T} c(x_t, \hat{x}_t)\right]$

we can solve $\min \mathbb{E}[c(x_t, \hat{x}_t)]$ for each t.

n.1 n.2 … n.t …

Multi-step decentralized estimation

Key observation **The problem at time t is a one-shot optimization problem**

Multi-step decentralized estimation

Key observation The problem at time t is a one-shot optimization problem

Optimal estimator Let $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid I^i_t]$ and $\hat{\Sigma}^{ij}_t = \text{cov}(\hat{x}^i_{\text{local},t}, \hat{x}^j_{\text{local},t})$. Then,

$$\hat{x}^i_t = L^i_t \, \hat{x}^i_{\text{local},t}, \qquad \text{vec}(L^i_t) = -\left[\hat{\Sigma}^{ij}_t \otimes R^{ij}\right]^{-1} \text{vec}(P^i \hat{\Sigma}^{ii}_t)$$

Multi-step decentralized estimation

Key observation The problem at time t is a one-shot optimization problem

Optimal estimator Let $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid I^i_t]$ and $\hat{\Sigma}^{ij}_t = \text{cov}(\hat{x}^i_{\text{local},t}, \hat{x}^j_{\text{local},t})$. Then,

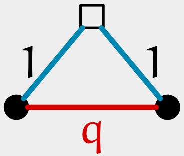
$$\hat{x}^i_t = L^i_t \, \hat{x}^i_{\text{local},t}, \qquad \text{vec}(L^i_t) = -\left[\hat{\Sigma}^{ij}_t \otimes R^{ij}\right]^{-1} \text{vec}(P^i \hat{\Sigma}^{ii}_t)$$

Remarks To compute the optimal solution, we only need to compute $\hat{x}^i_{\text{local},t}$ and $\hat{\Sigma}^{ij}_t$.

Recall, all random variables are jointly Gaussian. Pre-computing $\hat{\Sigma}^{ij}_t$ and keeping track of $\hat{x}^i_{\text{local},t}$ is trivial but for computational complexity.

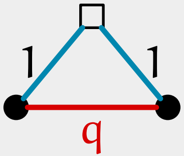
Almost same as standard Kalman filtering! Relatively straight forward to come up with recursive equations (but for notation!).

Example of multi-step estimation



$$\Gamma_t = \begin{bmatrix} (1+q)\hat{\Sigma}_t^{11} & -q\hat{\Sigma}_t^{12} \\ -q\hat{\Sigma}_t^{21} & (1+q)\hat{\Sigma}_t^{22} \end{bmatrix}. \qquad \eta_t = \begin{bmatrix} -\hat{\Sigma}_t^{11} \\ -\hat{\Sigma}_t^{22} \end{bmatrix}$$

Example of multi-step estimation



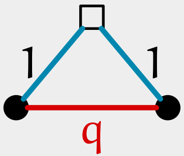
$$\Gamma_t = \begin{bmatrix} (1+q)\hat{\Sigma}_t^{11} & -q\hat{\Sigma}_t^{12} \\ -q\hat{\Sigma}_t^{21} & (1+q)\hat{\Sigma}_t^{22} \end{bmatrix}. \qquad \eta_t = \begin{bmatrix} -\hat{\Sigma}_t^{11} \\ -\hat{\Sigma}_t^{22} \end{bmatrix}$$

Assume a symmetric communication channel. So, $\hat{\Sigma}_t^{11} = \hat{\Sigma}_t^{22}$ and $\hat{\Sigma}_t^{12} = \hat{\Sigma}_t^{21}$.

Then $L = -\Gamma_t^{-1}\eta_t = \begin{bmatrix} (1+q) & -\alpha_t q \\ -\alpha_t q & (1+q) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \dfrac{1}{1+\bar{\alpha}_t q} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\alpha_t = \hat{\Sigma}_t^{12} / \hat{\Sigma}_t^{11}$ and $\bar{\alpha}_t = 1 - \alpha_t$.

Example of multi-step estimation



$$\Gamma_t = \begin{bmatrix} (1+q)\hat{\Sigma}_t^{11} & -q\hat{\Sigma}_t^{12} \\ -q\hat{\Sigma}_t^{21} & (1+q)\hat{\Sigma}_t^{22} \end{bmatrix}. \qquad \eta_t = \begin{bmatrix} -\hat{\Sigma}_t^{11} \\ -\hat{\Sigma}_t^{22} \end{bmatrix}$$

Assume a symmetric communication channel. So, $\hat{\Sigma}_t^{11} = \hat{\Sigma}_t^{22}$ and $\hat{\Sigma}_t^{12} = \hat{\Sigma}_t^{21}$.

Then $L = -\Gamma_t^{-1}\eta_t = \begin{bmatrix} (1+q) & -\alpha_t q \\ -\alpha_t q & (1+q) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \dfrac{1}{1+\bar{\alpha}_t q} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\alpha_t = \hat{\Sigma}_t^{12} / \hat{\Sigma}_t^{11}$ and $\bar{\alpha}_t = 1 - \alpha_t$.

d-regular graph

Suppose the communication graph is such that $\hat{\Sigma}_t^{ii}$ and $\hat{\Sigma}_t^{ij}$ are symmetric.

Then, $\hat{x}^i = \dfrac{1}{1 + d\bar{\alpha}_t q} \hat{x}^i_{\text{local}}$, $\qquad \bar{\alpha}_t = 1 - \dfrac{\hat{\Sigma}_t^{ij}}{\hat{\Sigma}_t^{ii}}$

Recursive computation of $\hat{x}^{i}_{local,t}$ and $\hat{\Sigma}^{ij}_{t}$.

Proof by examples . . .

Proof by examples . . .

No sharing of information $I_t^i = \{y_{1:t}^i\}$

Proof by examples . . .

No sharing of information

$$I_t^i = \{y_{1:t}^i\}$$

One step delay sharing

Complete communication graph with one unit communication delay.

$$I_t^i = \{y_t^i, y_{1:t-1}\}$$

Proof by examples . . .

No sharing of information

$I_t^i = \{y_{1:t}^i\}$

One step delay sharing

Complete communication graph with one unit communication delay.

$$I_t^i = \{y_t^i, y_{1:t-1}\}$$

d-step delay sharing

Complete communication graph with d units communication delay.

$$I_t^i = \{y_{t-d+1:t}^i, y_{1:t-d}\}$$

Proof by examples . . .

No sharing of information

$I_t^i = \{y_{1:t}^i\}$

One step delay sharing

Complete communication graph with one unit communication delay.

$$I_t^i = \{y_t^i, y_{1:t-1}\}$$

d-step delay sharing

Complete communication graph with d units communication delay.

$$I_t^i = \{y_{t-d+1:t}^i, y_{1:t-d}\}$$

General comm. graph

Assume a completely connected (directed) communication graph.

Can be effectively viewed as a d-step delay sharing, where d is the diameter of the graph.

No sharing of information

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \,|\, y^i_{1:t}]$. Then,

$$\hat{x}^i_{\text{local},t} = A\hat{x}^i_{\text{local},t-1} + K^i_t \big[y^i_t - C^i A \hat{x}^i_{\text{local},t-1} \big]$$

No sharing of information

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_{1:t}]$. Then,

$$\hat{x}^i_{\text{local},t} = A\hat{x}^i_{\text{local},t-1} + K^i_t \big[y^i_t - C^i A \hat{x}^i_{\text{local},t-1}\big]$$

Recursion for conditional covariance

Let $\Sigma^i_{t|t} = \text{var}(x_t - \hat{x}^i_{\text{local},t})$ and $\Sigma^i_{t+1|t} = A^\top \Sigma^i_{t|t} A + \text{var}(w^0)$. Then,

$$K^i_t = \Sigma^i_{t|t-1}\,(C^i)^\top \big[C^i\,\Sigma^i_{t|t-1}\,(C^t)^\top + \text{var}(w^i)\big]^{-1}$$

Let $\Lambda^i_t = I - K^i_t\,C^i$. Then,

$$\Sigma^i_{t|t} = \Lambda^i_t\,\Sigma^i_{t|t-1}\,(\Lambda^i_t)^\top + K^i_t\ \text{var}(w^i)\,(K^i_t)^\top$$

No sharing of information

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_{1:t}]$. Then,

$$\hat{x}^i_{\text{local},t} = A\hat{x}^i_{\text{local},t-1} + K^i_t\left[y^i_t - C^i A\hat{x}^i_{\text{local},t-1}\right]$$

Recursion for conditional covariance

Let $\Sigma^i_{t|t} = \text{var}(x_t - \hat{x}^i_{\text{local},t})$ and $\Sigma^i_{t+1|t} = A^\top \Sigma^i_{t|t} A + \text{var}(w^0)$. Then,

$$K^i_t = \Sigma^i_{t|t-1}\,(C^i)^\top \left[C^i\,\Sigma^i_{t|t-1}\,(C^t)^\top + \text{var}(w^i)\right]^{-1}$$

Standard Kalman filter

Let $\Lambda^i_t = I - K^i_t\,C^i$. Then,

$$\Sigma^i_{t|t} = \Lambda^i_t\,\Sigma^i_{t|t-1}\,(\Lambda^i_t)^\top + K^i_t\,\text{var}(w^i)\,(K^i_t)^\top$$

No sharing of information

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_{1:t}]$. Then,

$$\hat{x}^i_{\text{local},t} = A\hat{x}^i_{\text{local},t-1} + K^i_t\big[y^i_t - C^i A\hat{x}^i_{\text{local},t-1}\big]$$

Recursion for conditional covariance

Let $\Sigma^i_{t|t} = \text{var}(x_t - \hat{x}^i_{\text{local},t})$ and $\Sigma^i_{t+1|t} = A^\top \Sigma^i_{t|t} A + \text{var}(w^0)$. Then,

$$K^i_t = \Sigma^i_{t|t-1}\,(C^i)^\top \big[C^i\,\Sigma^i_{t|t-1}\,(C^t)^\top + \text{var}(w^i)\big]^{-1}$$

Standard Kalman filter

Let $\Lambda^i_t = I - K^i_t\,C^i$. Then,

$$\Sigma^i_{t|t} = \Lambda^i_t\,\Sigma^i_{t|t-1}\,(\Lambda^i_t)^\top + K^i_t\,\text{var}(w^i)\,(K^i_t)^\top$$

Covariance across agents

Let $\hat{\Sigma}^{ij}_t = \text{cov}(\hat{x}^i_t, \hat{x}^j_t)$ and $\tilde{\Sigma}^{ij}_{t|t} = \text{cov}(x_t - \hat{x}^i_t, x_t - \hat{x}^j_t)$.

Then, $\hat{\Sigma}^{ij}_t = \Sigma^x_t - \Sigma^i_{t|t} - \Sigma^j_{t|t} - \tilde{\Sigma}^{ij}_{t|t}$ and

$$\tilde{\Sigma}^{ij}_{t|t} = \Lambda^i_t \tilde{\Sigma}^{ij}_{t|t-1}(\Lambda^j_t)^\top, \qquad \text{where } \tilde{\Sigma}^{ij}_{t|t-1} = A^\top \tilde{\Sigma}^{ij}_{t-1} A + \text{var}(w^0)$$

One-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_t, y_{1:t-1}]$. Define $\hat{x}_{t|t-1} = \mathbb{E}[x_t \mid y_{1:t-1}]$. Then,

$$\hat{x}^i_{\text{local},t} = \hat{x}_{t|t-1} + K^i_t \left[y^i_t - C^i_t \hat{x}_{t|t-1} \right]$$

One-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_t, y_{1:t-1}]$. Define $\hat{x}_{t|t-1} = \mathbb{E}[x_t \mid y_{1:t-1}]$. Then,

$$\hat{x}^i_{\text{local},t} = \hat{x}_{t|t-1} + K^i_t \left[y^i_t - C^i_t \hat{x}_{t|t-1} \right]$$

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + AK_t \left[y_t - C\hat{x}_{t|t-1} \right]$$

One-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_t, y_{1:t-1}]$. Define $\hat{x}_{t|t-1} = \mathbb{E}[x_t \mid y_{1:t-1}]$. Then,

$$\hat{x}^i_{\text{local},t} = \hat{x}_{t|t-1} + K^i_t \left[y^i_t - C^i_t \hat{x}_{t|t-1} \right]$$

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + AK_t \left[y_t - C\hat{x}_{t|t-1} \right]$$

Recursion for conditional covariance

Let $\Sigma_{t|t-1} = \text{var}(x_t - \hat{x}_{t|t-1})$. The gains are given by

$$K^i_t = \Sigma_{t|t-1}(C^i_t)^\top \left[C^i \Sigma_{t|t-1}(C^i)^\top + \text{var}(w^i) \right]^{-1}$$

$$K_t = \Sigma_{t|t-1} C^\top \left[C\Sigma_{t|t-1} C^\top + \text{var}(w^1, \dots, w^n) \right]$$

Define $\Lambda_t = I - K_t C$.

$$\Sigma_{t+1|t} = A\Lambda_t \Sigma_{t|t-1} \Lambda_t^\top A^\top + \text{var}(w^0) + AK_t \,\text{var}(w^1, \dots, w^n) K_t^\top A^\top$$

One-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_t, y_{1:t-1}]$. Define $\hat{x}_{t|t-1} = \mathbb{E}[x_t \mid y_{1:t-1}]$. Then,

$$\hat{x}^i_{\text{local},t} = \hat{x}_{t|t-1} + K^i_t \left[y^i_t - C^i_t \hat{x}_{t|t-1} \right]$$

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + AK_t \left[y_t - C\hat{x}_{t|t-1} \right]$$

Recursion for conditional covariance

Let $\Sigma_{t|t-1} = \text{var}(x_t - \hat{x}_{t|t-1})$. The gains are given by

$$K^i_t = \Sigma_{t|t-1} (C^i_t)^\top \left[C^i \Sigma_{t|t-1} (C^i)^\top + \text{var}(w^i) \right]^{-1}$$

$$K_t = \Sigma_{t|t-1} C^\top \left[C\Sigma_{t|t-1} C^\top + \text{var}(w^1, \dots, w^n) \right]$$

Standard Kalman filter

Define $\Lambda_t = I - K_t C$.

$$\Sigma_{t+1|t} = A\Lambda_t \Sigma_{t|t-1} \Lambda_t^\top A^\top + \text{var}(w^0) + AK_t \,\text{var}(w^1, \dots, w^n) K_t^\top A^\top$$

One-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_t, y_{1:t-1}]$. Define $\hat{x}_{t|t-1} = \mathbb{E}[x_t \mid y_{1:t-1}]$. Then,

$$\hat{x}^i_{\text{local},t} = \hat{x}_{t|t-1} + K^i_t \left[y^i_t - C^i_t \hat{x}_{t|t-1} \right]$$

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + AK_t \left[y_t - C\hat{x}_{t|t-1} \right]$$

Recursion for conditional covariance

Let $\Sigma_{t|t-1} = \text{var}(x_t - \hat{x}_{t|t-1})$. The gains are given by

$$K^i_t = \Sigma_{t|t-1}(C^i_t)^\top \left[C^i \Sigma_{t|t-1}(C^i)^\top + \text{var}(w^i) \right]^{-1}$$

$$K_t = \Sigma_{t|t-1} C^\top \left[C\Sigma_{t|t-1} C^\top + \text{var}(w^1, \dots, w^n) \right]$$

Standard Kalman filter

Define $\Lambda_t = I - K_t C$.

$$\Sigma_{t+1|t} = A\Lambda_t \Sigma_{t|t-1} \Lambda_t^\top A^\top + \text{var}(w^0) + AK_t \,\text{var}(w^1, \dots, w^n) K_t^\top A^\top$$

Covariance across agents

$$\hat{\Sigma}^{ij}_t = K^i C^i \Sigma_{t|t-1} (C^j)^\top (K^j)^\top$$

d-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_{t-d+1:t}, y_{1:t-d}]$. Define $\hat{x}_{t-d+1|t-d} = \mathbb{E}[x_{t-d+1} \mid y_{1:t-d}]$.

$$\hat{x}^i_{\text{local},t} = A^{d-1}\hat{x}_{t-d+1|t-d} + K^i_t \left\{ \begin{bmatrix} y^i_t \\ y^i_{t-1} \\ \vdots \\ y^i_{t-d+1} \end{bmatrix} - \underbrace{\begin{bmatrix} C^i_t A^{d-1} \\ C^i_t A^{d-2} \\ \vdots \\ C^i_t \end{bmatrix}}_{\bar{C}^i_t} \hat{x}_{t-d+1|t-d} \right\}$$

d-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_{t-d+1:t}, y_{1:t-d}]$. Define $\hat{x}_{t-d+1|t-d} = \mathbb{E}[x_{t-d+1} \mid y_{1:t-d}]$.

$$\hat{x}^i_{\text{local},t} = A^{d-1}\hat{x}_{t-d+1|t-d} + K^i_t \left\{ \begin{bmatrix} y^i_t \\ y^i_{t-1} \\ \vdots \\ y^i_{t-d+1} \end{bmatrix} - \underbrace{\begin{bmatrix} C^i_t A^{d-1} \\ C^i_t A^{d-2} \\ \vdots \\ C^i_t \end{bmatrix}}_{\bar{C}^i_t} \hat{x}_{t-d+1|t-d} \right\}$$

Standard Kalman filter

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + AK_t\left[y_t - C\hat{x}_{t|t-1}\right]$$

and

$$\Sigma_{t+1|t} = A\Lambda_t \Sigma_{t|t-1}\Lambda_t^\top A^\top + \text{var}(w^0) + AK_t\,\text{var}(w^1, \dots, w^n) K_t^\top A^\top$$

d-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_{t-d+1:t}, y_{1:t-d}]$. Define $\hat{x}_{t-d+1|t-d} = \mathbb{E}[x_{t-d+1} \mid y_{1:t-d}]$.

$$\hat{x}^i_{\text{local},t} = A^{d-1}\hat{x}_{t-d+1|t-d} + K^i_t \left\{ \begin{bmatrix} y^i_t \\ y^i_{t-1} \\ \vdots \\ y^i_{t-d+1} \end{bmatrix} - \underbrace{\begin{bmatrix} C^i_t A^{d-1} \\ C^i_t A^{d-2} \\ \vdots \\ C^i_t \end{bmatrix}}_{\bar{C}^i_t} \hat{x}_{t-d+1|t-d} \right\}$$

Recursion for conditional covariance

$$K^i_t = \left[A^{d-1}\Sigma_{t-d+1|t-d}(\bar{C}^i)^\top + \bar{\Sigma}^{i0}_{t-k+1}\right]\left[\bar{C}^i\Sigma_{t-d+1|t-d}(\bar{C}^i)^\top + \bar{\Sigma}^{ii}_{t-d+1}\right]^{-1}$$

where $\bar{w}^i_{t-d+1} = W_i \operatorname{vec}(w^i_t, \dots, w^i_{t-d+1})$ and $\bar{\Sigma}^{ij}_t = \operatorname{cov}(\bar{w}^i_t, \bar{w}^j_t)$.

d-step delay sharing

Recursion for local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid y^i_{t-d+1:t}, y_{1:t-d}]$. Define $\hat{x}_{t-d+1|t-d} = \mathbb{E}[x_{t-d+1} \mid y_{1:t-d}]$.

$$\hat{x}^i_{\text{local},t} = A^{d-1}\hat{x}_{t-d+1|t-d} + K^i_t \left\{ \begin{bmatrix} y^i_t \\ y^i_{t-1} \\ \vdots \\ y^i_{t-d+1} \end{bmatrix} - \underbrace{\begin{bmatrix} C^i_t A^{d-1} \\ C^i_t A^{d-2} \\ \vdots \\ C^i_t \end{bmatrix}}_{\bar{C}^i_t} \hat{x}_{t-d+1|t-d} \right\}$$

Recursion for conditional covariance

$$K^i_t = \left[A^{d-1}\Sigma_{t-d+1|t-d}(\bar{C}^i)^\top + \bar{\Sigma}^{i0}_{t-k+1}\right]\left[\bar{C}^i\Sigma_{t-d+1|t-d}(\bar{C}^i)^\top + \bar{\Sigma}^{ii}_{t-d+1}\right]^{-1}$$

where $\bar{w}^i_{t-d+1} = W_i\,\text{vec}(w^i_t, \dots, w^i_{t-d+1})$ and $\bar{\Sigma}^{ij}_t = \text{cov}(\bar{w}^i_t, \bar{w}^j_t)$.

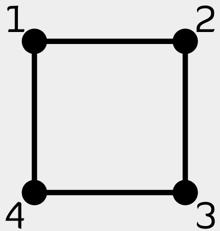
Covariance across agents

$$\hat{\Sigma}^{ij}_t = K^i_t\left[\bar{C}^i\Sigma_{t-d+1|t-d}(\bar{C}^j)^\top + \text{cov}(\bar{w}^i, \bar{w}^j)\right](K^j_t)^\top$$

General graph

Information structure

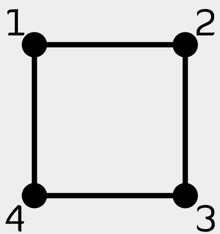
$$I_t^1 = \{y_{1:t}^1, y_{1:t-1}^2, y_{1:t-2}^3, y_{1:t-1}^4\}$$



General graph

Information structure

$$I^1_t = \{y^1_{1:t}, y^2_{1:t-1}, y^3_{1:t-2}, y^4_{1:t-1}\} = \{\underbrace{y^1_t, y^1_{t-1}, y^2_{t-1}, y^4_{t-1}}_{\text{local info}}, \underbrace{y_{1:t-2}}_{\text{common info}}\}$$

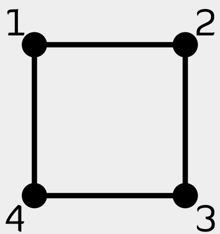


General graph

Information structure

$$I_t^1 = \{y_{1:t}^1, y_{1:t-1}^2, y_{1:t-2}^3, y_{1:t-1}^4\} = \{\underbrace{y_t^1, y_{t-1}^1, y_{t-1}^2, y_{t-1}^4}_{\text{local info}}, \underbrace{y_{1:t-2}}_{\text{common info}}\}$$

Local estimates



Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid I_t^i]$. Then,

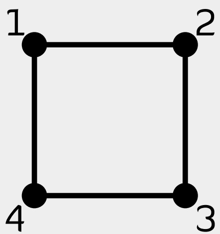
$$\hat{x}^1_{\text{local},t} = A\hat{x}_{t-1|t-2} + K_t^1 \left\{ \begin{bmatrix} y_t^1 \\ y_{t-1}^1 \\ y_{t-1}^2 \\ y_{t-1}^4 \end{bmatrix} - \begin{bmatrix} C^1 A_t \\ C^1 \\ C^2 \\ C^4 \end{bmatrix} \hat{x}_{t-1|t-2} \right\}$$

General graph

Information structure

$$I^1_t = \{y^1_{1:t}, y^2_{1:t-1}, y^3_{1:t-2}, y^4_{1:t-1}\} = \{\underbrace{y^1_t, y^1_{t-1}, y^2_{t-1}, y^4_{t-1}}_{\text{local info}}, \underbrace{y_{1:t-2}}_{\text{common info}}\}$$

Local estimates



Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid I^i_t]$. Then,

$$\hat{x}^1_{\text{local},t} = A\hat{x}_{t-1|t-2} + K^1_t \left\{ \begin{bmatrix} y^1_t \\ y^1_{t-1} \\ y^2_{t-1} \\ y^4_{t-1} \end{bmatrix} - \begin{bmatrix} C^1 A_t \\ C^1 \\ C^2 \\ C^4 \end{bmatrix} \hat{x}_{t-1|t-2} \right\}$$

Remarks

- Effectively equivalent to d-step delayed sharing.
- Each node keeps track of a delayed centralized estimator and innovation wrt common information.

Summary

One-shot decentralized estimation

Model

State of the world : $x \sim \mathcal{N}(0, \mathrm{var}(x))$

Observation of agent i: $y^i = C^i x + w^i_t, \quad w^i \sim \mathcal{N}(0, \mathrm{var}(w^i))$

Estimate of agent i : $\hat{x}^i = g^i(y^i)$. Let $\hat{x} = \mathrm{vec}(\hat{x}^1, \dots, \hat{x}^n)$

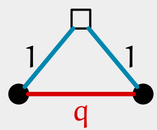
Objective

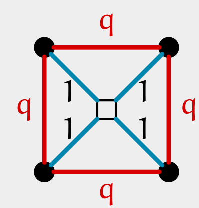
Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}[c(x, \hat{x})]$ where ...

$$c(x, \hat{x}) = \sum_{i=1}^{n} (x - \hat{x}^i)^\top M^{ii} (x - \hat{x}^i) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\hat{x}^i - \hat{x}^j)^\top M^{ij} (\hat{x}^i - \hat{x}^j)$$

$$(x-\hat{x}^1)^2 + (x-\hat{x}^2)^2 + q(\hat{x}^1-\hat{x}^2)^2$$

$$(x-\hat{x}^1)^2 + (x-\hat{x}^2)^2 + (x-\hat{x}^3)^2 + (x-\hat{x}^4)^2 + q(\hat{x}^1-\hat{x}^2)^2 + q(\hat{x}^2-\hat{x}^3)^2 + q(\hat{x}^3-\hat{x}^4)^4 + q(\hat{x}^4-\hat{x}^1)^2$$





Decentralized Kalman Filtering–(Mahajan)

Summary

One-shot decentralized estimation

Multi-step decentralized estimation (basic version)

Model

State of the world : $x_{t+1} = Ax_t + w^0_t, \quad w^0_t \sim \mathcal{N}(0, \text{var}(w^0))$

Observation of agent i: $y^i_t = C^i x_t + w^i_t, \quad w^i_t \sim \mathcal{N}(0, \text{var}(w^i))$

Estimate of agent i : $\hat{x}^i_t = g^i(y^i_{1:t})$. Let $\hat{x}_t = \text{vec}(\hat{x}^1_t, \dots, \hat{x}^n_t)$

Objective

Choose $(g^1, \dots, g^n)$ to minimize $\mathbb{E}\left[\sum_{t=1}^{T} c(x_t, \hat{x}_t)\right]$ where

$$c(x_t, \hat{x}_t) = \sum_{i=1}^{n} (x_t - \hat{x}^i_t)^\top M^{ii} (x_t - \hat{x}^i_t) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\hat{x}^i_t - \hat{x}^j_t)^\top M^{ij} (\hat{x}^i_t - \hat{x}^j_t)$$

General version

Neighbors can communicate to one another over a communication graph.

$\hat{x}^i_t = g^i(I^i_t)$, where $I^i_1 = y^i_1$ and for $t > 1$, $I^i_t = \text{vec}\left(y^i_t, I^i_{t-1}, \{I^j_{t-1}\}_{j \in N^i}\right)$.

Summary

Optimal solution for one-shot decentralized estimation

Translating Radner's result

Since the model is a static team, from Radner's result we can say that the optimal estimates are

$$\hat{x}^i = F^i y^i$$

However, this form of the solution does not work well for the multi-step case.

An alternative form of the solution

Let $\hat{x}^i_{\text{local}} = \mathbb{E}[x \,|\, y^i]$. Then, the optimal estimates are given by

$$\hat{x}^i = L^i \, \hat{x}^i_{\text{local}}, \qquad L = -\Gamma^{-1}\eta$$

where

- $L = \text{vec}(L^1, \dots, L^n)$
- $\hat{\Sigma}^{ij} = \text{cov}(\hat{x}^i, \hat{x}^j) = \Theta^i(\Sigma^{ii})^{-1}\Sigma^{ij}(\Sigma^{jj})^{-1}(\Theta^j)^\top$
- $\Gamma = [\Gamma^{ij}]$, where $\Gamma^{ij} = \hat{\Sigma}^{ij} \otimes R^{ij}$
- $\eta = \text{vec}(P^1\hat{\Sigma}^{11}, \dots, P^n\hat{\Sigma}^{nn})$

Summary

Multi-step decentralized estimation

Key observation The problem at time t is a one-shot optimization problem

Optimal estimator Let $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid I^i_t]$ and $\hat{\Sigma}^{ij}_t = \text{cov}(\hat{x}^i_{\text{local},t}, \hat{x}^j_{\text{local},t})$. Then,

$$\hat{x}^i_t = L^i_t \, \hat{x}^i_{\text{local},t}, \qquad \text{vec}(L^i_t) = -\big[\hat{\Sigma}^{ij}_t \otimes R^{ij}\big]^{-1} \text{vec}(P^i \hat{\Sigma}^{ii}_t)$$

Remarks To compute the optimal solution, we only need to compute $\hat{x}^i_{\text{local},t}$ and $\hat{\Sigma}^{ij}_t$.

Recall, all random variables are jointly Gaussian. Pre-computing $\hat{\Sigma}^{ij}_t$ and keeping track of $\hat{x}^i_{\text{local},t}$ is trivial but for computational complexity.

Almost same as standard Kalman filtering! Relatively straight forward to come up with recursive equations (but for notation!).

Summary

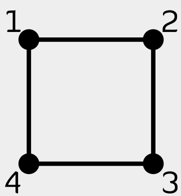
General graph

Information structure

$$I^1_t = \{y^1_{1:t}, y^2_{1:t-1}, y^3_{1:t-2}, y^4_{1:t-1}\} = \{\underbrace{y^1_t, y^1_{t-1}, y^2_{t-1}, y^4_{t-1}}_{\text{local info}}, \underbrace{y_{1:t-2}}_{\text{common info}}\}$$

Local estimates

Recall $\hat{x}^i_{\text{local},t} = \mathbb{E}[x_t \mid I^i_t]$. Then,



$$\hat{x}^1_{\text{local},t} = A\hat{x}_{t-1|t-2} + K^1_t \left\{ \begin{bmatrix} y^1_t \\ y^1_{t-1} \\ y^2_{t-1} \\ y^4_{t-1} \end{bmatrix} - \begin{bmatrix} C^1 A_t \\ C^1 \\ C^2 \\ C^4 \end{bmatrix} \hat{x}_{t-1|t-2} \right\}$$

Remarks

- Effectively equivalent to d-step delayed sharing.
- Each node keeps track of a delayed centralized estimator and innovation wrt common information.