Distributed consensus-based Bayesian estimation: sufficient conditions for performance characterization

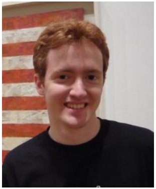
Damiano Varagnolo, Gianluigi Pillonetto, Luca Schenato

Department of Information Engineering - University of Padova

ACC - July 1^{st}, 2010

entirely written in LaTeX2ε using Beamer and Tik Z

Networked Control Systems Group - Padova, Italy



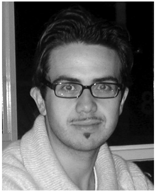
Saverio Bolognani

Angelo Cenedese

Alessandro Chiuso

Enrico Lovisari

Gianluigi Pillonetto

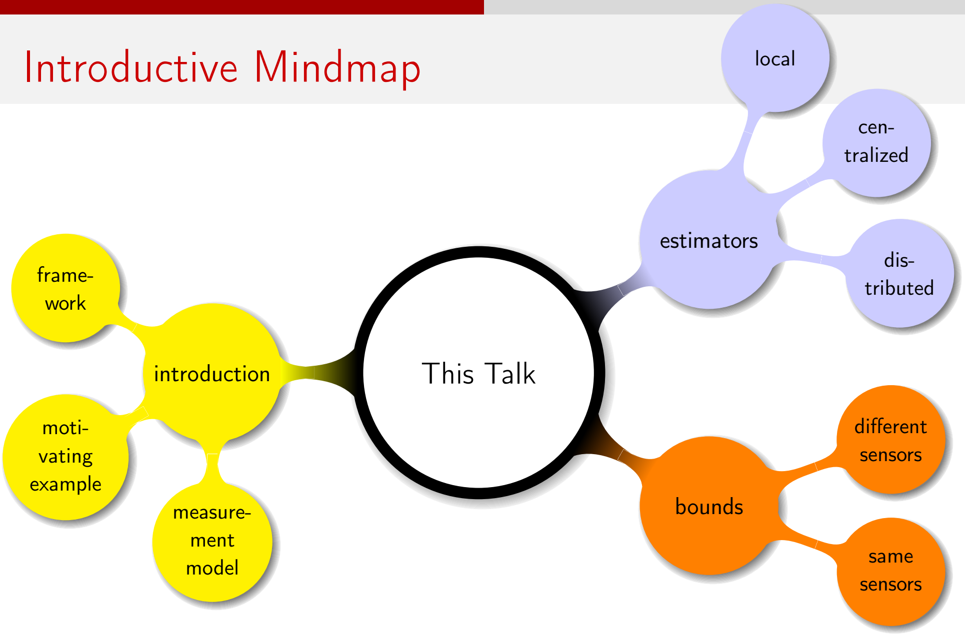


Luca Schenato

Damiano Varagnolo

Sandro Zampieri

Introductive Mindmap



Framework

Distributed Estimation

- measurements from many sensors
- no central coordinating unit
- no direct communications
- limited sensors knowledge
- no time dependency (no dynamic systems)

Distributed Algorithms should…

- do not rely on a-priori topology knowledge
- be robust to node failure and topology changes

A Motivating Example

(artificially built on ZebraNet project)

Network = { bunch of zebras with wireless sensors

Want to relate:

- temperature
- humidity
- thirstyness

(image: thundafunda.com)

We can use. . .

- *either* "local estimates" (bad since inexpensive sensors)
- *or* try to exploit sensors redundancy

Our contributions

Existing literature on:

- bounds on the number of sensors / measurements to obtain a desired level of accuracy (ex. [Kearns and Seung, 1995], [Yamanishi, 1997])

Our focus on:

- fully distributed algorithms
- comparisons between local and distributed algorithms

Main contribution: sufficient conditions assuring "to share is better"

Measurement Model

$$\mathbf{y}_i = C_i \mathbf{a} + \nu_i, \qquad i = 1, \ldots, S \tag{1}$$

where:

M : number of measurements

S : number of sensors

P : number of parameters $(M \gg P)$

$\mathbf{y}_i \in \mathbb{R}^M$: vector of measurements from sensor i

$\mathbf{a} \in \mathbb{R}^P$: vector of unknown parameters s.t. $\mathbf{a} \sim \mathcal{N}(0, \Sigma_{\mathbf{a}})$

$\nu_i \in \mathbb{R}^M$: vector of noises s.t. $\nu_i \sim \mathcal{N}(0, \sigma_i^2 I_M)$ (i.i.d.)

will focus on the case $C_i = C$

Standard Bayesian estimators

Original formulation

Local estimator:

$$\widehat{\mathbf{a}}_{\mathrm{loc},i} := \operatorname{cov}(\mathbf{a}, \mathbf{y}_i)\operatorname{var}(\mathbf{y}_i)^{-1}\mathbf{y}_i \tag{2}$$

Centralized estimator:

$$\widehat{\mathbf{a}}_{\mathrm{cent}} := \operatorname{cov}\left(\mathbf{a}, \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_S \end{bmatrix}\right)\operatorname{var}\left(\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_S \end{bmatrix}\right)^{-1}\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_S \end{bmatrix} \tag{3}$$

Standard Bayesian estimators

Numerical formulation

Local estimator:

$$\widehat{\mathbf{a}}_{\text{loc},i} := \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \sigma_i^2 I_M \right)^{-1} \mathbf{y}_i \tag{4}$$

Centralized estimator:

$$\widehat{\mathbf{a}}_{\text{cent}} := \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \left(\sum_{i=1}^{S} \frac{1}{\sigma_i^2} \right)^{-1} I_M \right)^{-1} \frac{\sum_{i=1}^{S} \frac{\mathbf{y}_i}{\sigma_i^2}}{\sum_{i=1}^{S} \frac{1}{\sigma_i^2}} \tag{5}$$

Distributing the Centralized Bayesian estimators

Definition: harmonic mean of the measurements noises variances:

$$h := \frac{S}{\sum_{i=1}^{S} \frac{1}{\sigma_i^2}} = \frac{1}{\frac{1}{S}\sum_{i=1}^{S} \frac{1}{\sigma_i^2}} \quad \rightarrow \quad \text{average consensus} \tag{6}$$

If all sensors know S then centralized estimator can be computing using average consensus:

$$\widehat{\mathbf{a}}_{\text{cent}} := \frac{\frac{1}{S}\sum_{i=1}^{S} \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \frac{h}{S} \cdot I_M \right)^{-1} \frac{\mathbf{y}_i}{\sigma_i^2}}{\frac{1}{S}\sum_{i=1}^{S} \frac{1}{\sigma_i^2}} \tag{7}$$

...and if the sensors do not know the number of sensors S?

use a guess $\overline{S}$: $$\widehat{\mathbf{a}}_{\text{dist}}\left(\overline{S}\right) := \frac{\dfrac{1}{S}\displaystyle\sum_{i=1}^{S} \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \frac{h}{\overline{S}} \cdot I_M \right)^{-1} \frac{\mathbf{y}_i}{\sigma_i^2}}{\dfrac{1}{S}\displaystyle\sum_{i=1}^{S} \frac{1}{\sigma_i^2}} \tag{8}$$

For sure behaves not better w.r.t. the centralized estimator:

$$\operatorname{var}\left(\widehat{\mathbf{a}}_{\text{dist}}\left(\overline{S}\right) - \mathbf{a}\right) \geq \operatorname{var}\left(\widehat{\mathbf{a}}_{\text{cent}} - \mathbf{a}\right) \qquad \forall\ \Sigma_{\mathbf{a}},\ h,\ C,\ P,\ M\ .$$

And w.r.t. the local one??

First bound: when a single sensor is sure to perform better with the distributed strategy

If:

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right] \tag{9}$$

then:

$$\mathrm{var}\left(\widehat{\mathbf{a}}_{\mathrm{dist}}\left(\overline{S}\right) - \mathbf{a}\right) < \mathrm{var}\left(\widehat{\mathbf{a}}_{\mathrm{loc},i} - \mathbf{a}\right) \qquad \forall\ \Sigma_{\mathbf{a}},\ h,\ C,\ P,\ M\ .$$

Second bound: when all the sensors are sure to perform better with the distributed strategy

Define $\sigma^2_{\min} := \min_i \{\sigma_i^2\}$. If:

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma^2_{\min}}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma^2_{\min}}} \right] \qquad (10)$$

then:

$$\mathrm{var}\left(\widehat{\mathbf{a}}_{\mathrm{dist}}\left(\overline{S}\right) - \mathbf{a}\right) < \mathrm{var}\left(\widehat{\mathbf{a}}_{\mathrm{loc},i} - \mathbf{a}\right) \qquad \forall\ i,\ \Sigma_{\mathbf{a}},\ h,\ C,\ P,\ M\ .$$

Third bound: when sensors "in average" perform better with the distributed strategy

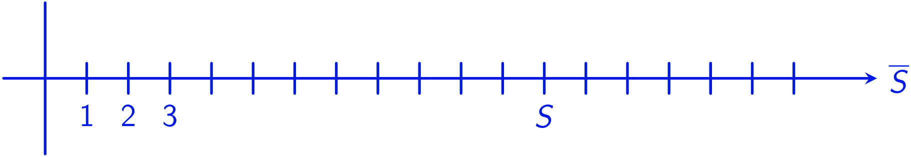
If:

$$\overline{S} \in \left[S - \sqrt{S^2 - S}, \; S + \sqrt{S^2 - S} \right] \tag{11}$$

then:

$$\operatorname{var}\left(\widehat{\mathbf{a}}_{\text{dist}}\left(\overline{S}\right) - \mathbf{a}\right) < \frac{1}{S} \sum_{i=1}^{S} \operatorname{var}\left(\widehat{\mathbf{a}}_{\text{loc},i} - \mathbf{a}\right) \qquad \forall \, \Sigma_{\mathbf{a}}, \; h, \; C, \; P, \; M \, .$$

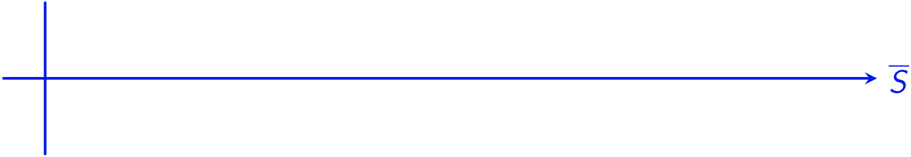
Graphical intuition of the bounds



Graphical intuition of the bounds

first bound: given sensor performs better with given strategy

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right]$$



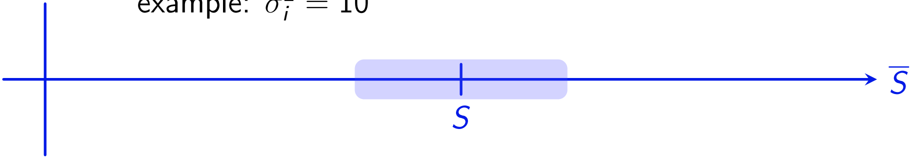
Graphical intuition of the bounds

first bound: given sensor performs better with given strategy

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right]$$

1) noise increases $\Rightarrow$ bound grows

example: $\sigma_i^2 = 10$



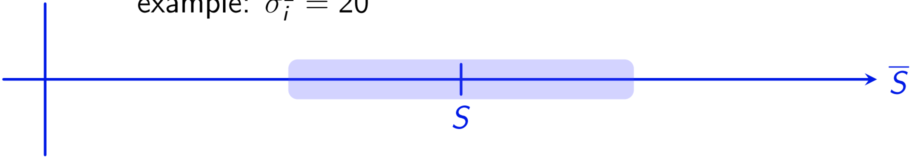
Graphical intuition of the bounds

first bound: given sensor performs better with given strategy

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right]$$

1) noise increases $\Rightarrow$ bound grows

example: $\sigma_i^2 = 20$



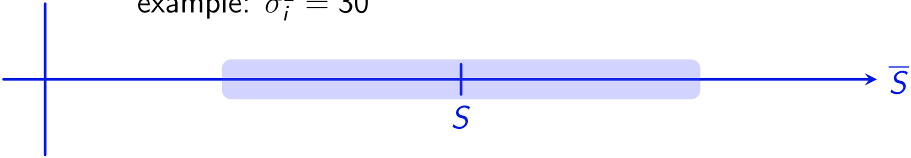
Graphical intuition of the bounds

first bound: given sensor performs better with given strategy

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right]$$

1) noise increases $\Rightarrow$ bound grows

example: $\sigma_i^2 = 30$



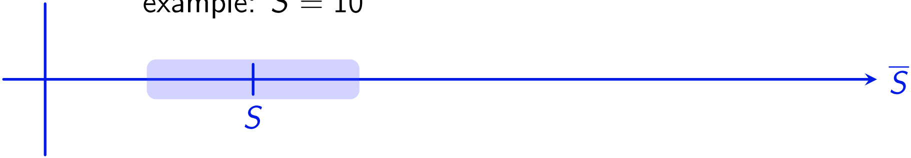
Graphical intuition of the bounds

first bound: given sensor performs better with given strategy

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right]$$

2) S increases $\Rightarrow$ bound shifts and grows

example: $S = 10$



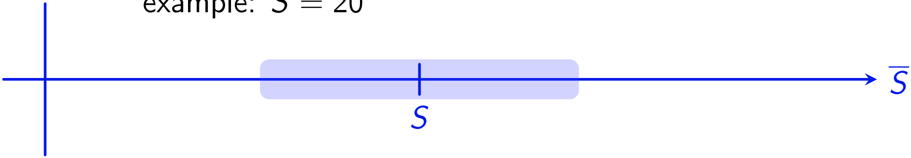
Graphical intuition of the bounds

first bound: given sensor performs better with given strategy

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right]$$

2) S increases $\Rightarrow$ bound shifts and grows

example: $S = 20$



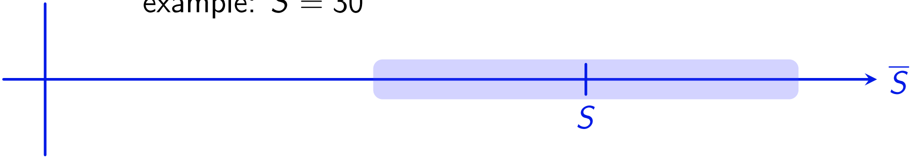
Graphical intuition of the bounds

first bound: given sensor performs better with given strategy

$$\overline{S} \in \left[S - \sqrt{S^2 - \frac{Sh}{\sigma_i^2}},\ S + \sqrt{S^2 - \frac{Sh}{\sigma_i^2}} \right]$$

2) S increases $\Rightarrow$ bound shifts and grows

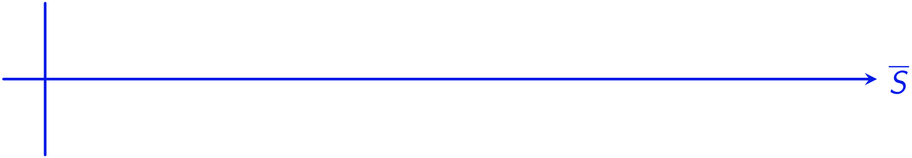
example: $S = 30$



Graphical intuition of the bounds

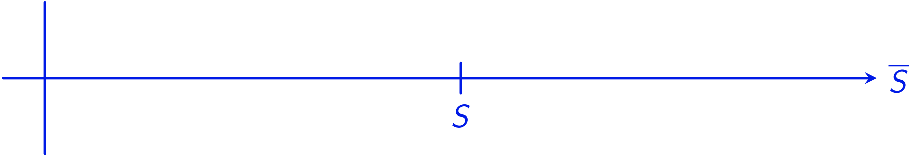
second and third bounds: similar behaviors

- noisiness increases ⇒ bounds grow
- S increases ⇒ bounds grow and shift



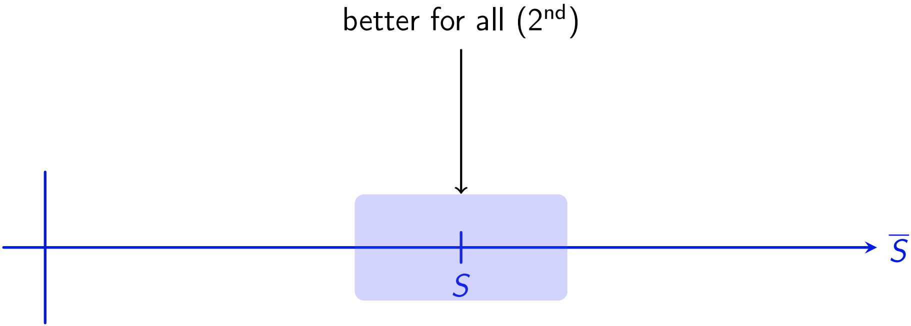
Graphical intuition of the bounds

relations between the bounds:



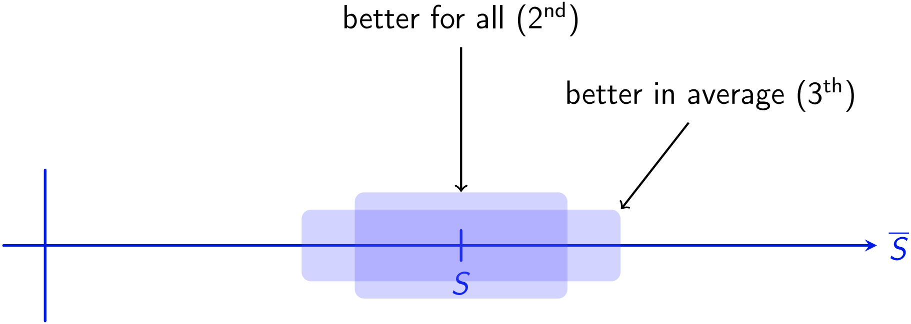
Graphical intuition of the bounds

relations between the bounds:



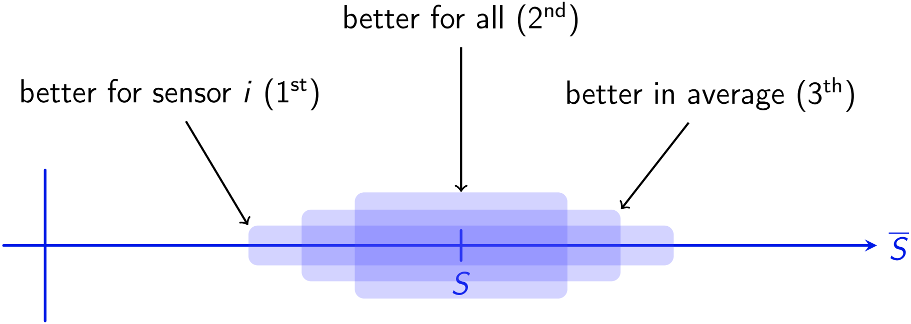
Graphical intuition of the bounds

relations between the bounds:



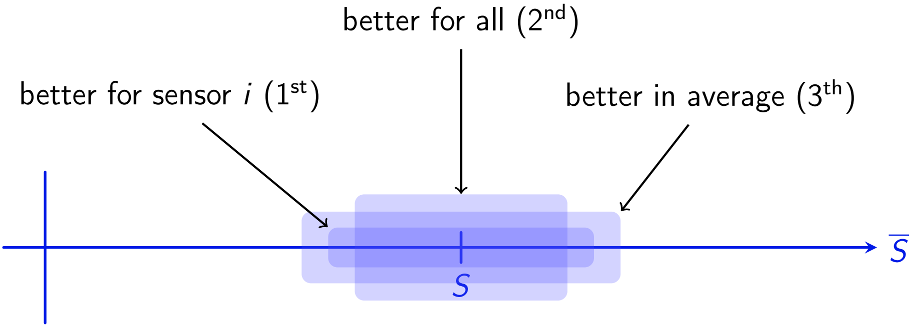
Graphical intuition of the bounds

relations between the bounds:



Graphical intuition of the bounds

relations between the bounds:



Simplified situation

When all sensors are evenly noisy: $\sigma_i^2 = \sigma^2$

Optimal centralized estimator has simplified structure:

$$\widehat{\mathbf{a}}_{\text{cent}} := \frac{1}{S} \sum_{i=1}^{S} \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \frac{\sigma^2}{S} I_M \right)^{-1} \mathbf{y}_i \qquad (12)$$

still requires the knowledge of S!

If not, use a guess $\overline{S}$:

$$\widehat{\mathbf{a}}_{\text{dist}} \left(\overline{S} \right) := \frac{1}{S} \sum_{i=1}^{S} \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \frac{\sigma^2}{\overline{S}} I_M \right)^{-1} \mathbf{y}_i \qquad (13)$$

Bounds for the simplified situation

How the bounds on $\overline{S}$ s.t.:

$$\text{var}\left(\widehat{\mathbf{a}}_{\text{dist}}\left(\overline{S}\right)-\mathbf{a}\right) < \text{var}\left(\widehat{\mathbf{a}}_{\text{loc},i}-\mathbf{a}\right) \qquad \forall\, \Sigma_{\mathbf{a}},\ \alpha,\ C,\ P,\ M$$

modify??

- directly previous results:

$$\overline{S} \in \left[S-\sqrt{S^2-S}, S+\sqrt{S^2-S}\right] \tag{14}$$

- using different proofs:

$$\overline{S} \in \left[1,\ \ 2\left(S-1\right)\right] \tag{15}$$

An useful corollary

$$\overline{S} \in [1,\ \ 2\,(S-1)] \quad \Rightarrow \quad \text{can choose } \overline{S} = 1$$

that imply:

$$\widehat{\mathbf{a}}_{\mathrm{loc},i} = \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \sigma^2 I_M \right)^{-1} \mathbf{y}_i \tag{16}$$

is worse than:

$$\widehat{\mathbf{a}}_{\mathrm{dist}}(1) = \frac{1}{S} \sum_{i=1}^{S} \Sigma_{\mathbf{a}} C^T \left(C \Sigma_{\mathbf{a}} C^T + \sigma^2 I_M \right)^{-1} \mathbf{y}_i = \frac{1}{S} \sum_{i=1}^{S} \widehat{\mathbf{a}}_{\mathrm{loc},i} \tag{17}$$

i.e. always better to share local optimal estimates (once computed)!

Conclusions and Future Works

Conclusions

- there exists bounds assuring distributed estimators to behave "better" than local ones (under mild assumptions)
- can use these bounds to justify naïve algorithms

Future Works

- instead of C consider sensor dependent C_i
- consider non-parametric function estimation (infinite-dimensional functions instead of finite-dimensional vectors)

Distributed consensus-based Bayesian estimation: sufficient conditions for performance characterization

Damiano Varagnolo, Gianluigi Pillonetto, Luca Schenato

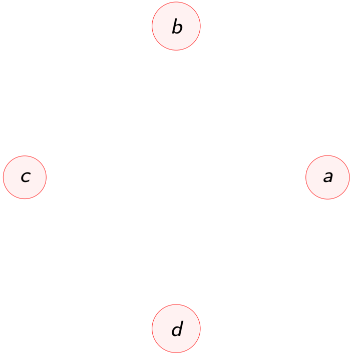
Department of Information Engineering - University of Padova

ACC - July 1^{st}, 2010

damiano.varagnolo@dei.unipd.it
www.dei.unipd.it/∼varagnolo/
google: damiano varagnolo

entirely written in LaTeX2ε using Beamer and TikZ

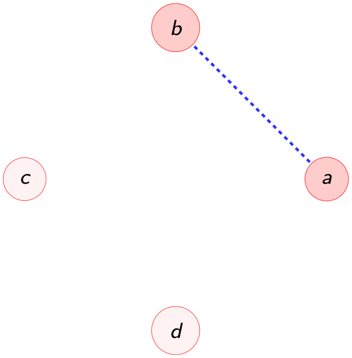
Example of how average consensus works [Cortés, 2008]



$t = 1$

a: 20.5
b: 13.5
c: 19.6
d: 18.0

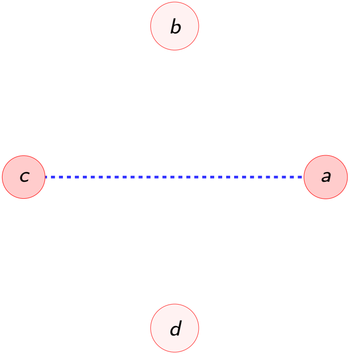
Example of how average consensus works [Cortés, 2008]



$t = 2$

a: 17.0
b: 17.0
c: 19.6
d: 18.0

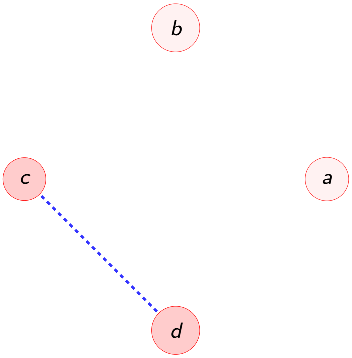
Example of how average consensus works [Cortés, 2008]



$t = 3$

a:	18.3
b:	17.0
c:	18.3
d:	18.0

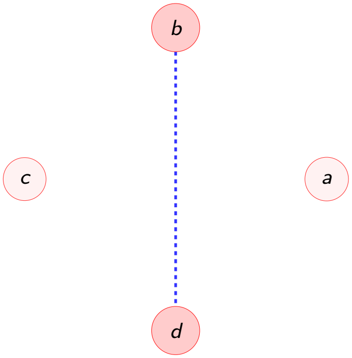
Example of how average consensus works [Cortés, 2008]



$t = 4$

a: 18.3
b: 17.0
c: 18.15
d: 18.15

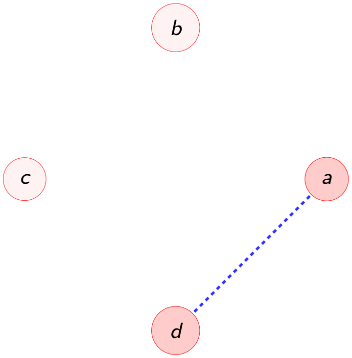
Example of how average consensus works [Cortés, 2008]



$t = 5$

a: 18.3
b: 17.575
c: 18.15
d: 17.575

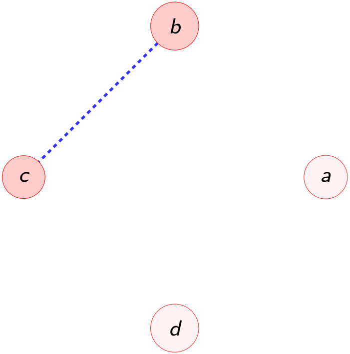
Example of how average consensus works [Cortés, 2008]



$t = 6$

a: 17.9375
b: 17.575
c: 18.15
d: 17.9375

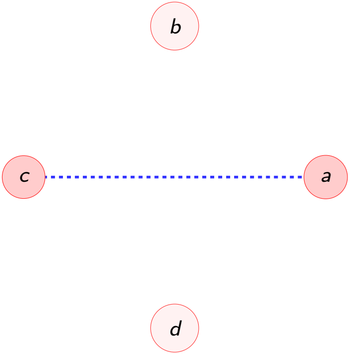
Example of how average consensus works [Cortés, 2008]



$t = 7$

a: 17.9375
b: 17.8625
c: 17.8625
d: 17.9375

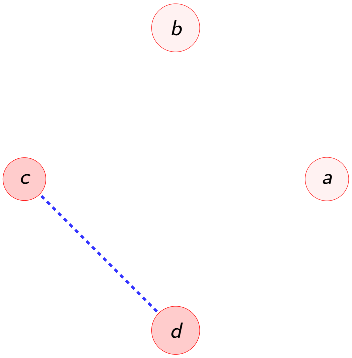
Example of how average consensus works [Cortés, 2008]



$t = 8$

a: 17.9
b: 17.8625
c: 17.9
d: 17.9375

Example of how average consensus works [Cortés, 2008]



$t = 9$

a: 17.9
b: 17.8625
c: 17.9187
d: 17.9187

Cortés, J. (2008).
Automatica .

Kearns, M. and Seung, H. S. (1995).
Machine Learning 18 (2-3), 255–276.

Yamanishi, K. (1997).
In: COLT '97: Proceedings of the tenth annual conference on Computational learning theory pp. 250–262, New York, NY, USA: ACM.