Distributed averaging on digital noisy networks

Paolo Frasca

Dipartimento di Matematica
Politecnico di Torino
Torino, Italy
http://calvino.polito.it/~frasca

joint work with R. Carli (UCSB), G. Como (MIT), F. Garin (INRIA)

Spring 2010 Thematic Program on Distributed Decision-Making and Control
LCCC, LTH, Lund, Sweden

April 9, 2010

Summary

The averaging problem

Given

- a strongly connected graph $\mathcal{G} = (V, E)$, $\quad |V| = N$
  - nodes are agents;
  - edges are available communication channels.
- $\forall\, v \in V$, $\theta_v \in \mathbb{R}$.

We want to compute the average

$$y = \frac{1}{N} \sum_{v \in V} \theta_v .$$

in spite of the communication limitations.

Linear averaging algorithm

Proposition

Let $\mathbf{x}(k+1) = P\,\mathbf{x}(k)$, *and* $\mathbf{x}(0) = \boldsymbol{\theta}$, *with* P *adapted to* $\mathcal{G}$. *If* P *is doubly stochastic and has positive diagonal* $\implies$ *the algorithm converges to the average:*

$$\lim_{k\to\infty} x_v(k) = y \quad \forall v \in V.$$

Moreover, let

$$n(\delta) := \inf\left\{n \in \mathbb{N} :\ N^{-1}\|\mathbf{x}(m) - y\mathbf{1}\|^2 \leq \delta, \forall m \geq n\right\}\ .$$

Then,

$$n(\delta) \leq C\frac{\log \delta^{-1}}{\log \rho^{-1}},$$

where ρ *is the second largest eigenvalue of* P*.*

But this algorithm requires communication of real numbers!

Can a network of digital noisy channels be used to compute averages?

To answer this question we need

- a model for digital communication and computation;
- useful convergence and performance notions;
- (possibly) algorithms solving the problem.

Digital broadcast communication

At each time instant $t \in \mathbb{N}$,

- every agent $v \in V$ broadcasts a binary signal $a_v(t) \in \{0, 1\}$ to its out-neighbourhood $\mathcal{N}_v^+$;
- every agent $w \in \mathcal{N}_v^+$ receives a possibly erased version $b_{v\to w}(t) \in \{0, 1, ?\}$ of $a_v(t)$;
- each agent $v \in V$ makes an estimate $\hat{y}_v(t)$ of y.

We assume that

- the communication network is memoryless;
- for every $v \in V$, $w \in \mathcal{N}_v^+$, and $t \in \mathbb{N}$,

$$b_{v\to w}(t) = \begin{cases} ? & \text{w.p.} \quad \varepsilon \\ a_v(t) & \text{w.p.} \quad 1-\varepsilon\,. \end{cases}$$

Performance measures I

Given a required precision $\delta \in]0,1]$, we define two complexity figures.

- *Communication complexity*

$$\tau(\delta) := \inf \left\{ t \in \mathbb{N} : \; N^{-1}\mathbb{E}\left[||\hat{\mathbf{y}}(s) - y\mathbf{1}||^2 \right] \leq \delta, \forall s \geq t \right\} \;,$$

the minimum number of binary transmissions each agent has to perform in order to guarantee that the average mean squared estimation error does not exceed δ.

Performance measures II

- *Computational complexity:*
  denote by $\kappa_v(t)$ the minimum number of operations required by agent v for to compute $\alpha_v(t)$ and $\hat{y}_v(t)$; define

$$\kappa(\delta) := \max \left\{ \sum_{1 \leq t \leq \tau(\delta)} \kappa_v(t) : \; v \in V \right\},$$

  the maximum, over all agents $v \in V$, of the total number of operations required to be performed, in order to achieve an average mean squared estimation error not exceeding δ.

Algorithm I

We want to adapt to the digital setting the linear update

$$x_v(j+1) = \sum_{w \in \mathcal{N}_v^-} P_{vw} x_w(j) + P_{vv} x_v(j) \,.$$

But x_w can not be transmitted as such on a noisy channel! Instead, $v \in \mathcal{N}_w^+$ can only obtain, after some binary transmissions, an estimate $\hat{x}_w^{(v)}$.

A solution consists in alternating

- transmission phases of increasing length $\ell_j = S_L\, j$
- averaging steps

$$x_v(j+1) = \sum_{w \in \mathcal{N}_v^-} P_{vw} \hat{x}_w^{(v)}(j) + P_{vv} x_v(j) \,.$$

Algorithm II

- During transmission, agents use *linear tree codes*, which [CFZ08] allow $u \in \mathbb{R}$ to be transmitted with precision

$$\mathbb{E}\left[(u - \hat{u}_\ell)^2\right] \leq \beta_L^{2\ell},$$

  requiring a number of operations $k_\ell^L \leq B\ell^3$, for all $\ell \geq 0$, where $\beta_L \in (0,1)$, and $B > 0$ are constants depending on the erasure probability ε only.
- *why increasing length of transmission phases?* Increasing length, precision is increased. If precision is not increased, the committed errors accumulate, and $\mathbf{x}(t)$ does not converge [XBK07].

Result

Theorem

For any S_L, there exists a real-valued random variable $\hat{y}$ s. t.

- $\mathbb{E}\left[(y-\hat{y})^2\right] \leq \dfrac{\alpha^2}{(1-\alpha)^2}$, *where* $\alpha = \beta_L^{S_L}$, *and*
- *with probability one,*

$$\lim_{t\to\infty} \hat{y}_v(t) = \hat{y}\,, \qquad \forall v \in V\,.$$

Moreover, for all $\delta \in]0,1]$, one can choose S_L so that

$$\tau_L(\delta) \leq C_1 + C_2 \frac{\log^3 \delta^{-1}}{\log^2 \rho^{-1}}\,, \qquad \kappa_L(\delta) \leq C_3 + C_4 \frac{\log^7 \delta^{-1}}{\log^4 \rho^{-1}}\,,$$

where $\{C_i : i = 1, \dots, 4\}$ are positive constants depending on ε only.

Algorithm variation

Using

- sub-exponential *repetition codes*, which have linear complexity,
- $\ell_j = S_R j^2$,

one obtains a performance

$$\tau_R(\delta) \leq C_5 + C_6 \frac{\log^5 \delta^{-1}}{\log^3 \rho^{-1}}, \qquad \kappa_R(\delta) \leq C_7 + C_8 \frac{\log^5 \delta^{-1}}{\log^3 \rho^{-1}},$$

for all $\delta \in]0,1]$.

Feedback channels

If the channels allow communication feedback, the state update step can be replaced by

$$x_v(j+1) = x_v(j) - \sum_{w \in \mathcal{N}_v^+} P_{wv} \hat{x}_v^{(w)}(j) + \sum_{w \in \mathcal{N}_v^-} P_{vw} \hat{x}_w^{(v)}(j) \,.$$

Theorem

For any S_L, then with probability one,

$$\lim_{t \to \infty} \hat{y}_v(t) = y \,, \qquad \forall v \in V \,.$$

Moreover, for all $\delta \in]0,1]$, one can choose S_L so that

$$\tau'_L(\delta) \leq C'_1 + C'_2 \frac{\log^2 \delta^{-1}}{\log^2 \rho^{-1}} \,, \qquad \kappa'_L(\delta) \leq C'_3 + C'_4 \frac{\log^4 \delta^{-1}}{\log^4 \rho^{-1}} \,,$$

where $\{C'_i : i = 1, \dots, 4\}$ are positive constants depending on ε only.

Remarks

About the presented algorithms:

- Results can be extended from erasure channels to symmetric memoryless channels.
- Why $\log^3(\delta)$ communication complexity?
  ($\#$ update steps $\times$ initial precision $\times$ increasing precision)
- Performance depends on topology, via ρ.
- No global knowledge (e.g., topology) is required to the agents.
- Communication feedback can be effectively exploited.

Open problem: Is there hope for an algorithm with $\log(\delta)$ communication complexity?

Further reading

R. Carli, G. Como, P. Frasca, and F. Garin.
Average consensus on digital noisy networks.
In *IFAC NeCSys*, Venice, Italy, September 2009.

R. Carli, F. Bullo, and S. Zampieri.
Quantized average consensus via dynamic coding/decoding schemes.
Int. J. Robust and Nonlinear Control, 20(2):156-175, 2010.

G. Como, F. Fagnani, and S. Zampieri.
Anytime reliable transmission of real-valued information through digital noisy channels.
SIAM JCO, In press, February 2010.

P. Frasca, R. Carli, F. Fagnani, and S. Zampieri.
Average consensus on networks with quantized communication.
Int. J. Robust and Nonlinear Control, 19(16):1787–1816, 2009.

L. Xiao, S. Boyd, and S.-J. Kim.
Distributed average consensus with least-mean-square deviation.
J. Parallel and Distributed Computing, 67(1):33-46, 2007