PSC: Progettazione di sistemi di controllo **a.a. 2010-2011**

Lezione 20 — 30 Novembre 2010

Docente: Luca Schenato *Stesori: Maran F., Marcon R., Marcassa A., Zanella F.*

20.1 Tipologie di comunicazione

Dalla lezione precedente, si è visto come l'evoluzione dello stato di una rete tempo-variante rispetti l'equazione di aggiornamento

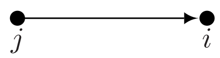
$$x(t+1) = P(t)x(t)$$

con $P(0),\ P(1),\dots$ i.i.d., $P(t) \in \mathcal{P}$, $\mathbb{P}\left[P(t) = P_i\right] = p_i$; la randomizzazione viene usata per modellare l'imprecisione dovuta all'ambiente. $x(t)$ è una sequenza aleatoria, essendolo le $P(t)$, e il *probabilistic consensus* corrisponde a $x(t) \to \alpha\mathbb{1}$ c.p.1 con α vettore aleatorio, $\alpha = \rho(\omega)^T x(0)$.

Teorema 20.1. *Se le $P(t)$ hanno diagonale positiva c.p.1 e il grafo relativo a $\bar{P} = \mathbb{E}\left[P(t)\right]$, $\mathcal{G}_{\bar{P}}$, è fortemente connesso, allora si ha probabilistic consensus.*

Esempi di applicazione del teorema su algoritmi randomizzati
Se w_{ij} è la probabilità che si accenda l'edge da j a i, si può dimostrare che $\mathcal{G}_W = \mathcal{G}_{\bar{P}}$ (a parte la diagonale).

1. *Asymmetric Gossip*: supposto

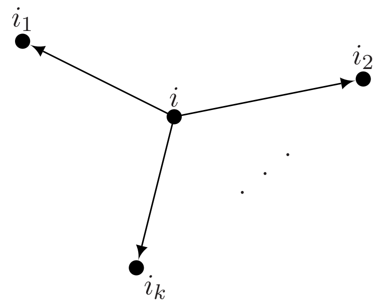
   

   l'algoritmo è del tipo

   $$\begin{aligned} x_i(t+1) &= (1-q)x_i(t) + qx_j(t) \\ x_l(t+1) &= x_l(t) \end{aligned}$$

   e non mantiene la media. $P(t)$ ha sempre la diagonale positiva, e questo verifica la prima condizione; per la seconda, è sufficiente che W porti a un grafo fortemente connesso.

2. *Broadcast*: si considera un grafo orientato in cui ad ogni istante si accende un nodo (non un edge) e trasmette a tutti i suoi vicini; tutti quelli che sentono prendono l'informazione e ne fanno la media.

   $$\begin{aligned} x_{i_1}(t+1) &= (1-q)x_{i_1}(t) + qx_i(t) \\ x_{i_2}(t+1) &= (1-q)x_{i_2}(t) + qx_i(t) \\ &\vdots \\ x_{i_k}(t+1) &= (1-q)x_{i_k}(t) + qx_i(t) \end{aligned}$$

Anche in questo caso la media non viene mantenuta. $P(t)$ ha sempre diagonale positiva (prima condizione verificata); inoltre, se $w_i > 0\ \forall i$ allora tutti i nodi hanno probabilità di accendersi e si può dimostrare che $\mathcal{G}_{\bar{P}} = \mathcal{G}$ grafo di partenza (seconda condizione verificata).

In entrambi gli algoritmi non c'è parallelismo nell'accensione degli edge (se ne attiva sempre uno alla volta).

20.2 Analisi delle prestazioni

Si vuole analizzare la velocità di convergenza al consensus: questo viene comunemente fatto sulla base di due metriche, l'analisi *a regime* e *in transitorio*.

20.2.1 Analisi a regime

Idealmente, si dovrebbe convergere al baricentro delle condizioni iniziali: ci si chiede quanto il valore asintotico sia lontano. Definiamo il baricentro al tempo t come

$$x_A(t) \triangleq \frac{1}{N}\sum_i x_i(t) = \frac{1}{N}\mathbb{1}^T x(t)$$

(nel symettric gossip $x_A(t)$ è costante). Il parametro di interesse è $\beta(t) = |x_A(t) - x_A(0)|^2$, la prestazione a regime è determinata da $\beta(\infty)$:

$$\begin{aligned}
\beta(\infty) &= |x_A(t) - x_A(0)|^2 \\
&= \left|\frac{1}{N}\mathbb{1}^T x(\infty) - \frac{1}{N}\mathbb{1}^T x(0)\right|^2 \\
x(\infty) = \mathbb{1}\rho^T x(0) \rightarrow &= \left|\underbrace{\left(\rho^T - \frac{1}{N}\mathbb{1}^T\right)}_{\triangleq v^T} x(0)\right|^2 \\
v^T x(0) \text{ è scalare} \rightarrow &= x(0)^T \underbrace{vv^T}_{\triangleq B(\omega)} x(0) \\
&= x(0)^T B(\omega) x(0)
\end{aligned}$$

che è una forma quadratica, essendo $B(\omega) = \left(\rho(\omega) - \frac{1}{N}\mathbb{1}\right)\left(\rho(\omega) - \frac{1}{N}\mathbb{1}\right)^T$ semi-definita positiva di rango 1; la dipendenza da ω indica che c'è ancora aleatorietà.

20.2.2 Analisi in transitorio

Per vedere la velocità di convergenza, la metrica da considerare è

$$d'(t) = \frac{1}{N}\|x(t) - x(\infty)\|^2;$$

ovviamente $d'(t) \rightarrow 0$. Nel caso deterministico la convergenza è esponenziale, ma in condizioni di stocasticità la $d'(t)$ dipende anche dal parametro ω. Può essere riscritta come

$$d'(t) = \frac{1}{N}\sum_i |x_i(t) - x_i(\infty)|^2;$$

che mette in evidenza come ad ogni t sia la media della distanza di ogni singolo agente dal suo valore finale.

Un'altra metrica da considerare è

$$d(t) = \frac{1}{N}\|x(t) - \mathbb{1}x_A(t)\|^2,$$

distanza dello stato rispetto al baricentro istantaneo; anche $d(t) \to 0$. Vale il seguente

Lemma 20.1. *$d(t) \leq d'(t) \leq (1+\sqrt{N})^2 d(t)$, quindi se $d(t) \to 0$ esponenzialmente anche $d'(t) \to 0$ esponenzialmente (teorema dei carabinieri).*

Alla luce di questo risultato, è preferibile studiare $d(t)$ (più semplice).

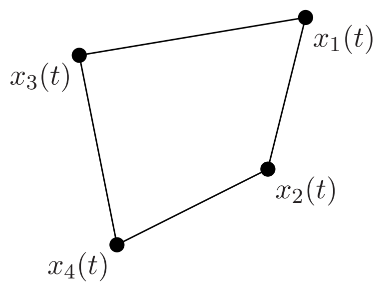
Teorema 20.2. *Nelle condizioni di probabilistic consensus,*

$$\lim_{t\to\infty} d(t)^{\frac{1}{t}} = \lambda \in \mathbb{R}$$

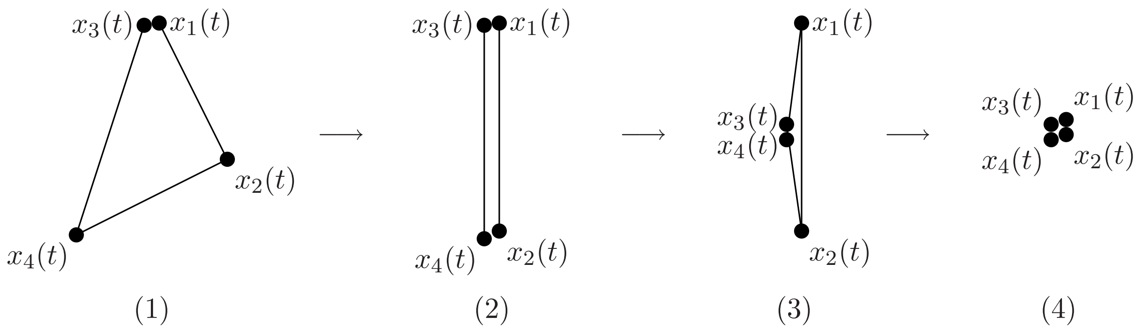
dove $d(t)^{\frac{1}{t}}$ (aleatorio) è il rate di convergenza*.*

λ è detto *secondo esponente di Lyapunov* per $d(t)$ (il primo è 1). È un valore difficile da calcolare, e pone inoltre un secondo problema. Si consideri l'algoritmo di symmetric gossip con grafo completo, $q = \frac{1}{2}$, in cui tutti gli edge hanno la stessa probabilità di accendersi $w_{ij} = \frac{1}{N}$: prendiamo $N = 2^\nu$. Da esperimenti numerici si ottiene, per $N = 16$, una decrescenza esponenziale con $\lambda \simeq 0.93$.

La teoria di Lyapunov mostra invece che $\lambda = 0$ (dead-beat: dopo un numero finito di passi si arriva al consenso). Prendiamo infatti $N = 4$, e supponiamo che dopo t passi i vari $x_i(t)$ siano così disposti



una possibile evoluzione è



e quindi in quattro azioni successive si giunge al consensus, ovvero $d(t) = 0$ da un certo punto in poi. Ciò significa che

$$\underbrace{P(0)P(1)\cdots P(t-1)}_{x_1,x_2,x_3,x_4 \text{ di partenza}} \cdot \underbrace{P(t)P(t+1)P(t+2)\overset{\text{consenso}}{\overset{\downarrow}{P(t+3)}}}_{\text{con prob. piccola questa sequenza arriva}} .$$

Per $N = 4$ il sistema in effetti è dead-beat; lo è anche per $N = 16$, ma il tempo in cui $d(t) = 0$ è molto più grande.

Ci si può dunque chiedere cosa sia lo 0.93 ricavato empiricamente: esso emerge dall'analisi della varianza d'errore (analisi in media). La risposta si può ricercare facendo l'analisi dei momenti a regime.

Si suppone $\beta(\infty) = x(0)^T B(\omega)x(0)$ dove $B(\omega) = (\rho - \frac{1}{N}\mathbf{1})(\rho - \frac{1}{N}\mathbf{1})^T$ in questo modo si riesce a calcolare il valore atteso: $\mathbb{E}[\beta(\infty)] = x(0)^T \bar{B}x(0)$ dove:

$$\begin{aligned}\bar{B} &= \mathbb{E}[B] \\ &= \mathbb{E}[\rho\rho^T] - \frac{1}{N}\mathbb{E}[\rho]\mathbf{1}^T - \frac{1}{N}\mathbf{1}\mathbb{E}[\rho]^T + \frac{1}{N^2}\mathbf{1}\mathbf{1}^T\end{aligned}$$

$\mathbb{E}[\rho]$ è l'autovettore sinistro di $\bar{P}$ cioè $\mathbb{E}[\rho]^T\bar{P} = \mathbb{E}[\rho]^T d$ Anche se l'algoritmo non mantiene la media la $\bar{P}$ è comunque doppiamente stocastica. In tal caso $\mathbb{E}[\rho] = \frac{1}{N}\mathbf{1}$ e la risultante $\bar{B} = \mathbb{E}[\rho\rho^T] - \frac{1}{N^2}\mathbf{1}\mathbf{1}^T$

Per quanto riguarda il transitorio si vuole capire quanto velocemente tende a zero l'aspettazione: $\mathbb{E}[d(t)] \to 0$. La velocità di convergenza è di tipo esponenziale; si può notare questo da come è fatta la funzione $d(t)$:

$$\begin{aligned}d(t) &= \frac{1}{N}||x(t) - x_A(t)\mathbf{1}||^2 \\ &= \frac{1}{N}||x(t) - \frac{\mathbf{1}\mathbf{1}^T}{N}x(t)||^2 \\ &= \frac{1}{N}||(I - \frac{\mathbf{1}\mathbf{1}^T}{N})x(t)||^2 \\ &= \frac{1}{N}||\Omega x(t)||^2\end{aligned}$$

Quindi, tenendo conto delle propritá di Ω:

1. Ω è simmetrico
2. Ω è una proiezione $\Omega = \Omega \cdot \Omega$

e del fatto che $x(t) = Q(t)x(0)$, si può calcolare l'aspettazione $\mathbb{E}[d(t)]$:

$$\begin{aligned}\mathbb{E}[d(t)] &= \frac{1}{N}\mathbb{E}[x(t)^T\Omega^T\Omega x(t)] \\ &= \frac{1}{N}\mathbb{E}[x(t)^T\Omega x(t)] \\ &= \frac{1}{N}x(0)^T\mathbb{E}[Q(t)^T\Omega Q(t)]x(0) \\ &= x(0)^T\Delta(t)x(0)\end{aligned}$$

dove $\Delta(t) = \frac{1}{N}\mathbb{E}[Q(t)^T\Omega Q(t)]$.
$\Delta(t)$ evolve come un sistema lineare tempo discreto e tende a zero in maniera esponenziale.

$$\begin{aligned}
\Delta(t+1) &= \mathbb{E}\left[P(0)^T P(1)^T \cdots P(t)^T \Omega P(t) \cdots P(0)\right] = \\
&= \mathbb{E}\left[P(0)^T P(1)^T \cdots P(t)^T \Omega P(t) \cdots P(0)|P(0)\right] = \\
&= \mathbb{E}\left[P(0)^T \mathbb{E}\left[P(1)^T \cdots P(t)^T \Omega P(t) \cdots P(1)|P(0)\right] P(0)\right] = \\
&= \mathbb{E}\left[P(0)^T \Delta(t) P(0)\right]
\end{aligned}$$

avendo sfruttato la proprietà della probabilità condizionata $\mathbb{E}[z] = \mathbb{E}\left[\mathbb{E}[z|x]\right]$.
Si definisce quindi $L : \Re^{N\times N} \to \Re^{N\times N}$:

$$\Delta \to \mathbb{E}\left[P(0)^T \Delta P(0)\right] \to \begin{cases} \Delta(t+1) &= L(\Delta(t)) \\ \Delta(0) &= \Omega \end{cases}$$

Si prenda ora lo spazio vettoriale $R =< \Omega, \alpha(\Omega), \alpha^2(\Omega), \cdots > \to L_R$, e

$$L(\Delta) = \mathbb{E}\left[P(0)^T \Delta P(0)\right] = \sum_i p_i P_i^T \Delta P_i$$

dove $P(0) = P_i$ con probabilità p_i; L è un operatore lineare anche se mappa matrici in matrici.
Si definisce ora $\delta(t) = \mathrm{vect}(\Delta(t))$, dove l'operatore *vect* è dato da

$$\mathrm{vect}(E|M) = \mathbb{E}\left[\mathrm{vect}M\right]$$

L'aggiornamento di δ si ottiene come:

$$\delta(t+1) = \mathrm{vect}\left(\mathbb{E}\left[P(0)^T \Delta(t) P(0)\right]\right) = \mathbb{E}\left[\mathrm{vect}\left(P(0)^T \Delta(t) P(0)\right)\right] \tag{20.1}$$

Esiste inoltre una relazione tale per cui, date tre matrici A, B, $C \in \Re^{n\times n}$,

$$\mathrm{vect}(ABC) = (C^T \otimes A)\mathrm{vect}(B)$$

Applicando questa proprietà alla relazione riportata in (20.1) si trova che

$$\delta(t+1) = \mathbb{E}\left[P(0)^T \otimes P(0)\right] \delta(t)$$

con $E\left[P(0)^T \otimes P(0)\right] = L$ matrice $\in \Re^{N^2\times N^2}$ che rappresenta la media di tutti i prodotti di Kroenecker.
Si può dimostrare che L^T è stocastica, visto che valgono le due proprietà circa le matrici stocastiche:

1. se L ha diagonale > 0 con probabilità 1

2. G_L è fortemente connesso

con queste due proprietà si arriva al consensus: si ha un autovalore pari a 1, mentre tutti gli altri sono contenuti nel cerchio unitario.