PSC: Progettazione di sistemi di controllo	**a.a. 2010-2011**
Lezione 19 — 26 Novembre 2010	
Docente: Luca Schenato	*Stesori: Maran F., Marcon R., Marcassa A., Zanella F.*

19.1 Algoritmi di consensus randomizzati

Finora si sono sempre considerati sistemi *tempo-invarianti*, ovvero descritti dalla legge

$$x^+ = Px$$

con P costante nel tempo: l'obiettivo è trovare P tale che tutte le componenti del vettore $x(t)$ convergano al medesimo valore, cioè $x(t) \to \alpha\mathbb{1}$. Ciò accade se il grafo associato a P, $\mathcal{G}_P$, è *fortemente connesso*: in tal caso $\alpha = w^T x(0)$, dove w è l'*autovettore sinistro* di P relativo all'autovalore 1 ($\Leftrightarrow w^T P = w^T$; l'autovettore *destro* relativo a 1 è $\mathbb{1}$). Se P è *doppiamente stocastica*, allora

$$\alpha = \frac{1}{N}\sum_i x_i(0) \Leftrightarrow w^T = \frac{1}{N}\mathbb{1}^T;$$

vale anche il viceversa (condizione *necessaria e sufficiente*).

Nei due esempi successivi la $P(t)$ varia in modo aleatorio nel tempo.

Esempi

1. *Algoritmo gossip*: è un algoritmo randomizzato (condizione imposta dall'esterno in modo da evitare la collisione di messaggi, lo scambio di informazioni avviene in tempi diversi).

   Si parte da un grafo $\mathcal{G}$, che si suppone per semplicità *completo* (N^2 edge) e *non orientato*: la matrice di adiacenza è simmetrica e completa. Ad ogni istante di tempo, si attiva un edge (nell'insieme ε di lati) con probabilità $w_{ij} \geq 0$: dato che se ne accende uno solo alla volta, si suppone $\sum_{i,j} w_{ij} = 1$.

   Si consideri ad esempio $w_{ij} = \frac{1}{N^2}\ \forall i,j$, in modo da usare *tutti* gli edge. Quando si attiva un arco $j \longrightarrow i$, $i \neq j$, i nodi j e i si scambiano l'informazione riguardante il proprio stato, calcolando il nuovo stato come *combinazione convessa*

   $$x_i(t+1) = (1-q)x_i(t) + qx_j(t) \tag{19.1}$$
   $$x_j(t+1) = qx_i(t) + (1-q)x_j(t); \tag{19.2}$$

   tutti gli altri stati non ricevono informazione

   $$x_l(t+1) = x_l(t), \quad \forall l \neq i,j. \tag{19.3}$$

   P è casuale, dipende da come si accendono gli edge, e le attivazioni sono indipendenti (ciò che succede al tempo $t+1$ è indipendente da ciò che succede al tempo t).

Non si è certi della convergenza dell'algoritmo, ma in caso positivo converge alla media delle condizioni iniziali (*average consensus*): infatti il valore $\frac{1}{N}\sum_i x_i(t) = \frac{1}{N}\mathbb{1}^T x(t)$ ($\leftarrow$ *baricentro*) resta sempre lo stesso ne tempo, dato che

$$
\begin{aligned}
\frac{1}{N}\sum_l x_l(t+1) &= \frac{1}{N}\left\{\sum_{l\neq i,j} x_l(t+1) + x_i(t+1) + x_j(t+1)\right\} \\
\text{da (19.1), (19.2), (19.3)} \rightarrow &= \frac{1}{N}\left\{\sum_{l\neq i,j} x_l(t) + (1-q)x_i(t) + qx_j(t) + qx_i(t) + (1-q)x_j(t)\right\} \\
&= \frac{1}{N}\sum_l x_l(t).
\end{aligned}
$$

Ergo se $x_l(t) \overset{t\to\infty}{\longrightarrow} \alpha\ \forall l$, allora α è tale che $\frac{1}{N}\sum_l \underbrace{x_l(\infty)}_{\alpha} = \frac{1}{N}\sum_l x_l(0)$.

Si dimostra che questo è l'unico algoritmo randomizzato che preserva la media: ci sono tuttavia svariati casi in cui, se N è sufficientemente elevato, la distanza fra valore finale e media è piccola (idealmente coincidono per $N \to \infty$).

2. *Perdita di pacchetti*: si considera un tradizionale algoritmo di consenso in cui però l'informazione da j a i può non arrivare.

   Si parte in questo caso da una matrice P ben precisa, ad esempio

$$
P_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & & & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} \end{bmatrix} \tag{19.4}
$$

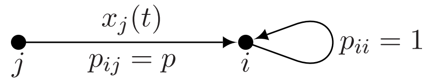
   circolante, oppure

$$
P_2 = \begin{bmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \ddots & \vdots \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix} \tag{19.5}
$$

   completa ($\Rightarrow$ convergenza dead-beat al baricentro).

   In generale, se j vuole trasmettere informazioni a i all'istante t si tiene conto di una *probabilità di ricezione* p_{ij} definita come

$$
p_{ij} = \begin{cases} 1 & \text{se } i = j \\ p & \text{se } i \neq j \end{cases}
$$



Se viene perso un pacchetto, si devono adattare i pesi (la loro somma deve restare unitaria $\Rightarrow$ P deve rimanere stocastica): un modo semplice è aggiungere al peso del nodo che non riceve informazione quello del nodo dal quale non l'ha ricevuta. Si introduce così la variabile

$$L_{ij}(t) \triangleq \begin{cases} 1 & \text{se l'informazione da } j \text{ a } i \text{ arriva} \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} 1 & \text{con probabilità } p \\ 0 & \text{con probabilità } 1-p \end{cases}, \quad i \neq j;$$

$L_{ii}(t) \equiv 1 \ \forall i$; $L_{ij}(t)$ è indipendente dai nodi e nel tempo.

L'algoritmo di compensazione dei pesi diviene

$$x_i(t+1) = \left[P_{ii} + \sum_{j \neq i} P_{ij} \left(1 - L_{ij}(t)\right) \right] x_i(t) + \sum_{j \neq i} P_{ij} L_{ij}(t) x_j(t),$$

in contrapposizione al caso senza perdita di pacchetti in cui si ha

$$x_i(t+1) = \sum_j P_{ij} x_j(t) = P_{ii} x_i(t) + \sum_{j \neq i} P_{ij} x_j(t).$$

Usando P_1 definita in (19.4), si ha

$$x_i(t+1) = \frac{1}{2} x_i(t) + \frac{1}{2} x_{i+1}(t) = \begin{cases} \frac{1}{2} x_i(t) + \frac{1}{2} x_{i+1}(t) & \text{se l'informazione da } j \text{ a } i \text{ arriva} \\ x_i(t) & \text{altrimenti} \end{cases}.$$

Se si è in condizioni di agreement, l'algoritmo *non* fa perdere l'accordo ($x_i(t) = \alpha \Rightarrow x_i(t+1) = \alpha, \ \forall i$).

19.2 Definizioni e condizioni di convergenza

Il sistema diventa del tipo

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

con $P(0), P(1), \ldots$ sequenza di variabili aleatorie a valori matriciali indipendenti. Si suppone inoltre che $P(t)$ assuma valori in un insieme $\mathcal{P} \subseteq \mathbb{R}^{N \times N}$ *finito*, ergo $\mathcal{P} = \{P_1, \ldots, P_h\}$ è una lista di valori da cui il sistema sceglie a caso, e che $\mathbb{P}[P(t) = P_i] = p_i$ ($\sum_i p_i = 1$); infine, le $P(t)$ sono i.i.d..

Quindi $P(t, \omega)$ è funzione del parametro $\omega \in \Omega$ (spazio degli eventi) che decide quale sequenza è selezionata.

Definizione 19.1 (Probabilistic Consensus). *L'algoritmo sopra descritto dà* probabilistic consensus *se $x(t) \to \alpha\mathbb{1}$ c.p.1 (*almost surely*).*

α dipende da $x(0)$ (come nel caso deterministico) e da $\omega \in \Omega$ (a parte nel caso del gossip) $\Rightarrow \alpha(x(0), \omega)$. *Fissato* ω (ovvero fissata la sequenza delle $P(t)$), $\alpha(x(0), \omega)$ è lineare in $x(0)$ $\Rightarrow$ $\alpha(x(0), \omega) = \rho(\omega)^T x(0)$. La linearità è facile da dimostrare.

***Dimostrazione*:** consideriamo la condizione iniziale $x(0)$ come somma di due condizioni iniziali $x_1(0),\ x_2(0)$:

$$x(0) = x_1(0) + x_2(0) \Rightarrow x(t) + x_1(t) + x_2(t) \xrightarrow{t \to \infty} x(\infty) = x_1(\infty) + x_2(\infty)$$

e quindi

$$\alpha\mathbb{1} = x(\infty) = x_1(\infty) + x_2(\infty) = \alpha_1\mathbb{1} + \alpha_2\mathbb{1} \Rightarrow \alpha = \alpha_1 + \alpha_2;$$

perciò α è lineare (la moltiplicazione per costante è banale). $\square$

Quindi, $x(t) \to \left(\rho^T(\omega)x(0)\right)\mathbb{1}$.

Definiamo

$$Q(t,\omega) \triangleq P(t-1,\omega)\cdots P(1,\omega)\cdot P(0,\omega);$$

è facile vedere che $x(t,\omega) = Q(t,\omega)x(0)$.

Proposizione 19.1. *$Q(t,\omega) \to \mathbb{1}\rho(\omega)^T$ c.p.1 è la condizione equivalente a $x(t) \to \alpha\mathbb{1}$ c.p.1 (probabilistic consensus).*

***Dimostrazione*:** se $Q(t,\omega) \to \mathbb{1}\rho(\omega)^T$,

$$Q(t,\omega)x(0) \to \mathbb{1}\,\underbrace{\rho(\omega)^T x(0)}_{\triangleq\alpha} = \mathbb{1}\alpha;$$

se $x(t) \to \alpha\mathbb{1}$,

$$Q(t,\omega)x(0) = x(t) \to \underbrace{\rho(\omega)^T x(0)}_{\text{scalare}}\,\mathbb{1} = \mathbb{1}\rho(\omega)^T x(0)$$

e prendendo $x(0)$ come vettore della base canonica,

$$Q(t,\omega) \to \mathbb{1}\rho(\omega)^T.$$

$\square$

Teorema 19.1. *Si supponga che*

1. *$P(t)$ ha gli elementi della diagonale positivi c.p.1 ($P(t) = P_i$ con probabilità p_i);*
2. *sia $\bar{P} = \mathbb{E}[P(t)] = \sum_i p_i P_i$: $\mathcal{G}_{\bar{P}}$ è fortemente connesso;*

allora si ha probabilistic consensus.

Osservazione 1

Se le condizioni 1 e 2 del teorema precedente sono verificate, allora $Q(t) \to \mathbb{1}\rho^T$ c.p.1, quindi

$$\mathbb{E}[Q(t)] \to \mathbb{E}\left[\mathbb{1}\rho^T\right] = \mathbb{1}\left[\rho^T\right];$$

per l'indipendenza delle $P(t)$ e dalla condizione 2, si ha

$$\mathbb{E}[Q(t)] = \mathbb{E}[P(t-1)]\,\mathbb{E}[P(t-2)]\cdots = \bar{P}^t$$

e quindi

$$\bar{P}^t \to \mathbb{1}\mathbb{E}\left[\rho^T\right].$$

Se si usasse $\bar{P}$ si avrebbe consenso deterministico, con $w^T = \mathbb{E}\left[\rho^T\right]$.

Osservazione 2

Se $\bar{P}$ è doppiamente stocastica, allora $\mathbb{E}\left[\rho^T\right] = \frac{1}{N}\mathbb{1}^T$ (*baricentro*). Non è detto però che se $\bar{P}$ è doppiamente stocastica l'algoritmo converga al baricentro $\forall\omega$ (solo in media). Per avere l'average probabilistic consensus, condizione necessaria e sufficiente è che $P(t)$ sia doppiamente stocastica c.p.1 (P_i doppiamente stocastica $\forall i$) (l'average probabilistic consensus corrisponde alla convergenza alla media c.p.1).

Si ha inoltre che $P(t)$ doppiamente stocastica $\Leftrightarrow$ $\bar{P}$ doppiamente stocastica.

19.2.1 Applicazioni del teorema

1. *Algoritmo gossip*: definiamo una matrice R^{ij} come

$$
R^{ij} = \begin{array}{cc}
 & \begin{array}{ccccccccccccc} & & & & i & & & & & j & & & \end{array} \\
\begin{array}{c} \\ \\ \\ \\ i \\ \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & \cdot & & & & & \cdot & & & \\
 & 1 & & & \cdot & & & & & \cdot & & & \\
 & & \ddots & & \cdot & & & & & \cdot & & & \\
 & & & 1 & \cdot & & & & & \cdot & & & \\
\cdot & \cdot & \cdot & \cdot & 1-q & \cdot & \cdot & \cdot & \cdot & q & & & \\
 & & & & \cdot & 1 & & & & \cdot & & & \\
 & & & & \cdot & & 1 & & & \cdot & & & \\
 & & & & \cdot & & & \ddots & & \cdot & & & \\
 & & & & \cdot & & & & 1 & \cdot & & & \\
\cdot & \cdot & \cdot & \cdot & q & \cdot & \cdot & \cdot & \cdot & 1-q & & & \\
 & & & & & & & & & & 1 & & \\
 & & & & & & & & & & & \ddots & \\
 & & & & & & & & & & & & 1
\end{bmatrix}
\end{array}
$$

con 0 nelle posizioni non indicate; è evidente che

$$
R^{ij} \begin{bmatrix} x_1(t) \\ \vdots \\ x_i(t) \\ \vdots \\ x_j(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \begin{bmatrix} x_1(t) \\ \vdots \\ (1-q)x_i(t) + qx_j(t) \\ \vdots \\ qx_i(t) + (1-q)x_j(t) \\ \vdots \\ x_N(t) \end{bmatrix}.
$$

Supponiamo che $\sum_{i,j} w_{ij} = 1$: tutte le R^{ij} sono doppiamente stocastiche (essendo simmetriche e stocastiche). Supponiamo inoltre che $0 < q < 1$: in tal caso tutti gli elementi della diagonale sono positivi (condizione 1 del teorema verificata); dimostriamo che $\mathcal{G}_{\bar{P}} = \mathcal{G}_{W+W^T}$, dove $W = [w_{ij}]$.

***Dimostrazione*:** gli elementi sulla diagonale di $\bar{P}$ sono tutti positivi (tutti loop);

$$
j \bullet\!\longrightarrow\!\bullet\, i \ \text{ in } \mathcal{E}_W \Leftrightarrow j \bullet\!\longrightarrow\!\bullet\, i \ \text{ in } \mathcal{E}_{\bar{P}} \qquad \text{se } i \neq j
$$

infatti $\bar{P} = \sum_{i,j} w_{ij} R^{ij}$, $\left(\bar{P}\right)_{hk} = \sum_{i,j} w_{ij} \left(R^{ij}\right)_{hk}$; prendo $h \neq k$,

$$
\left(R^{ij}\right)_{hk} = \begin{cases} q & \text{se } (i,j) = (h,k) \text{ oppure } (i,j) = (k,h) \\ 0 & \text{altrove} \end{cases}
$$

quindi $\left(\bar{P}\right)_{hk} = \sum_{i,j} w_{ij} \left(R^{ij}\right)_{hk} = qw_{hk} + qw_{kh} > 0$ se e solo se $w_{hk} > 0$ o $w_{kh} > 0 \Leftrightarrow \left(W + W^T\right)_{hk} > 0$. $\square$

Si può dimostrare che se $\mathcal{G}_{\bar{P}} = \mathcal{G}_{W+W^T}$ non vale non si può avere consenso, quindi è anche condizione necessaria.

2. *Perdita di pacchetti*: da quanto visto precedentemente, per $i \neq j$ $P_{ij}(t) = L_{ij}(t)P_{ij}$, mentre $P_{ii}(t) = P_{ii} + \sum_i (1 - L_{ij}(t)) P_{ij} \geq P_{ii}$, quindi scegliendo la diagonale positiva resta sempre > 0 e dunque è verificata la prima condizione del teorema.

   Inoltre, se $i \neq j$ $\left(\bar{P}\right)_{ij} = \mathbb{E}\left[P_{ij}(t)\right] = (1-p) \cdot 0 + p \cdot P_{ij} = pP_{ij}$: finchè $p > 0$, $\mathcal{G}_{\bar{P}} = \mathcal{G}_P$ fuori diagonale.