Performance of linear average-consensus algorithm in large-scale networks

FEDERICA GARIN

(NeCS team, INRIA Rhône-Alpes and gipsa-lab, Grenoble, France)

joint work with:

SANDRO ZAMPIERI, ENRICO LOVISARI, RUGGERO CARLI

(DEI, Univ. di Padova, Italy)

LCCC Focus Period ‘Information and Control in Networks’, Lund Univ. , Oct. 2012

Distributed estimation and control

An active research trend in the control-theory community

- Wireless sensor networks, e.g.,
  - fire alarms in forests
  - irrigation of large green-houses
  - camera networks: surveillance, motion capture
- mobile multi-agent coordination
  - robots or drones (Unmanned Aerial Vehicles, Autonomous Underwater Vehicles, smart cars)
  - perform formation control, patrolling, source seeking
- model of animal or social behavior
  - opinion dynamics in social networks
  - animal flocking and herding

(Average) consensus

- Problem: all agents need to agree on a value
  Moreover, they need to (approx.) compute a given fct. of initial values, usually the average.
- Why do we care?
  - Toy example of distributed task. Hope to get deep understanding of fundamental limitations, and hints for further research on more challenging problems
  - Building block necessary to perform more complicated tasks: distributed estimation (e.g., Kalman filter, least squares regression), sensor calibration (e.g., clock synchronization), distributed optimization, formation control
  - Model of social aggregation and flocking

(Average) consensus continued

- Distributed: agents need to agree in a distributed way
  - Simplest scenario: a graph describes allowed communications. Agents can exchange messages with neighbors. Time-invariant graph, synchronous exchanges.
  - Imperfection of communication: quantization of messages, noise, delays
  - Randomly time-varying graph (gossip): model for link failures or randomized algorithm not requiring synchronization. Edges are activated at random, e.g., with independent Poisson clocks.
  - State-dependent time-varying graph: model of social or animal interaction, or mobile robots.
    Agents move to the computed position,
    graph depends on distances.

Some references

- Classic book: Bertsekas and Tsitsiklis, Parallel and distributed computation: Numerical methods, Prentice Hall, 1989
- Classic book (computer science point of view): Lynch, Distributed algorithms, Morgan Kaufmann, 1997
- Seminal paper (1): Olfati-Saber, Murray, Consensus problems in networks of agents with switching topology and time delays, IEEE TAC, 2004
- Seminal paper (2): Moreau, Stability of multi-agent systems with time-dependent communication links, IEEE TAC, 2005
- Book on mobile agents coordination: Bullo, Cortés, Martínez, Distributed Control of Robotic Networks, Princeton, 2009
- Survey on consensus in distributed estimation or control: Garin, Schenato, A survey on distributed estimation and control applications using linear consensus algorithms, in Networked Control Systems, Springer LNCIS, 2011
- Survey on gossip: Dimakis, Kar, Moura, Rabbat, Scaglione, Gossip algorithms for distributed signal processing, Proc. of the IEEE, 2011
- Survey on opinion dynamics: Acemoglu, Ozdaglar, Opinion dynamics and learning in social networks, Dynamic Games and Applications, 2011

Linear Average Consensus (discrete-time LTI)

- Simple setting: time-invariant communication graph, perfect and synchronous communication
- Discrete-time linear algorithm:
  State update = convex combination of neighbors' states
  $x_u(t) = \sum_v P_{uv} x_v(t)$
  Can use only neighbors' states: $P_{uv} = 0$ if $u \nrightarrow v$.
- In vector notation:

$$\boxed{x(t+1) = P\,x(t)}$$

- Design of P:
  - consistent with the graph: $P_{uv} = 0$ if $u \nrightarrow v$.
  - doubly-stochastic: $P_{uv} \geq 0$, row-sum=column-sum=1
  - primitive (strongly connected and aperiodic graph)

Classical performance analysis

From Markov chains literature, Perron-Frobenius theorem

- Assume:
  - P primitive (strongly connected and aperiodic graph);
  - P doubly-stochastic: $P_{ij} \geq 0 \, \forall i, j$, $1^T P = 1^T$, $P1 = 1$
- Eigenvalues of P:
  - 1 with multiplicity 1;
  - $|\lambda| < 1$ for all other eigenvalues
- $\lim_{t \to \infty} x(t) = \frac{1}{N} \sum_i x_i(0)$
- speed of convergence: ρ_{ess}^t
  where $\rho_{\text{ess}} =$ 2nd largest eigenvalues' modulus

New performance indices

- Why?
  - different costs describe different objectives (consensus used in different contexts)
  - in large-scale networks, tools for choosing the correct scaling of N = # nodes and t = time (number of iterations)

- What index?
  - LQ cost (ℓ^2-norm of transient);
  - quadratic estimation error in averaging measures;
  - quadratic error in distributed Kalman filter
  - . . . (taylored to your problem!)

LQ cost (ℓ^2-norm of transient)

- Consensus algorithm $x(t+1) = Px(t)$
- Initial condition $x(0)$ = random variable
  $\mathbb{E}[x(0)] = 0$ and $\mathbb{E}\left[x(0)x^T(0)\right] = I$

- Transient performance evaluation by ℓ^2-norm

$$J_{\mathsf{LQ}}(P) := \frac{1}{N}\sum_{t\geq 0}\mathbb{E}\|x(t) - x_{\text{ave}}\mathbf{1}\|^2 \quad x_{\text{ave}} = \frac{1}{N}\sum_{i=1}^{N} x_i(0)$$

- $J_{\mathsf{LQ}}(P) = \frac{1}{N}\sum_{t\geq 0}\operatorname{trace}\left[(P^t - \frac{1}{N}\mathbf{1}\mathbf{1}^T)^T(P^t - \frac{1}{N}\mathbf{1}\mathbf{1}^T)\right]$
  If P is normal (e.g. symmetric), with notation $\lambda_1 = 1$

$$J_{\mathsf{LQ}}(P) = \frac{1}{N}\sum_{i=2}^{N}\frac{1}{1-|\lambda_i|^2}$$

Other reasons to study the LQ cost

The same cost arises from different problems
For example:

- Consensus with noise in the state update:

  $$x(t+1) = Px(t) + n(t)$$

  Cost = asymptotic variance of distance from consensus
  [Xiao, Boyd, Kim, Distributed average consensus with least mean square deviation, J. Parall. Distrib. Comp, 2007]

- Formation control (platooning)
  Cost = formation coherence
  [Bamieh et al, Coherence in large-scale networks: Dimension dependent limitations of local feedback, TAC 2010]

Quadratic error in distributed estimation

N sensors measure same $y \in \mathbb{R}$ + indep. noises:

$$x_i(0) = y + w_i \qquad \forall i = 1, \ldots, N$$

indep. noises $w_1, \ldots, w_n$, average = 0, variance= 1

- Best estimate of y: the average $\hat{y} = \frac{1}{N} \sum_{i=1}^{N} x_i(0)$
  Compute $\hat{y}$ with consensus: $x(t+1) = P\,x(t)$
- Cost = average quadratic error

$$J_e(P,t) = \frac{1}{N} \mathbb{E}\left[e(t)^T e(t)\right], \quad e_i(t) = x_i(t) - y$$

- $J_e(P,t) = \frac{1}{N} \operatorname{trace}\left[(P^T)^t P^t\right]$
  If P is normal (e.g. symmetric)

$$J_e(P,t) = \frac{1}{N} \sum_{i=1}^{N} |\lambda_i|^{2t}$$

Other costs

- Average distance from consensus in the presence of quantization or noise
- estimation or prediction error in distributed Kalman filter

  . . .

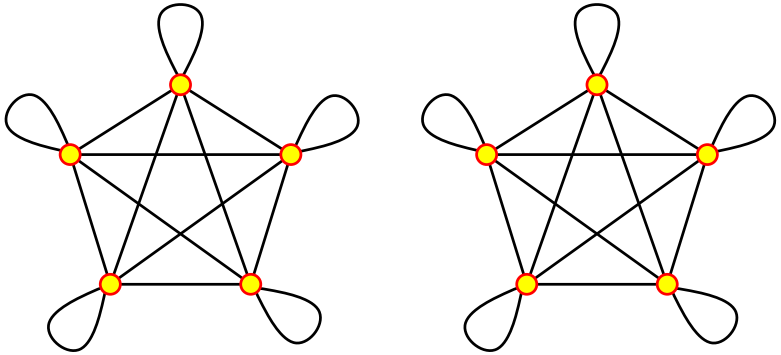
See book chapter:

F. Garin and L. Schenato, A survey on distributed estimation and control applications using linear consensus algorithms, in “Networked Control Systems”, Springer LNCIS, 2010

Example: contrasting performance indices

Toy example where ρ_{ess} very bad, estimation very good:

2 disconnected complete graphs of $n = N/2$ nodes each.



$$P = \left[\begin{array}{c|c} \frac{1}{n}11^T & 0 \\ \hline 0 & \frac{1}{n}11^T \end{array}\right]$$

- eigenvalues: 1 with multipl. 2, 0 with multipl. $N-2$
- NO convergence! (disconnected graph, $\rho_{\text{ess}} = 1$)
- Estim. error: $J_e(P,t) = \frac{1}{N}\sum_i |\lambda_i|^{2t} = \frac{2}{N}\ \forall t \geq 1$
  Almost as good as optimal centralized estimation
  (variance of $\hat{y} = 1/N$).

Consensus and spectral graph theory

- Choice of coefficients also matters, but many properties depend on the graph.
- Spectral graph theory studies eigenvalues of matrices associated with graphs (Adjacency, Laplacian)
- Most literature focused on spectral gap = $1 - \rho_{\text{ess}}(P)$.

  Very interesting results: spectral gap related to a geometric property (expansion).

  There exists expander graphs, with non-vanishing spectral gap ($\rho_{\text{ess}}(P)$ bounded away from 1) despite bounded number of neighbors
- We consider costs depending on all eigenvalues. Must find new results

Consensus and Markov chains

- Doubly-stochastic matrix $P \leftrightarrow$
  Markov chain with uniform invariant measure
- Costs describing consensus performance can be interesting for Markov chains.

  For example, if P is symmetric

  $$J_{\mathrm{LQ}}(P) = \frac{1}{N}\big(\text{average first hitting time of } P^2\big)$$

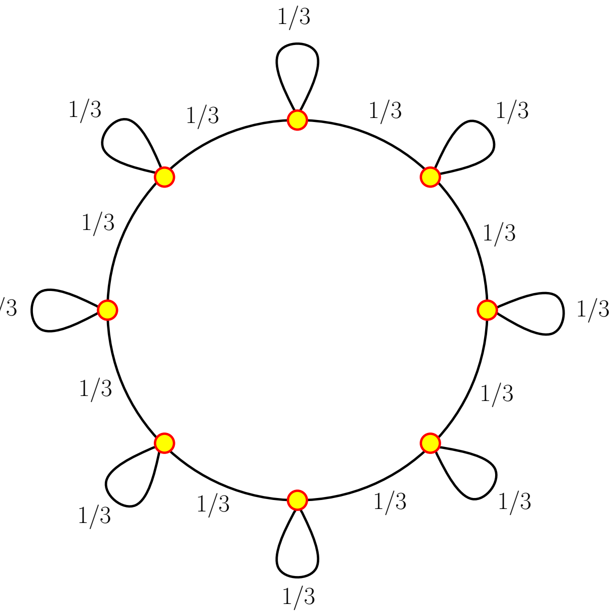
  Average first hitting time = $\frac{1}{N^2}\sum_{u,v} E_{uv}$

  $E_{uv} = \mathbb{E}\big(\min\{t \geq 0 : X_t = v\}\big| X_0 = u\big)$
  X_t Markov chain with transition matrix P^2

Our goals

- Understand effect of graph topology on performance
- Study **large scale** graphs
- Understand the effect of **local interactions**:
  - bounded number of neighbours;
  - some geographical notion of near neighbours (e.g., exclude De Bruijn and other expander graphs, small-word networks etc., because they require some long-range communication)
  - towards a realistic model for sensor networks, even if starting from simplified examples

Simple local communication: circular graph

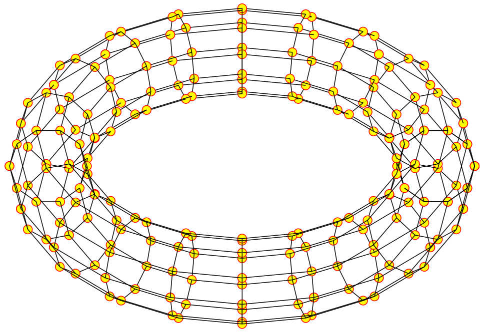


$$P = \begin{bmatrix} 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 \end{bmatrix}$$

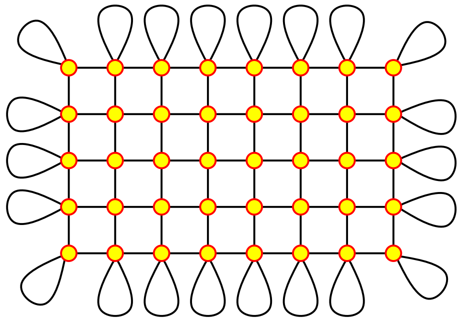
- eigenvalues: $\lambda_h = \frac{1}{3} + \frac{2}{3}\cos(\frac{2\pi}{N}h),\ h = 0, \ldots, N-1$
- 2nd largest $|\lambda|$: $\rho_{\text{ess}} \to 1$ as $1 - \frac{c}{N^2}$
- LQ cost: $J_{\text{LQ}}(P) \asymp N$
- Estim. error: $J_e(P,t) \asymp \max\left(\frac{1}{N}, \frac{1}{\sqrt{t}}\right) \to 0$

Grids (on d-dimensional tori and cubes)

- Generalization of circles:



grid on d-dim. torus
(Abelian Cayley graph)



grid on d-dim. cube
(project. of torus [Boyd et al.])

- 2nd largest $|\lambda|$: $\rho_{\text{ess}} \to 1$ as $1 - \frac{c}{N^{2/d}}$
- LQ cost:

$$J_{\text{LQ}}(P) = \frac{1}{N} \sum_{\lambda \neq 1} \frac{1}{1 - |\lambda|^2} \asymp \begin{cases} N & \text{if } d = 1 \\ \log N & \text{if } d = 2 \\ 1 & \text{if } d \geq 3 \end{cases}$$

- estim. error: $J_e(P,t) = \frac{1}{N} \sum_\lambda |\lambda|^{2t} \asymp \max\left(\frac{1}{N}, \frac{1}{t^{d/2}}\right)$

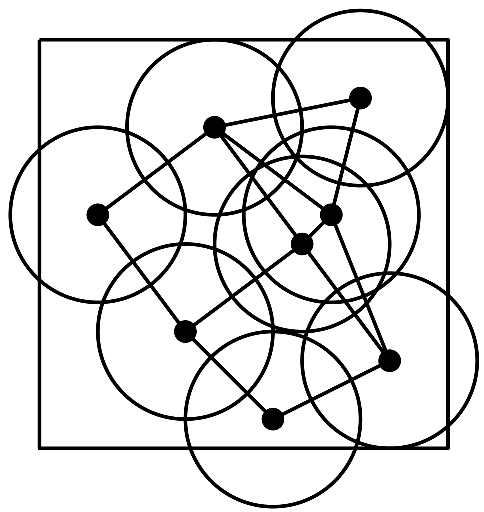
Why Cayley graphs and grids? What's next?

- Why?
  - Elegant mathematical framework:
    Fourier transform on Abelian groups (general. DFT),
    explicit expression for eigenvalues.
  - Example of geographically local interactions
- More realistic models of sensor networks:
  - Random geometric graphs
  - (Deterministic) perturbations of regular grids

  Question: are the scaling laws mostly due to the symmetries, or to some notion of geographically local interaction in d-dimensional Euclidean space?

Random geometric graphs

- Introduced [Gilbert '63], model for wireless sensor networks [Franceschetti, Meester '07]
- Probabilistic model:
  - N points unif. at random within a cube $\subset \mathbb{R}^d$
  - bi-directional edge within points at distance $\leq r$



Random geometric graphs (continued)

- From our **simulations**:
  same behaviour as grids for our quadratic costs
  (connected realizations of random geom. graphs with constant average degree)

- Mathematical results:
  - Well-studied: connectivity threshold (percolation) [Penrose book 2003]
  - Few results on spectrum:
    for simple random walk, above connect. threshold
    - $\rho_{ess} \to 1$ same as grid [Boyd et al. '06]
    - spectral density concentrates to the grid's [Sanatan Rai, PhD thesis, 2005]

Deterministic geometric graphs

- Perturbation of regular grids. Not trivial!
  - Not classical matrix perturbation analysis: not continuous variation of all matrix entries, but significant modification of few entries (e.g., cutting one edge = zeroing one entry)
  - Modifying few edges might significantly change performance (e.g., if disconnects graph)
- F. Fagnani (Polit. Torino), G. Como (Lund) and J.-C. Delvenne (Louvain) study 'democracy' of Markov Chains: how perturbations influence invariant measure, i.e. left eigenvector of eigenvalue 1
- We assume: modified P remains primitive (str. connected graph) and symmetric ($\Rightarrow$ uniform inv. measure)

A powerful tool

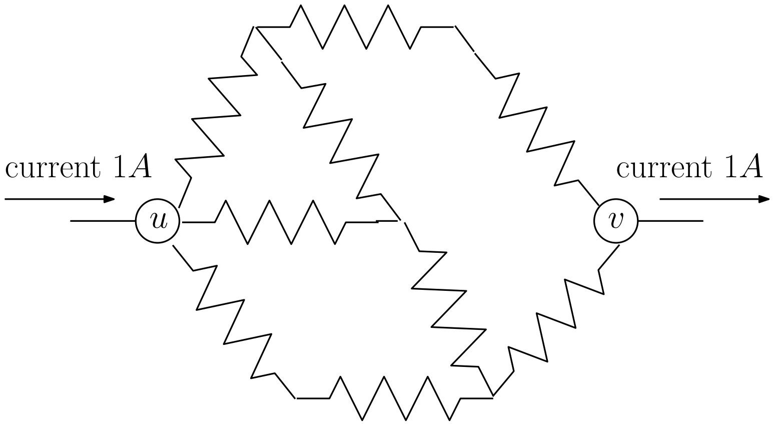
Equivalence:
reversible Markov chains $\leftrightarrow$ resistive electrical networks

- Introduced:
  [Doyle, Snell, Random Walks and Electric Networks, book, 1984]
  Recently used in distributed estimation and control:
  [Barooah, Estimation and control with relative measurements: algorithms and scaling laws, PhD thesis, UCSB, 2007]
  [Ghosh, Boyd, Saberi, Minimizing eff. resist. of a graph, SIAM '08]
- For the symmetric case:

P symmetric stochastic matrix	$\leftrightarrow$	electrical network: • graph associated with P; • on edge (u, v), resistance $R_{uv} = 1/P_{uv}$.

Effective resistance: definition

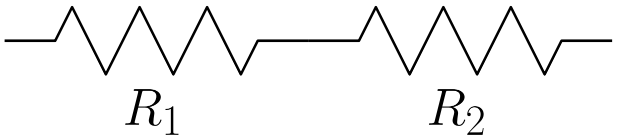
Effective resistance between nodes u, v in the network:



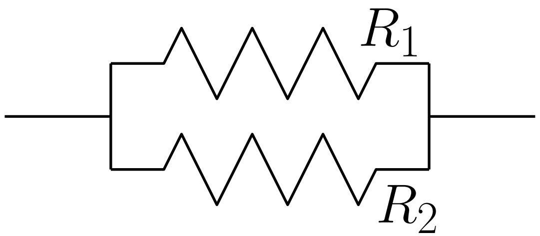
equivalent to:

i.e., $\mathcal{R}^{\text{eff}}_{uv} = V_v - V_u$.

Simple examples:



$$\mathcal{R}^{\text{eff}} = R_1 + R_2$$



$$\mathcal{R}^{\text{eff}} = \frac{R_1 R_2}{R_1 + R_2}$$

Why do we care about effective resistances?

We study the cost

$$J_{\mathsf{LQ}}(P) = \frac{1}{N}\sum_{t\geq 0} \operatorname{trace}\left(P^{2t} - \frac{1}{N}11^T\right)$$

Construct the electrical network associated with P^2.
Then:

$$J_{\mathsf{LQ}}(P) = \frac{1}{N^2}\sum_{u,v} \mathcal{R}^{\text{eff}}_{uv}$$

Cost $J_{\mathsf{LQ}}(P)$ = average effective resistance $\bar{\mathcal{R}}^{\text{eff}}$.

Why do we care about effective resistances? (2)

- Properties of the effective resistances:
  - Monotonicity: if you add an edge, or if you decrease the resistance on an existing edge, then all effective resistances in the network will be decreased or same.
  - Scaling: if all resistances are multiplied by α, then all effective resistances are multiplied by α.

- Bound on eff. resist. using eff. resist. of ‘similar’ network.

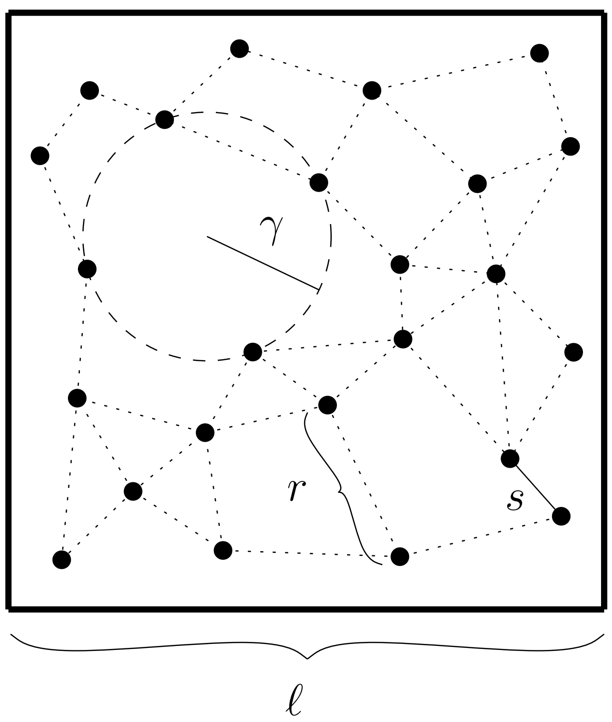
  This is the tool we need to study $J_{\text{LQ}}(P) = \bar{\mathcal{R}}^{\text{eff}}$ of perturbed grids!

Deterministic geometric graphs

Geometric graph:

[Barooah, PhD th. ’07], [Lovisari, Zampieri, Annual Reviews in Control ’12]

- vertices = points in $\mathbb{R}^d$
- 5 geometric parameters:



- ○ ℓ = edge of hypercube containing all nodes;
- ○ s = min. Euclidean dist. between two nodes;
- ○ r = max. Euclidean dist. between two nodes;
- ○ γ = radius of largest empty ball;
- ○ ρ = minimum ratio graphical dist. / Euclid. dist.

Geometric graphs behave like grids

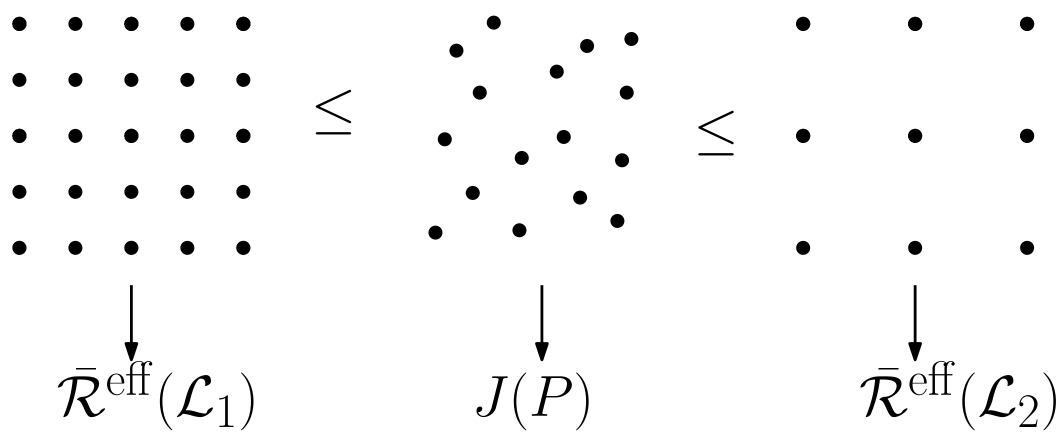
Theorem:

P symm. stoch. primitive, associated with geom. graph $\mathcal{G}$

$\Rightarrow \exists$ two grids $\mathcal{L}_1$ and $\mathcal{L}_2$ (with the same dimension) s.t.

$$c_1 \bar{\mathcal{R}}^{\text{eff}}(\mathcal{L}_1) \leq J_{\text{LQ}}(P) \leq c_2 \bar{\mathcal{R}}^{\text{eff}}(\mathcal{L}_2)$$

c_1, c_2 depend only on the geometric parameters of $\mathcal{G}$ and on min and max non-zero entries of P.



Geometric graphs behave like grids (2)

$$c_1 \bar{\mathcal{R}}^{\text{eff}}(\mathcal{L}_1) \leq J_{\mathsf{LQ}}(P) \leq c_2 \bar{\mathcal{R}}^{\text{eff}}(\mathcal{L}_2)$$

- c_1, c_2 depend only on the geometric parameters of $\mathcal{G}$ and on min and max non-zero entries of P.
  Interesting case: c_1, c_2 indep. of N, size of $\mathcal{L}_1, \mathcal{L}_2 \asymp N$
  i.e., $\mathcal{G}$ roughly looks like d-dimensional grid
- Recall assumed P primitive and symm.
  Can generalize: reversible Markov chain + assumption on inv. meas. (stronger than ‘democratic’: all entries $\sim c/N$)
- restrictive assumptions, but easy to find suitable graphs and construct symm. P e.g. with Metropolis weights
- such examples show that grid’s performance is due to local interactions (bounded number of neighbours + bounded distances), not to symmetries

Conclusion

- Different performance indices for consensus algorithm
  [Garin, Schenato, book chapter, 2011]
- We study performance in large-scale 'geometric' graphs:
  - rigorous results for regular grids
    [Garin, Zampieri, SIAM J. Contr. and Opt. 2012]
  - simulations: random geom. graphs behave as grids
    [Carli, Garin, Zampieri, ITA Workshop'09]
  - a class of deterministic geometric graphs behave as grids
    [Lovisari, Zampieri, Annual Reviews in Control, 2012]
    [Lovisari, Garin, Zampieri, CDC'10 and submitted SICON]

`http://necs.inrialpes.fr/people/garin/publications`

`http://automatica.dei.unipd.it/people/lovisari/publications.html`