On graph sequences with finite-time consensus

Xin Jiang
Department of Industrial and Systems Engineering
Lehigh University

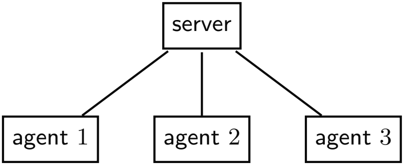
joint work with
Edward D. H. Nguyen (Rice), César A. Uribe (Rice), Bicheng Ying (Google)

2024 INFORMS Optimization Society Conference
March 24, 2024

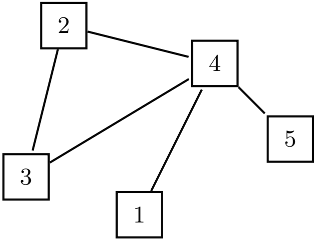
Distributed optimization

$$\text{minimize} \quad f(x) := \frac{1}{n}\sum_{i=1}^{n} f_i(x)$$

- **distributed** methods perform computation over a network (broader class)
- **decentralized** methods do so without central coordination (a subclass)



centralized setting



decentralized setting

Network topology in decentralized optimization

Classic assumptions on network topology

- static and defined beforehand, *e.g.*, network sensor localization
- dynamic/time-varying: bounded eigenvalues

$$\lambda_{\mathsf{min}} I \preceq W^{(k)} \preceq \lambda_{\mathsf{max}} I, \quad \text{for all iterations } k$$

- agents are *equidistant*

Modern scenarios (*e.g.*, high-performance computing (HPC), GPU)

- networks are flexible and cheaply rearranged
- networks are time-varying and might be disconnected
- agents are formed in clusters: intra-cluster communication is cheaper

This talk:

design new time-varying topologies with desirable properties

Decentralized average consensus

Mixing matrix $W \in \mathbb{R}^{n\times n}$ in decentralized optimization algorithms

- associated with a graph $G = (V, E)$: $W_{ij} = 0$ if $\{i, j\} \notin E$
- a round of communication is represented as matrix–vector product

$$(Wy)_i = \sum_{j=1}^{n} W_{ij} y_j = \sum_{j\in\mathcal{N}_i} W_{ij} y_j;$$

Decentralized average consensus

- suppose each agent $i \in V$ contains a vector $x_i \in \mathbb{R}^d$
- goal: to compute the average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ in a decentralized manner
- decentralized averaging with mixing matrix $W \in \mathbb{R}^{n\times n}$

$$X^{(k+1)} = WX^{(k)}, \quad \text{where } X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T \in \mathbb{R}^{n\times d}$$

- it converges asymptotically for all $X^{(0)}$ if and only if

$$W\mathbb{1} = \mathbb{1}, \qquad W^T\mathbb{1} = \mathbb{1}, \qquad 1 = |\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$$

Graph sequence with finite-time consensus property

the **finite-time consensus** property is defined for a given sequence of graphs

$$\{G^{(l)} \equiv (V, W^{(l)}, E^{(l)})\}_{l=0}^{\tau-1}$$

Consensus perspective: decentralized averaging converges in τ iterations

$$X^{(\tau)} = W^{(\tau-1)} W^{(\tau-2)} \dots W^{(1)} W^{(0)} X^{(0)} = \mathbb{1}\bar{x}^T$$

Matrix perspective: $\{W^{(l)}\}_{l=0}^{\tau-1} \subset \mathbb{R}^{n\times n}$ are doubly stochastic and

$$W^{(\tau-1)} W^{(\tau-2)} \dots W^{(1)} W^{(0)} = \frac{1}{n}\mathbb{1}\mathbb{1}^T =: J$$

Preview

we study three classes of graph sequences with finite-time consensus

graph sequence	size n	τ
one-peer exponential	$n = 2^\tau$	$\log_2 n$
p-peer hyper-cuboids	any $n \in \mathbb{N}_{\geq 2}$	# prime factors
SDS factor graphs	any $n \in \mathbb{N}_{\geq 2}$	flexible*

SDS: sequential doubly stochastic; *: τ is related to a partition $n = \sum_{k=1}^{\tau} n_k$

in the first two classes, we use the following convention to index $W \in \mathbb{R}^{n \times n}$

$$W = [w_{ij}], \quad i, j = 0, 1, \ldots, n-1$$

Outline

One-peer exponential graphs

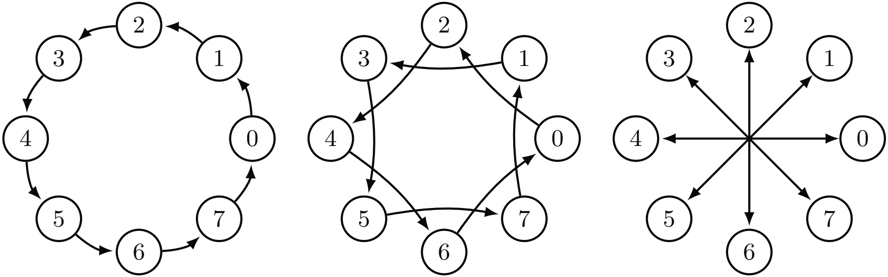
p-Peer hyper-cuboids

Hierarchical banded factor graphs

One-peer exponential graphs

- for $n \in \mathbb{N}_{\geq 2}$, define $\tau := \lfloor \log_2 n \rfloor$ and $\{G^{(l)}\}_{l=0}^{\tau-1}$ with weight matrices

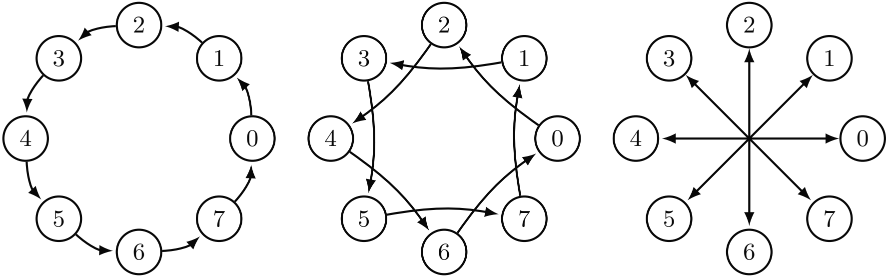
$$w_{ij}^{(l)} = \begin{cases} \frac{1}{2} & \text{if } \operatorname{mod}(j-i,n) = 2^{\operatorname{mod}(l,\tau)} \\ \frac{1}{2} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$



- if $n = 2^\tau$ for some $\tau \in \mathbb{N}_{\geq 1}$, then $\{W^{(l)}\}_{l=0}^{\tau-1}$ has finite-time consensus

[ALBR'19, YYC+'21, NJYU'23]

One-peer exponential graphs



$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \quad \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \quad \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

Detour: circulant matrix

- the $n \times n$ circulant matrix associated with $c = \big(c_0, c_1, \ldots, c_{n-1}\big)$ is

$$C = \mathrm{Circ}(c_0, c_1, \ldots, c_{n-1}) = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & & c_2 \\ \vdots & c_1 & c_0 & \ddots & \ddots \\ c_{n-2} & & \ddots & \ddots & c_{n-1} \\ c_{n-1} & c_{n-2} & \ddots & c_1 & c_0 \end{bmatrix}$$

- all circulant matrices share the same eigenvectors:

$$C = \big(\tfrac{1}{\sqrt{n}}F\big) \cdot \big(\mathrm{diag}(Fc)\big) \cdot \big(\tfrac{1}{\sqrt{n}}F^H\big),$$

  where F is the $n \times n$ DFT matrix

- the eigenvalues are complex numbers:

$$\lambda_i = c_0 + c_1\omega^i + c_2\omega^{2i} + \cdots + c_{n-1}\omega^{(n-1)i}, \quad i = 0, 1, \ldots, n-1,$$

  where $\omega = \exp\big(\frac{2\pi \hat{\jmath}}{n}\big)$ is a primitive n-th root of unity

Proof for finite-time consensus

- the mixing matrices of one-peer exponential graphs are circulant, and

$$W^{(\tau-1)} \cdots W^{(1)} W^{(0)} = \left(\tfrac{1}{\sqrt{n}} F\right) \cdot \left(\Lambda^{(\tau-1)} \cdots \Lambda^{(1)} \Lambda^{(0)}\right) \cdot \left(\tfrac{1}{\sqrt{n}} F^H\right),$$

  where $\Lambda^{(l)} = \mathrm{diag}(Fc^{(l)})$ and $c^{(l)}$ is the first column of $W^{(l)}$
- the first entry in $Fc^{(l)}$ is always 1 because $F_{1,:} = \mathbb{1}^T$
- it implies the first entry in $\Lambda := \Lambda^{(\tau-1)} \cdots \Lambda^{(1)} \Lambda^{(0)}$ is 1
- the other (diagonal) entries in Λ, Λ_{ii}, are

$$\begin{aligned}
&\frac{1}{2^\tau}\Big((1+\omega^{(n-1)(i)})(1+\omega^{(n-2)(i)})(1+\omega^{(n-4)(i)})\cdots(1+\omega^{(n-2^{\tau-1})(i)})\Big) \\
&= \frac{1}{2^\tau}\Big((1+\omega^{(-1)(i)})(1+\omega^{(-2)(i)})(1+\omega^{(-4)(i)})\cdots(1+\omega^{(-2^{\tau-1})(i)})\Big) \\
&= \frac{1}{2^\tau}\sum_{l=0}^{n-1}\omega^{-il} = \frac{1}{2^\tau}\left(\frac{1-\omega^{-in}}{1-\omega^{-i}}\right) = 0
\end{aligned}$$

Outline

One-peer exponential graphs

p-Peer hyper-cuboids

Hierarchical banded factor graphs

One-peer hyper-cube

- given $n = 2^\tau$ with some $\tau \in \mathbb{N}_{\geq 1}$, define

$$w_{ij}^{(l)} = \begin{cases} \frac{1}{2} & \text{if } (i \wedge j) = 2^{\mathrm{mod}(l,\tau)} \\ \frac{1}{2} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

  where $i \wedge j$ represents the bit-wise XOR operation between i and j

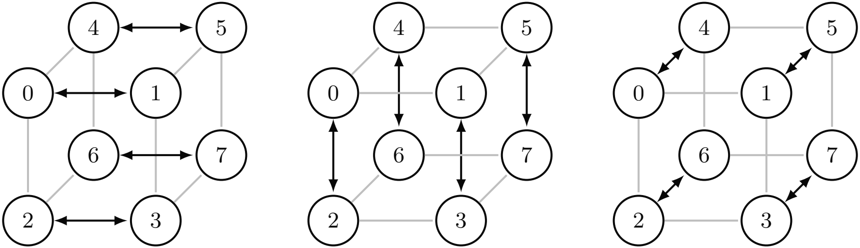
- represent i in its binary form $(i_{\tau-1} i_{\tau-2} \ldots i_0)_2$, and the first if-condition is

$$(i_{\tau-1} i_{\tau-2} \cdots i_0)_2 \wedge (j_{\tau-1} j_{\tau-2} \cdots j_0)_2 = (0 \cdots 0\, 1\, \underbrace{0 \, \cdots \, 0}_{\mathrm{mod}(l,\tau)})_2;$$

  only the $(\mathrm{mod}(l,\tau) + 1)$-th digit in i's and j's binary form is different

[SLJJ'16]

One-peer hyper-cube



$$\left[\begin{array}{cc|cc|cc|cc} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{array}\right] \quad \left[\begin{array}{cc|cc|cc|cc} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \hline \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \hline 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right] \quad \left[\begin{array}{cc|cc|cc|cc} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \hline 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \hline \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \hline 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{array}\right]$$

Multi-base representation of integers

- extension of one-peer hyper-cube to arbitrary matrix size n relies on:

  multi-base integer representation

- $(p_{\tau-1}, p_{\tau-2}, \ldots, p_0)$-based representation is an element in

$$\mathbb{N}_{p_{\tau-1}} \times \mathbb{N}_{p_{\tau-2}} \times \cdots \times \mathbb{N}_{p_0},$$

  where $\mathbb{N}_{p_j}$ is the group of nonnegative integers modulo $p_j \in \mathbb{N}_{\geq 2}$

- for example, $(2, 2, \ldots, 2)$-based representation is binary representation

- $(2, 3)$-based representation maps any integer in $\{0, 1, \ldots, 5\}$ to

$$0 \to \{0\}_2 \times \{0\}_3 \qquad 1 \to \{0\}_2 \times \{1\}_3 \qquad 2 \to \{0\}_2 \times \{2\}_3$$
$$3 \to \{1\}_2 \times \{0\}_3 \qquad 4 \to \{1\}_2 \times \{1\}_3 \qquad 5 \to \{1\}_2 \times \{2\}_3$$

- overload the notation as $(i_{p_{\tau-1}} \cdots i_{p_1} i_{p_0})_{p_{\tau-1},\ldots,p_1,p_0}$

p-Peer hyper-cuboid

- suppose the prime factorization of $n \in \mathbb{N}_{\geq 2}$ is $n = p_{\tau-1} \cdots p_1 p_0$; then

$$w_{ij}^{(l)} = \begin{cases} \frac{1}{p_{\mathrm{mod}(l,\tau)}} & \text{if } (i \wedge_{p_{\tau-1},\ldots,p_1,p_0} j) = (0, \cdots, 0, 1, \underbrace{0, \cdots, 0}_{\mathrm{mod}(l,\tau)})_{p_{\tau-1},\ldots,p_1,p_0} \\ \frac{1}{p_{\mathrm{mod}(l,\tau)}} & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

  where $i \wedge_{p_{\tau-1},\ldots,p_1,p_0} j$ denotes the bit-wise XOR operation between the $(p_{\tau-1}, \ldots, p_1, p_0)$-based representation of i and j

- *e.g.*, the prime factor set of $n = 12$ is $(p_2, p_1, p_0) = (2, 2, 3)$, with $\tau = 3$

- $i = 8$ and $j = 11$ are mapped in the $(2, 2, 3)$-based representation as

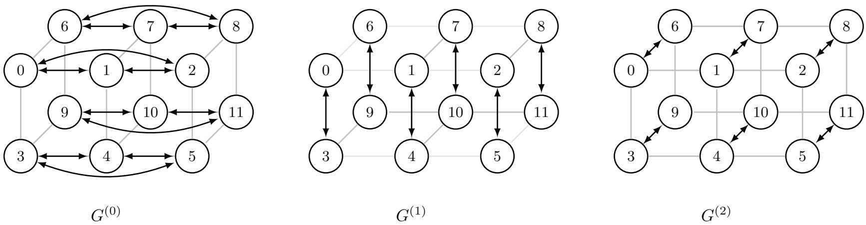
$$8 \to \{1\}_2 \times \{0\}_2 \times \{2\}_3, \qquad 11 \to \{1\}_2 \times \{1\}_2 \times \{2\}_3$$

- they differ only at the sub-group $\mathbb{N}_{p_1} = \mathbb{N}_2$

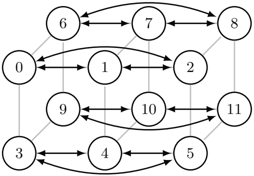
- when $l = 1$, agents $i = 8$ and $j = 11$ are connected with $w_{8,11}^{(1)} = \frac{1}{p_1} = \frac{1}{2}$

Example

$$(n,\tau) = (12,3),\ (p_2,p_1,p_0) = (2,2,3)$$



Example



$G^{(0)}$

$$\left[\begin{array}{ccc|ccc|ccc|ccc}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{array}\right]$$

p-Peer hyper-cuboid: Kronecker representation

p-peer hyper-cuboids of size $n = \prod_{k=0}^{\tau-1} p_k$ can be rewritten as

$$W^{(l)} = \widetilde{W}_{\tau-1}^{(l)} \otimes \cdots \otimes \widetilde{W}_1^{(l)} \otimes \widetilde{W}_0^{(l)},$$

where each $p_k \times p_k$ matrix $\widetilde{W}_k^{(l)}$ is defined by

$$\widetilde{W}_k^{(l)} = \begin{cases} I_{p_k} & \text{if } \mathrm{mod}(l, \tau) \neq k \\ \frac{1}{p_k} \mathbb{1}\mathbb{1}^T & \text{if } \mathrm{mod}(l, \tau) = k \end{cases}$$

Finite-time consensus

$$\begin{aligned} \prod_{l=0}^{\tau-1} W^{(l)} &= \prod_{l=0}^{\tau-1} \left(\widetilde{W}_{\tau-1}^{(l)} \otimes \widetilde{W}_{\tau-2}^{(l)} \otimes \cdots \otimes \widetilde{W}_0^{(l)} \right) \\ &\overset{\blacktriangle}{=} \left(\prod_{l=0}^{\tau-1} \widetilde{W}_{\tau-1}^{(l)} \right) \otimes \left(\prod_{l=0}^{\tau-1} \widetilde{W}_{\tau-2}^{(l)} \right) \otimes \cdots \otimes \left(\prod_{l=0}^{\tau-1} \widetilde{W}_0^{(l)} \right) \\ &= \left(\frac{1}{p_{\tau-1}} \mathbb{1}_{p_{\tau-1}} \mathbb{1}_{p_{\tau-1}}^T \right) \otimes \cdots \otimes \left(\frac{1}{p_0} \mathbb{1}_{p_0} \mathbb{1}_{p_0}^T \right) = \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T \end{aligned}$$

($\blacktriangle$) uses the property $(A \otimes B)(C \otimes D) = AC \otimes BD$

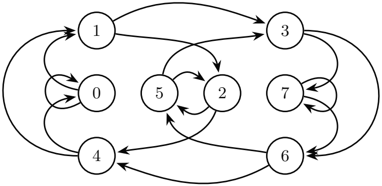
de Bruijn graphs

for $n = p^\tau$, the de Bruijn graph $G_{\mathsf{db}} = (V, W_{\mathsf{db}}, E_{\mathsf{db}})$ is defined by

$$w_{ij} = \begin{cases} \frac{1}{p} & \text{if } (i_{\tau-2} i_{\tau-3} \dots i_0)_p = (j_{\tau-1} j_{\tau-2} \dots j_1)_p \\ 0 & \text{otherwise,} \end{cases}$$

where $(i_{\tau-1} i_{\tau-2} \dots i_0)_p$ is the p-based representation of i

- example: $n = 8$, $p = 2$, $\tau = 3$



- connection between de Bruijn graphs and p-peer hyper-cuboids

$$W_{\mathsf{hc}}^{(l)} = P^{(l)} W_{\mathsf{db}} (Q^{(l)})^T \quad \text{for all } l = 0, 1, \dots, \tau - 1,$$

where $\{(P^{(l)}, Q^{(l)})\}$ are permutation matrices

[deBruijn'46, DCZ09]

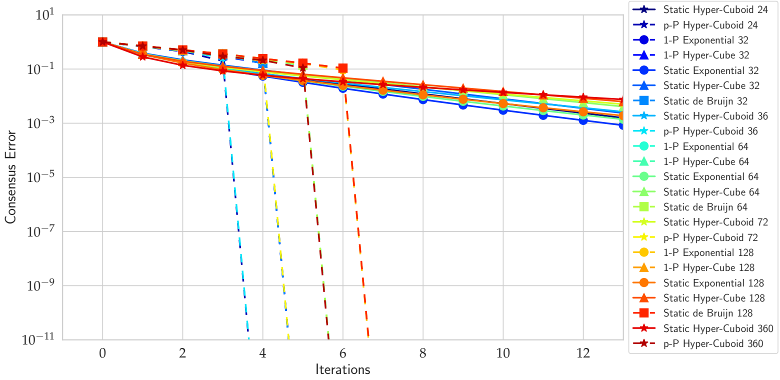
Numerical demonstration: decentralized average consensus

- decentralized average consensus iterations

$$x_i^{(k+1)} = W^{(k)} x_i^{(k)}, \quad \text{for } i = 1, \dots, n \text{ in parallel}$$

- we plot the consensus error

$$\Xi^{(k)} = \frac{1}{n} \sum_{i=1}^{n} \|x_i^{(k)} - x_{\text{avg}}^{(0)}\|_2^2$$



Outline

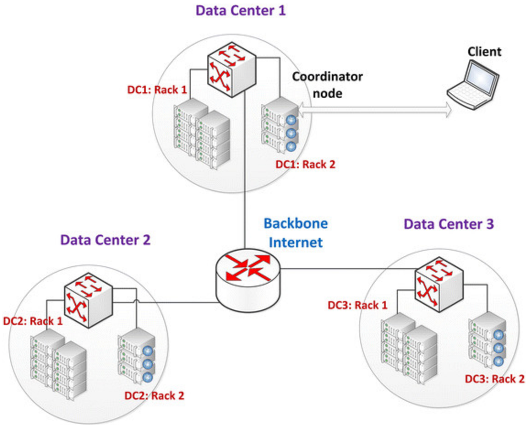
One-peer exponential graphs

p-Peer hyper-cuboids

Hierarchical banded factor graphs

Motivation

- p-peer hyper-cuboids revert to fully-connected graphs when n is *prime*
- data centers are not equidistant but formed in clusters
  - intra-cluster communication is cheap, flexible and can be varied
  - inter-cluster communication is expensive and should be minimized



Three-phase communication protocol

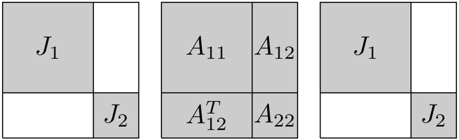
- phase 1: intra-cluster communication achieving finite-time consensus
- phase 2: limited inter-cluster communication
- phase 3: intra-cluster communication achieving finite-time consensus

we now focus on reducing the communication cost in phase 2

A two-block example

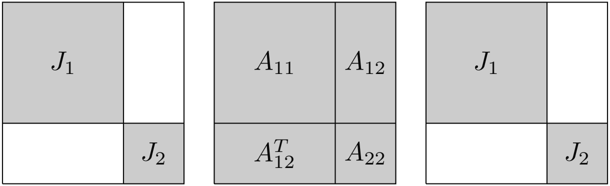
$$J = \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} = \begin{bmatrix} J_1 A_{11} J_1 & J_1 A_{12} J_2 \\ (J_1 A_{12} J_2)^T & J_2 A_{22} J_2 \end{bmatrix},$$

where $n = n_1 + n_2$ with $n_1 \geq n_2$, $J_1 = \frac{1}{n_1}\mathbb{1}_{n_1}\mathbb{1}_{n_1}^T$, and $J_2 = \frac{1}{n_2}\mathbb{1}_{n_2}\mathbb{1}_{n_2}^T$



A two-block example

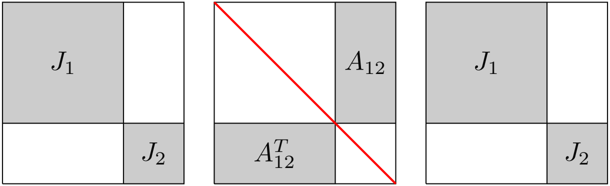
$$J = \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} = \begin{bmatrix} J_1 A_{11} J_1 & J_1 A_{12} J_2 \\ (J_1 A_{12} J_2)^T & J_2 A_{22} J_2 \end{bmatrix}$$



additional conditions can be imposed to increase the sparsity of A

A two-block example

$$J = \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} = \begin{bmatrix} J_1 A_{11} J_1 & J_1 A_{12} J_2 \\ (J_1 A_{12} J_2)^T & J_2 A_{22} J_2 \end{bmatrix}$$

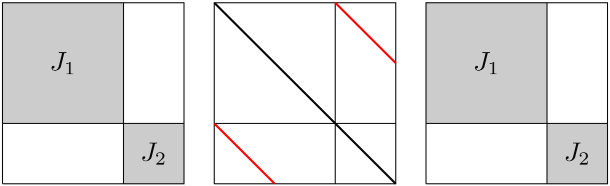


additional conditions can be imposed to increase the sparsity of A

- no intra-cluster communication: A_{11} and A_{22} are diagonal

A two-block example

$$J = \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} = \begin{bmatrix} J_1 A_{11} J_1 & J_1 A_{12} J_2 \\ (J_1 A_{12} J_2)^T & J_2 A_{22} J_2 \end{bmatrix}$$

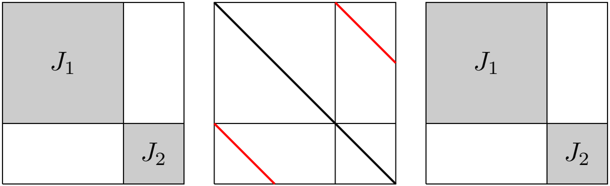


additional conditions can be imposed to increase the sparsity of A

- no intra-cluster communication: A_{11} and A_{22} are diagonal
- "one-to-one" inter-cluster communication

A two-block example

$$J = \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \begin{bmatrix} J_1 & \\ & J_2 \end{bmatrix} = \begin{bmatrix} J_1 A_{11} J_1 & J_1 A_{12} J_2 \\ (J_1 A_{12} J_2)^T & J_2 A_{22} J_2 \end{bmatrix}$$



additional conditions can be imposed to increase the sparsity of A

- no intra-cluster communication: A_{11} and A_{22} are diagonal
- nonzeros in A_{12} only appear on the diagonal and are the same

$$A = \left[\begin{array}{cc|c} \frac{n_2}{n} I_{n_2} & 0 & \frac{n_1}{n} I_{n_2} \\ 0 & I_{n_1 - n_2} & 0 \\ \hline \frac{n_1}{n} I_{n_2} & 0 & \frac{n_2}{n} I_{n_1} \end{array}\right]$$

Option 1: A_{12} is only nonzero in the first entry

$$A = \left[\begin{array}{ccccc|cccc} \alpha_1 & & & & & \beta & & & \\ & 1 & & & & & 0 & & \\ & & 1 & & & & & \ddots & \\ & & & \ddots & & & & & 0 \\ & & & & \ddots & & & & \\ & & & & 1 & & & & \\ \hline \beta & & & & & \alpha_2 & & & \\ & 0 & & & & & 1 & & \\ & & \ddots & & & & & \ddots & \\ & & & 0 & & & & & 1 \end{array}\right]$$

where

$$\alpha_1 = \frac{n_1^2}{n} - n_1 + 1, \qquad \alpha_2 = \frac{n_2^2}{n} - n_2 + 1, \qquad \beta = \frac{n_1 n_2}{n}$$

Option 2: the nonzero entries in A_{12} are the same

recall $n = n_1 + n_2$ and $n_1 \geq n_2$

$$A = \left[\begin{array}{cc|c} \frac{n_2}{n} I_{n_2} & 0 & \frac{n_1}{n} I_{n_2} \\ 0 & I_{n_1 - n_2} & 0 \\ \hline \frac{n_1}{n} I_{n_2} & 0 & \frac{n_2}{n} I_{n_1} \end{array}\right]$$

observe that this A is doubly stochastic

The general case

$$J = J_0 A J_0$$

- this factorization relies on a partition of $n \in \mathbb{N}_{\geq 2}$:

$$n = \sum_{k=1}^{\tau} n_k \quad \text{with } n_k \geq \sum_{j=k+1}^{\tau} n_j \text{ for all } k \in [\tau - 1]$$

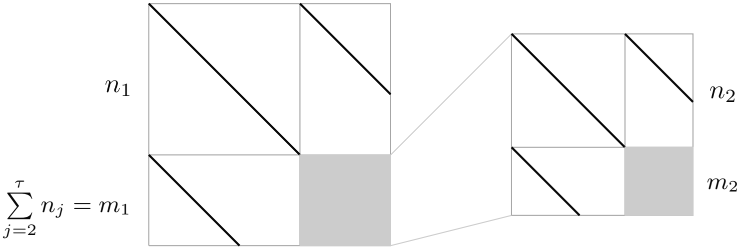
- $J_0 := J_1 \oplus \cdots \oplus J_\tau$ is block diagonal with $J_k := \frac{1}{n_k} \mathbb{1}\mathbb{1}^T \in \mathbb{R}^{n_k \times n_k}$

- $\oplus$ the direct sum of two matrices: $X \oplus Y = \mathrm{blkdiag}(X, Y)$

- each J_k can be further decomposed into, *e.g.*, p-peer hyper-cuboids

- we provide two options for the A-factor
  - A can be *hierarchically* partitioned as banded matrices
  - A can be decomposed as product of several banded matrices

Hierarchically banded (HB) factorization



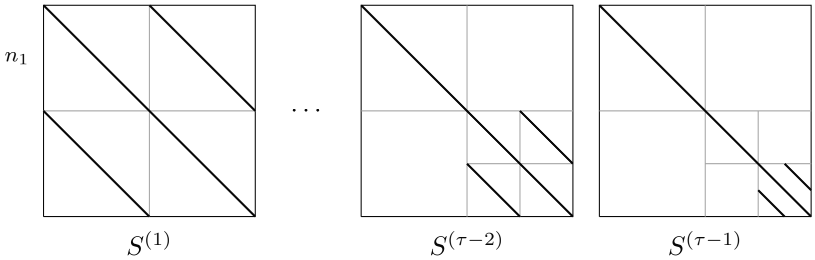
$$J = J_0 A J_0$$

- (density) reduced hierarchically banded (RHB) factorization
  - A_{RHB} has limited nonzeros in each band
- doubly stochastic hierarchically banded (DSHB) factorization
  - A_{DSHB} is symmetric, doubly stochastic, and hierarchically banded

Sequential doubly stochastic (SDS) factorization

$$J = J_0 A_{\mathsf{L}} J_0 \qquad \text{with } A_{\mathsf{L}} = S^{(1)} S^{(2)} \cdots S^{(\tau-1)}$$
$$J = J_0 A_{\mathsf{R}} J_0 \qquad \text{with } A_{\mathsf{R}} = S^{(\tau-1)} S^{(\tau-2)} \cdots S^{(1)}$$

where $\{S^{(k)}\} \subset \mathbb{S}^n$ are symmetric and doubly stochastic with banded pattern



Summary: graph sequences with finite-time consensus

- one-peer exponential graphs [ALBR'19, YYC+'21, NJYU'23]
  - $n = 2^\tau$, maximum degree is 1
  - they share the same eigenspace

- p-peer hyper-cuboids [NJYU'23]
  - any $n \in \mathbb{N}_{\geq 2}$, τ is the number of prime factors
  - maximum degree is the largest prime factor of n
  - includes one-peer hyper-cubes [SLJJ'16] as special cases

- sparse factorization of J of the form [JNUY'24]

$$J = J_0 A J_0, \quad \text{where } J_0 = J_1 \oplus \cdots \oplus J_\tau$$

  - (density) reduced hierarchically banded factorization: A_{RHB}
  - doubly stochastic hierarchically banded factorization: A_{DSHB}
  - sequential doubly stochastic (SDS) factorization: A_{L} and A_{R}

$$A_{\mathsf{L}} = S^{(1)} S^{(2)} \cdots S^{(\tau)}, \qquad A_{\mathsf{R}} = S^{(\tau)} S^{(\tau-1)} \cdots S^{(1)},$$

  where $\{S^{(k)}\} \subset \mathbb{S}^n$ are doubly stochastic with banded pattern

Summary

Graph sequences with finite-time consensus

topology	size n	max. deg.	τ
one-peer exponential	power of 2	1	$\log_2 n$
p-peer hyper-cuboid	arbitrary	largest prime factor	# of prime factors
one-peer hyper-cube	power of 2	1	$\log_2 n$
de Bruijn	power of p	p	$\log_p n$

Sparse factorization $J = J_0 A J_0$

matrices in phase 2	A_{RHB}	A_{DSHB}	A_{L}	A_{R}	S-factors
nnz	$n + \tau(\tau - 1)$	$\sum_{k=1}^{\tau} k n_k$	$\sum_{k=1}^{\tau} (2^k - 1) n_k$	$\sum_{k=1}^{\tau} (2^k - 1) n_k$	$n_k + 2 \sum_{i=k+1}^{\tau} n_i$
$d_{\max}$	τ	τ	τ	$2^{\tau-1}$	2
# iter in phase 2	1	1	1	1	$\tau - 1$

What is forthcoming

- introduce graph sequences with finite-time consensus (this talk)

- incorporate such graphs into existing decentralized algorithms
  (talk 2 by Edward D. H. Nguyen)

- design new decentralized algorithms that allow time-varying topologies
  (talk 3 by Bicheng Ying)

References

- [NJYU'23] On graphs with finite-time consensus and their use in gradient tracking, arXiv:2311.01317
- [JNUY'24] Sparse factorization of the square all-ones matrix of arbitrary order, arXiv:2401.14596