Distributed Computing

Nitin Vaidya

Distributed computing in Nature

What is distributed computing?

A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable.

Leslie Lamport

What is distributed computing?

Parallel computing versus *distributed* computing

Example:

To add N numbers where N very large
use 4 processors, each adding up N/4,
then add the 4 partial sums

Parallel or distributed ?

What is distributed computing?

- *Parallel* computing versus *distributed* computing
- Role of uncertainty in distributed systems
  - Clock drift
  - Network delays
  - Network losses
  - Asynchrony
  - Failures

Clocks

- Notion of *time* very useful in real life, and so it is in distributed systems

- Example …

    Submit programming assignment
    by e-mail by 11:59 pm Monday

Clocks

- Notion of *time* very useful in real life, and so it is in distributed systems

- Example …

Submit programming assignment
by e-mail by 11:59 pm Monday

By which clock ?

Clocks

- Notion of *time* very useful in real life, and so it is in distributed systems

- Example …

Submit programming assignment
by e-mail by 11:59 pm Monday

If it reaches at 12:01, how do we know it was sent by 11:59 pm?

How to synchronize clocks?

How to synchronize clocks?

Role of delay uncertainty

Ordering of Events

- If we can’t have “perfectly” synchronized clocks, can we still accurately determine *what happened first*?

Distribute Storage

How to improve system availability?

- Potentially large network delays … network partition

- Failures

Replication is a common approach

- If data stored only in one place, far away user will incur significant access delay

➔ Store data in multiple replicas,

Clients prefer to access “closest” replica

Replicated Storage

- How to keep replicas “consistent” ?
- What does “consistent” really mean?

Consistency Model

- Since shared memory may be accessed by different processes concurrently, we need to define how the updates are observed by the processes

- *Consistency model* captures these requirements

Consistency #1

Alice: My cat was hit by a car.

Alice: But luckily she is fine.

Bob: That’s great!

What should Calvin observe?

Consistency #1

Alice: My cat was hit by a car.

Alice: But luckily she is fine.

Bob: That's great!

What should Calvin observe?

Consistency #2

Alice: My cat was hit by a car.

Alice: But luckily she is fine.

Bob: That's terrible!

What should Calvin observe?

Agreement

- Where to meet for dinner?

Agreement with Failure

- Non-faulty nodes must agree

Agreement with
Crash Failure & Asynchrony

What if nodes misbehave?

- Crash failures are benign
- Other extreme … Byzantine failures

“Local” Algorithms

Consensus

- Each node has an input (scalar or vector)

Consensus

- Each node has an input (scalar or vector)

- Consensus: Output in *convex hull* of the inputs

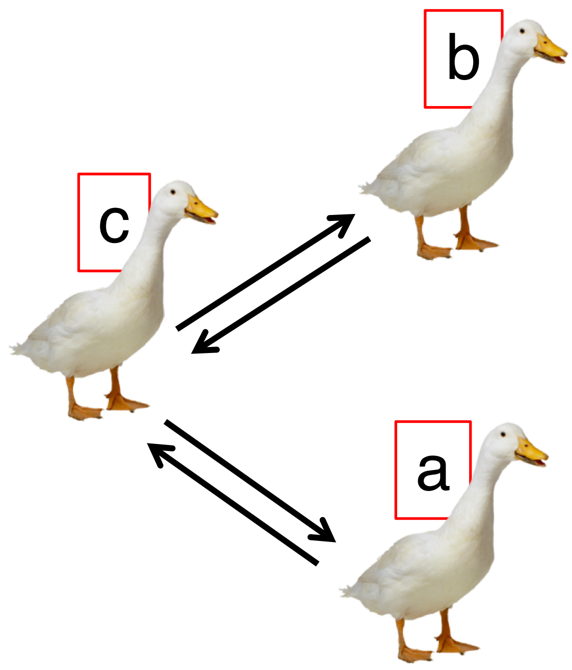
Consensus

convex
hull
Baseball Fields
GEORGETOWN
Georgetown University McDonough School of...
Georgetown Day School Lower/Middle School
Georgetown University
Foundry Branch Valley Park
Georgetown Cupcake
Canal Rd NW
Filomena Ristorante
Wawa
Baked & Wired
Trader Joe's
WEST END
Rock Creek and Potomac Pkwy NW
Three Sisters
Potomac River
George Washington Memorial Pkwy
Georgetown Waterfront Park
AMC Georgetown 14
George Washington University Hospital
Fort CF Smith Park
Fort Bennett Park and Palisades Trail
Key Bridge Marriott
N Spout Run Pkwy
The Mole
Watergate Complex Real Estate
Gateway Park
NORTH HIGHLAND
Custis Memorial Pkwy (Toll road)
DISTRICT OF COLUMBIA
VIRGINIA
McDonald's
John F. Kennedy Center for the Performing Arts
Pho 75
COLONIAL VILLAGE
Wilson Blvd
Clarendon Blvd
Easby Point
Tupelo Honey
16th Rd N
US Marine
Theodore Roosevelt Bridge
Albert Einstein Memoria

Consensus

Initially, state = input

$c = (a+b+c)/3$

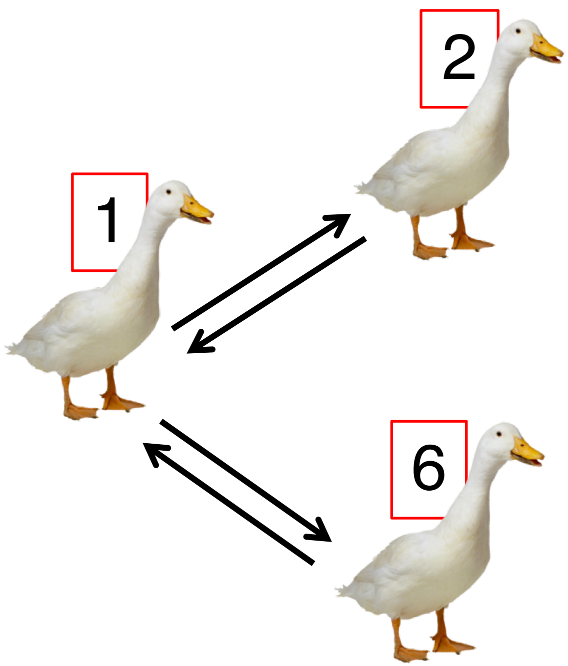


$b = (b+c)/2$

$a = (a+c)/2$

Consensus

$c = (1+2+6)/3 = 3$



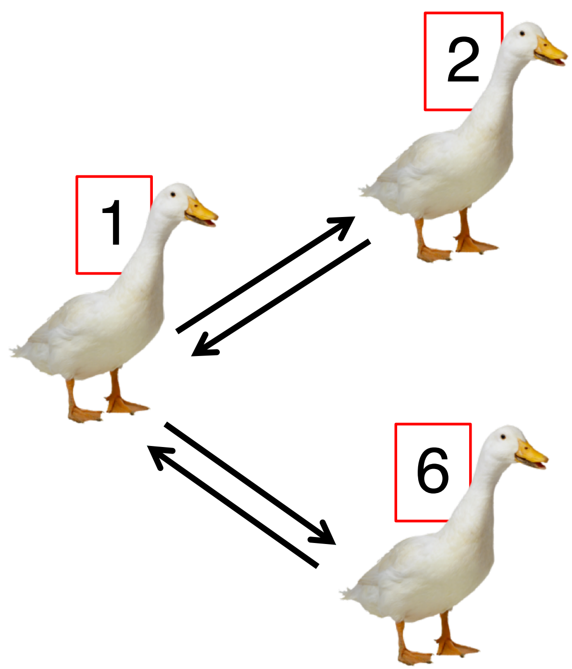
$b = (1+2)/2 = 3/2$

$a = (1+6)/2 = 7/2$

Consensus

As time → ∞, values become identical

$c = (1+2+6)/3 = 3$



$b = (1+2)/2 = 3/2$

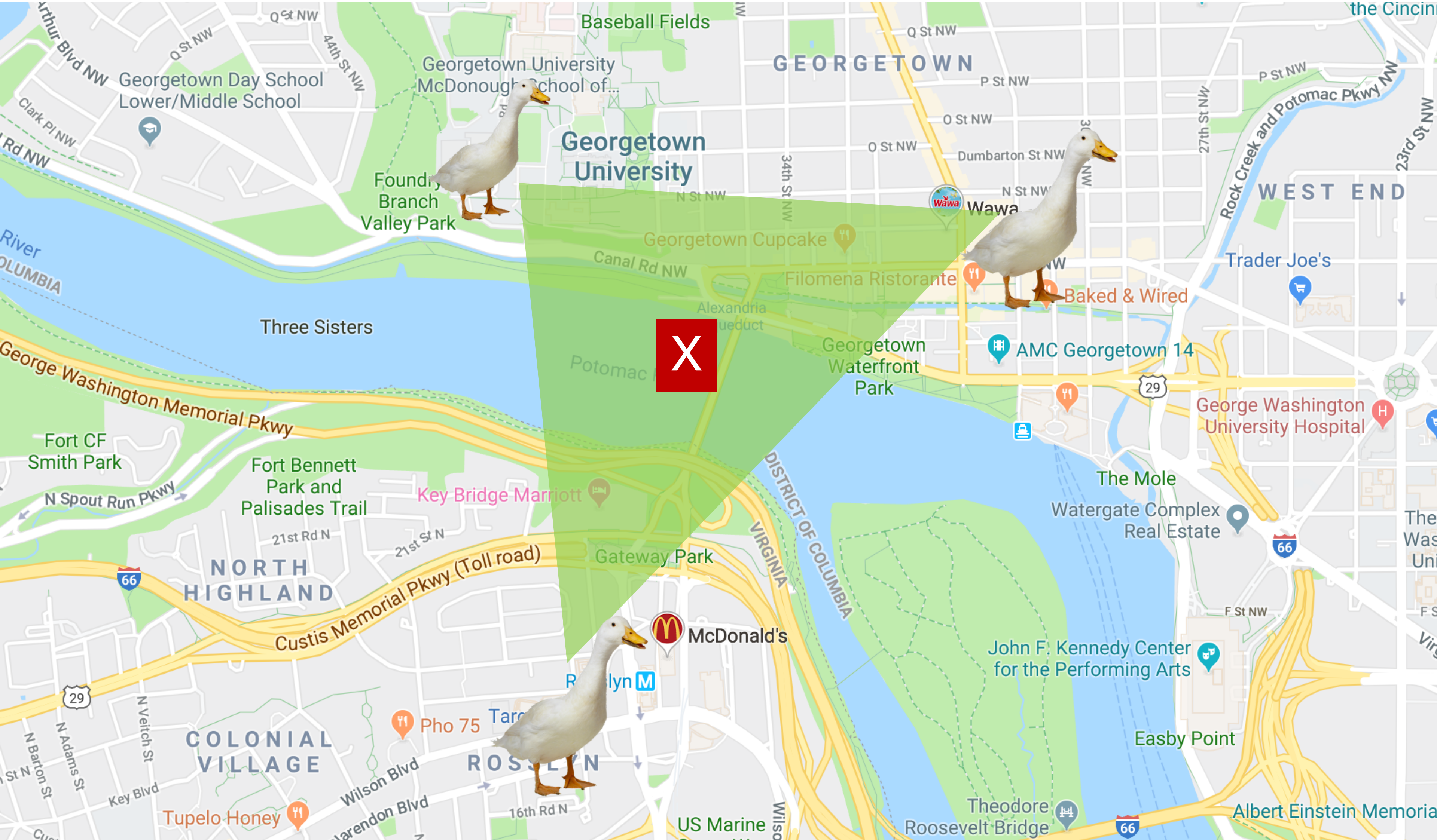
$a = (1+6)/2 = 7/2$

Average Consensus

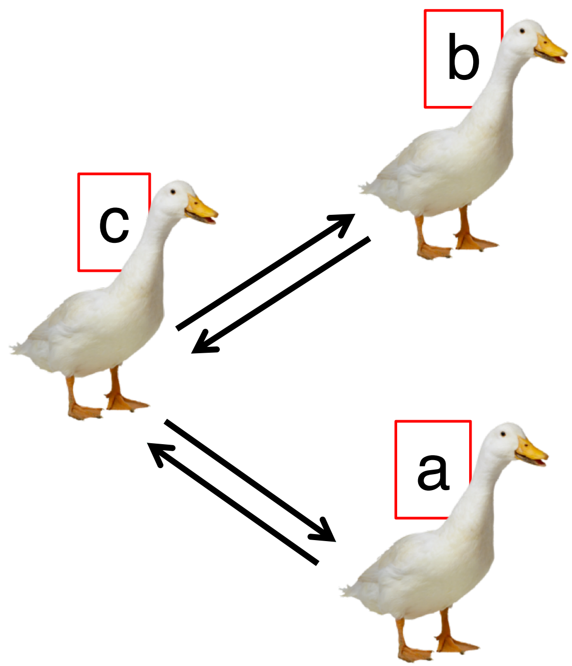
- Each node has an input (scalar or vector)

- Average consensus: Output = average of inputs

Average Consensus



Average Consensus
Change of Weights

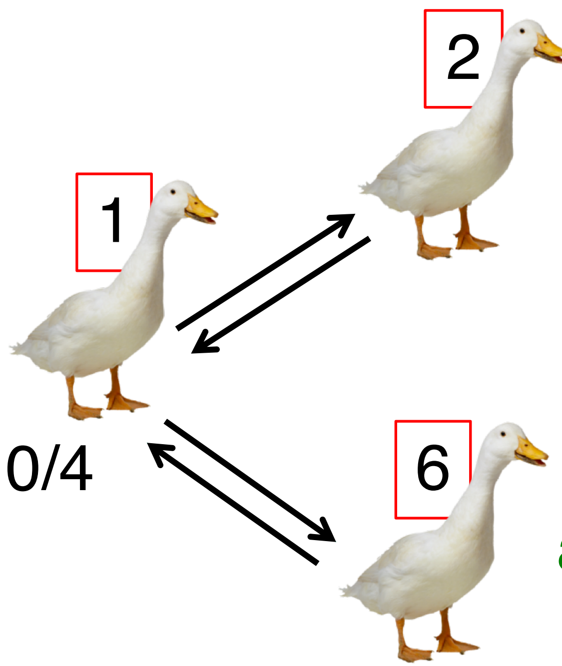


$b = 3b/4 + c/4$

$c = a/4 + b/4 + c/2$

$a = 3a/4 + c/4$

Average Consensus



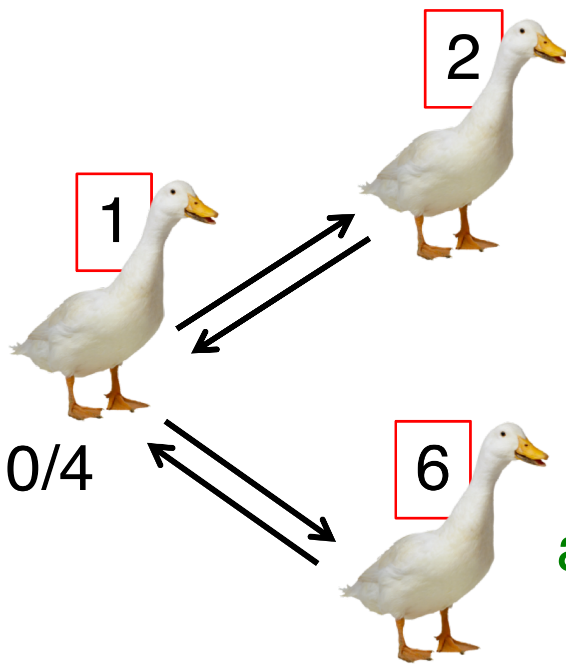
b = (6/4+1/4)=7/4

c = (2/4+6/4+1/2) = 10/4

a = (18/4+1/4) = 19/4

Average Consensus

As time → ∞, values converge to *average* of inputs



b = (6/4+1/4)=7/4

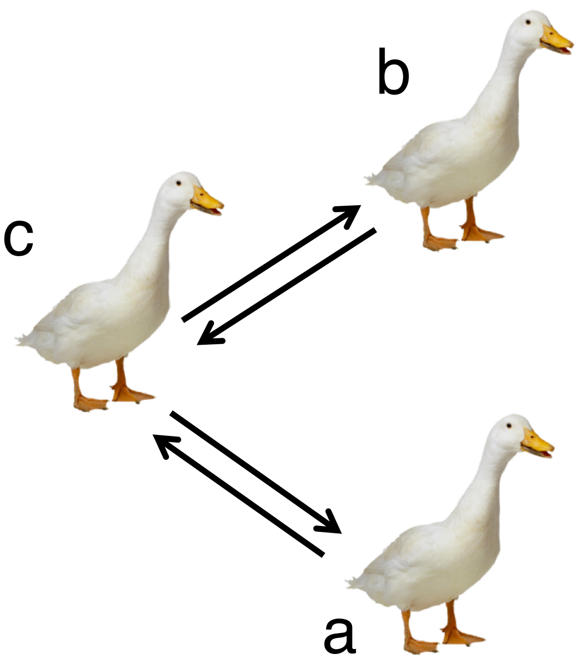
c = (2/4+6/4+1/2) = 10/4

a = (18/4+1/4) = 19/4

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := \underbrace{\begin{pmatrix} 3/4 & 0 & 1/4 \\ 0 & 3/4 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}}_{M} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = M \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

b = 3b/4+ c/4

c = a/4+b/4+c/2



a = 3a/4+ c/4

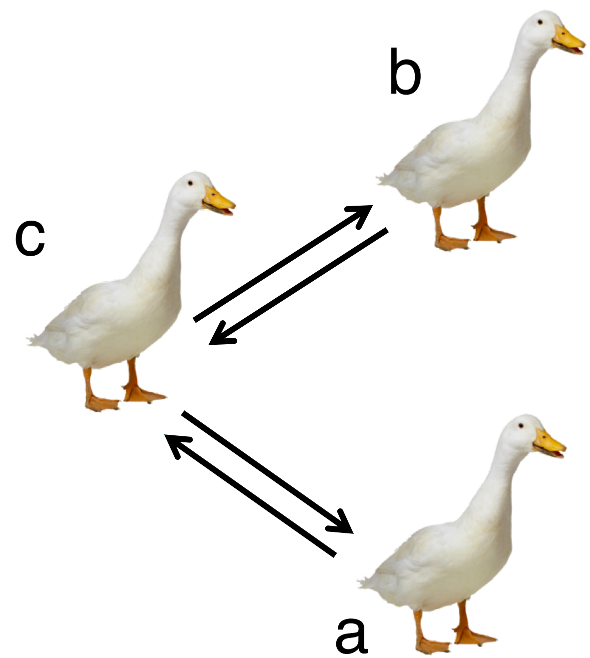
after 2 iterations

after 1 iteration

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := M \left[M \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right] = M^2 \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

b = 3b/4+ c/4

c = a/4+b/4+c/2



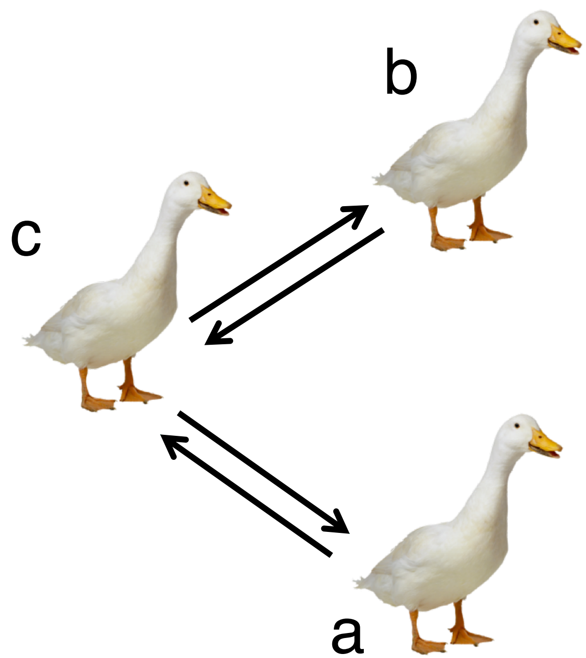
a = 3a/4+ c/4

after k iterations

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := M^k \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

b = 3b/4+ c/4

c = a/4+b/4+c/2



a = 3a/4+ c/4

Connected Undirected Graphs

after k iterations

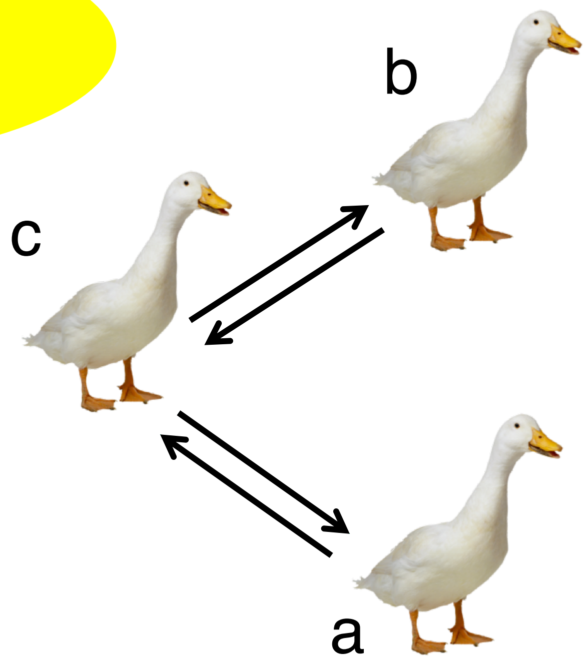
$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := M^k \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

- Consensus if M row stochastic
  - Matrix elements in [0,1]
  - M_{ij} non-zero if link (i,j) exists
  - Each row adds to 1

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := M^k \begin{pmatrix} a \\ b \\ c \end{pmatrix} \rightarrow \begin{pmatrix} p & q & r \\ p & q & r \\ p & q & r \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Row stochastic M

c = a/4+b/4+c/2



b = 3b/4+ c/4

a = 3a/4+ c/4

Row stochastic M

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := M \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Due to stochastic rows, each new state in convex hull of old states

Connected Undirected Graphs

after k iterations

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := M^k \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

- *Average* consensus

  if M doubly stochastic

  - Matrix elements in [0,1]
  - M_{ij} non-zero if link (i,j) exists
  - Each row & each column adds to 1

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := \begin{pmatrix} 3/4 & 0 & 1/4 \\ 0 & 3/4 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathrm{M} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

M

Due to stochastic rows,
each new state
in convex hull
of old states

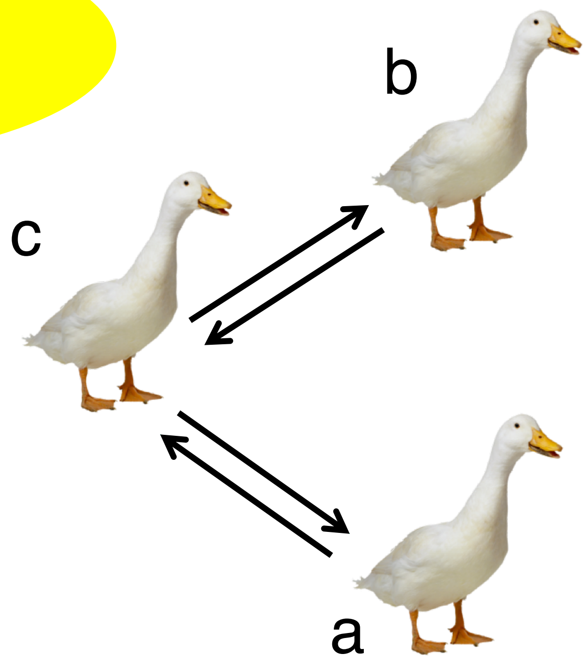
Due to stochastic columns,
total "mass"
(sum of states)
is preserved

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} := M^k \begin{pmatrix} a \\ b \\ c \end{pmatrix} \rightarrow \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Doubly stochastic M

$b = 3b/4 + c/4$

$c = a/4 + b/4 + c/2$

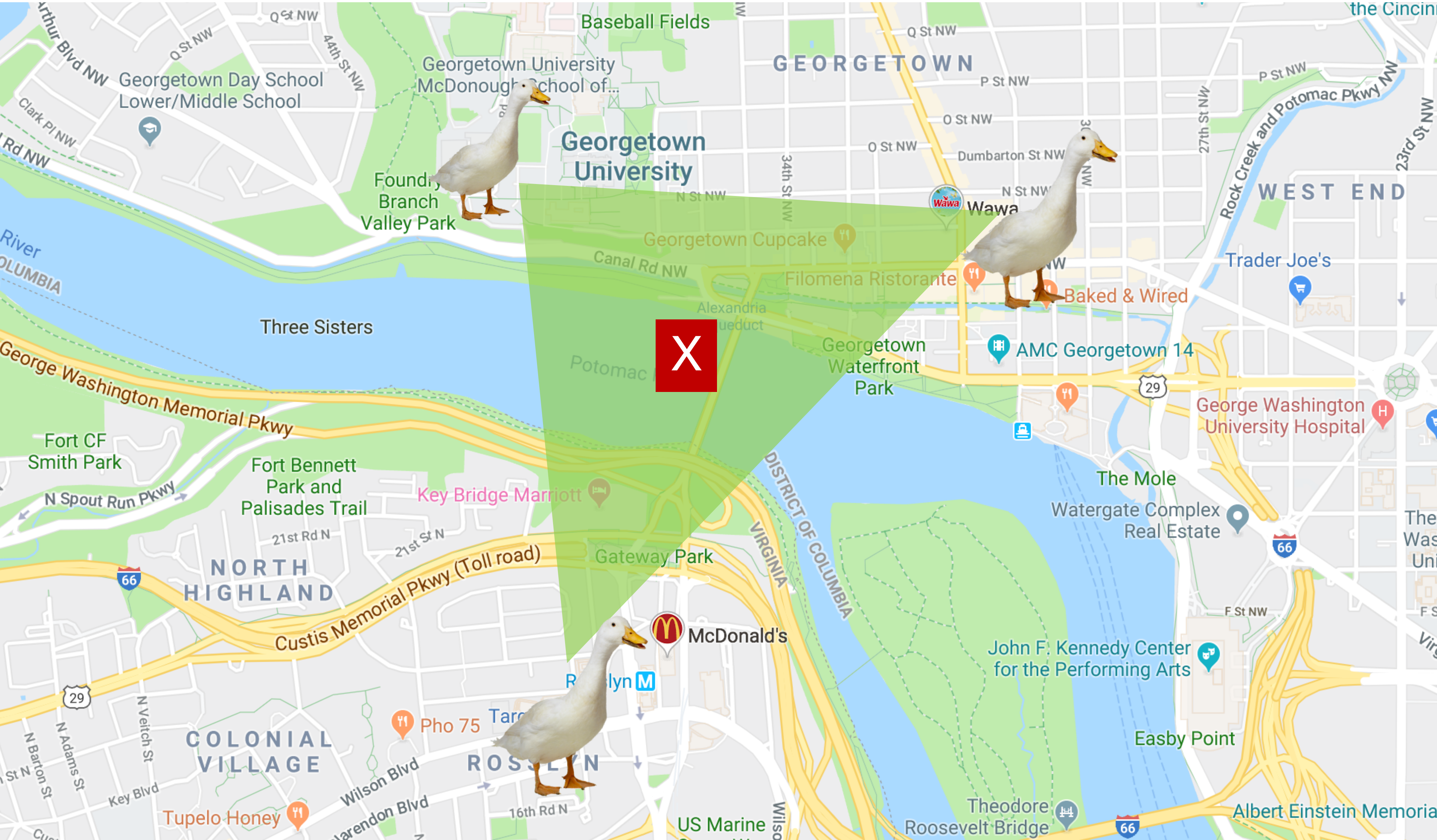


$a = 3a/4 + c/4$

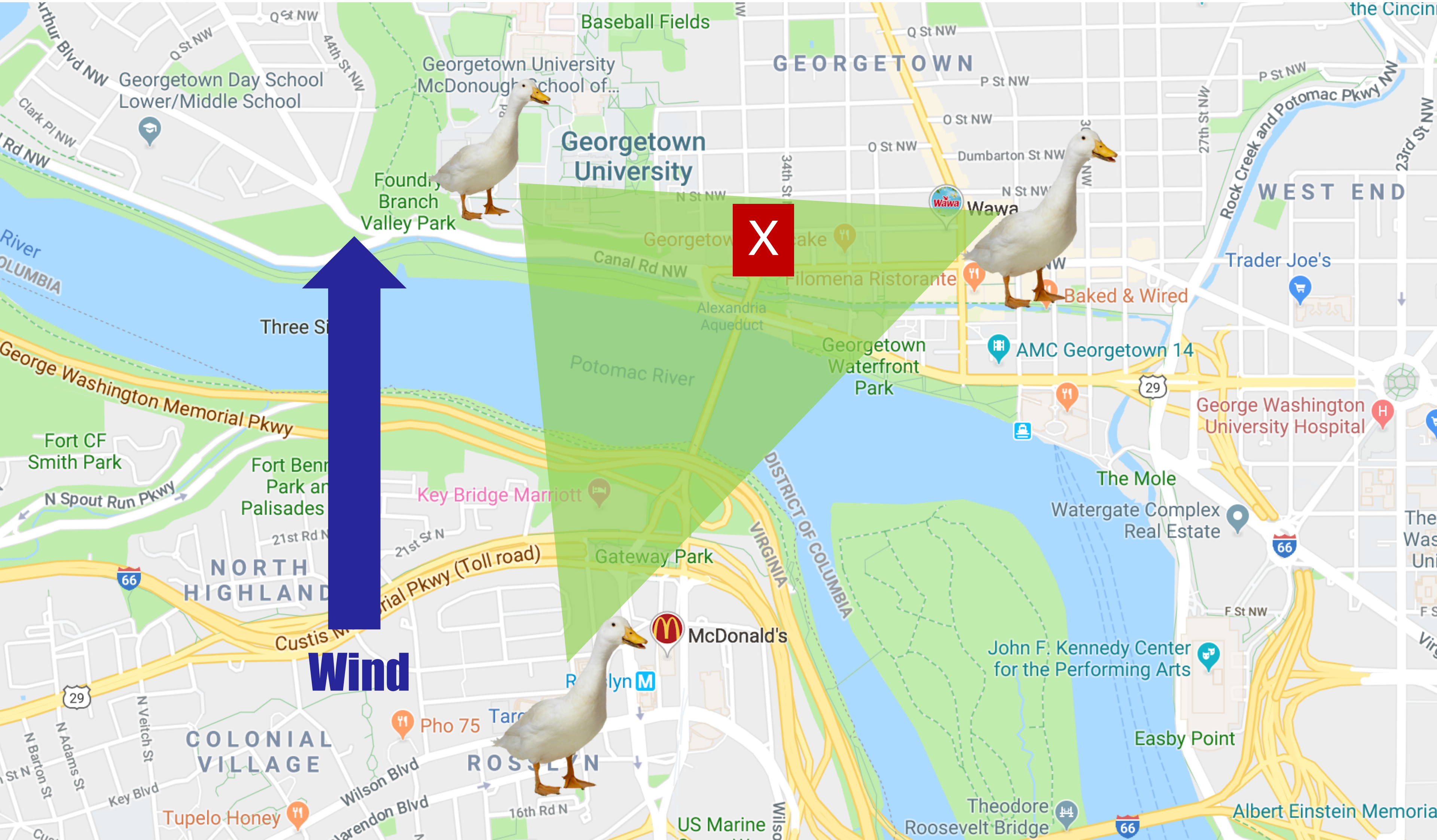
Consensus


convex
hull
Baseball Fields
GEORGETOWN
Georgetown University
McDonough School of...
Georgetown Day School
Lower/Middle School
Georgetown
University
Foundry
Branch
Valley Park
Georgetown Cupcake
Canal Rd NW
Filomena Ristorante
Wawa
Baked & Wired
Trader Joe's
WEST END
Rock Creek and Potomac Pkwy NW
Three Sisters
Potomac River
Georgetown
Waterfront
Park
AMC Georgetown 14
George Washington Memorial Pkwy
George Washington
University Hospital
Fort CF
Smith Park
Fort Bennett
Park and
Palisades Trail
Key Bridge Marriott
N Spout Run Pkwy
The Mole
Watergate Complex
Real Estate
NORTH
HIGHLAND
Custis Memorial Pkwy (Toll road)
Gateway Park
VIRGINIA
DISTRICT OF COLUMBIA
McDonald's
John F. Kennedy Center
for the Performing Arts
COLONIAL
VILLAGE
Pho 75
Wilson Blvd
Clarendon Blvd
Tupelo Honey
Easby Point
Theodore
Roosevelt Bridge
US Marine
Albert Einstein Memoria

Average Consensus



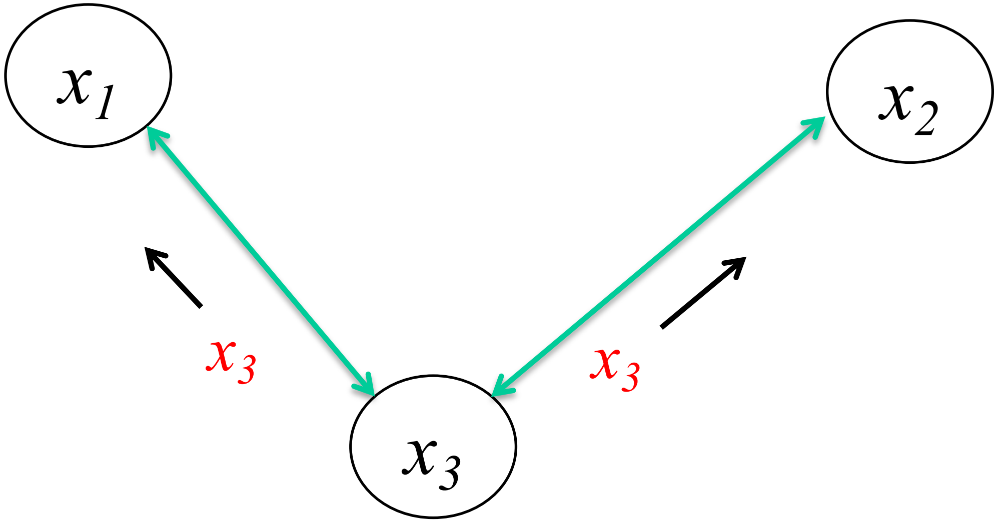
Optimization $\text{argmin} \sum f_i(x)$



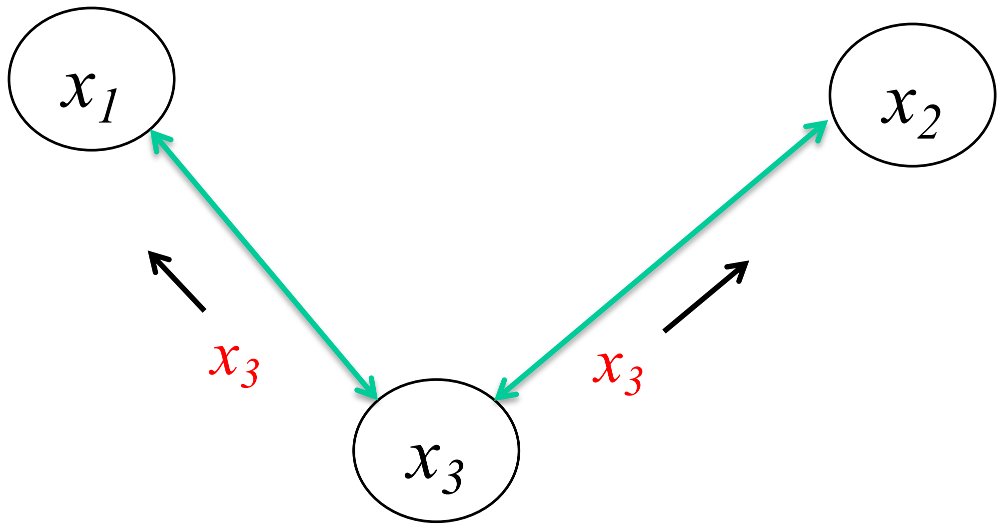
Distributed Optimization

$$f(x) = \sum f_i(x)$$

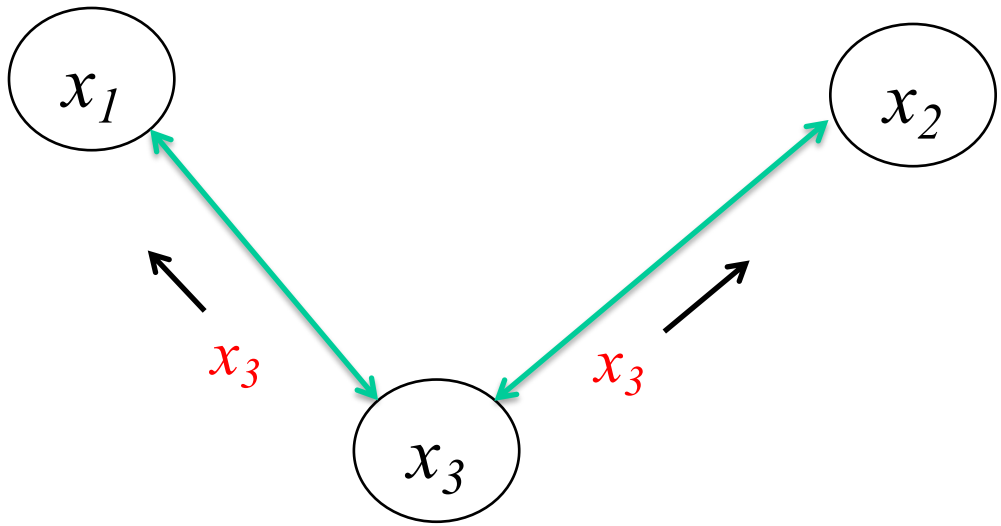
- Each agent maintains an estimate
- Local estimates shared with neighbors & updated
- Estimates converge to optimum



Example based on [Nedic and Ozdaglar, 2009]

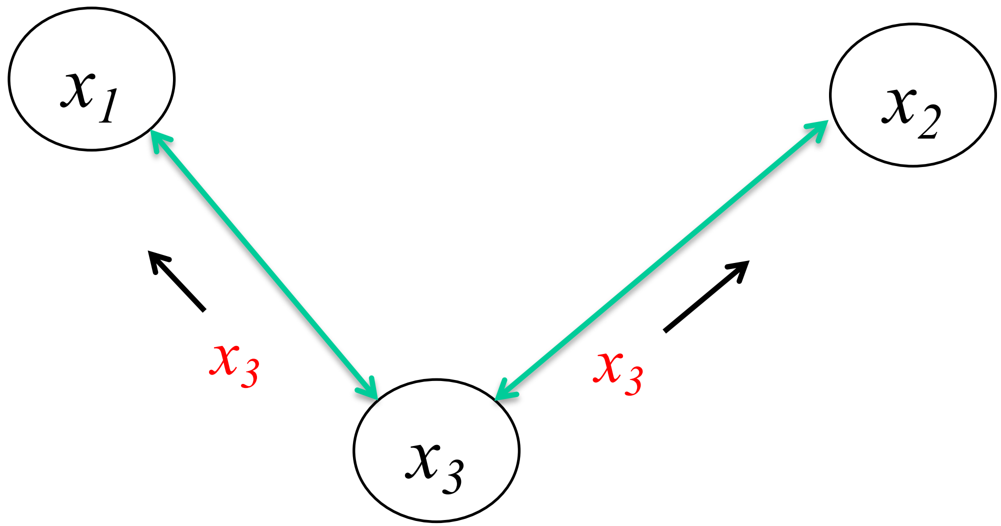


$$x_1 \longleftarrow \frac{2}{3}x_1 + \frac{1}{3}x_3 - \alpha \nabla f_1(x_1)$$



$$x_1 \leftarrow \frac{2}{3}x_1 + \frac{1}{3}x_3 - \alpha \nabla f_1(x_1)$$

$$x_3 \leftarrow \frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3 - \alpha \nabla f_3(x_3)$$



$$x_1[t+1] \longleftarrow \frac{2}{3}x_1[t] + \frac{1}{3}x_3[t] - \alpha \nabla f_1(x_1[t])$$

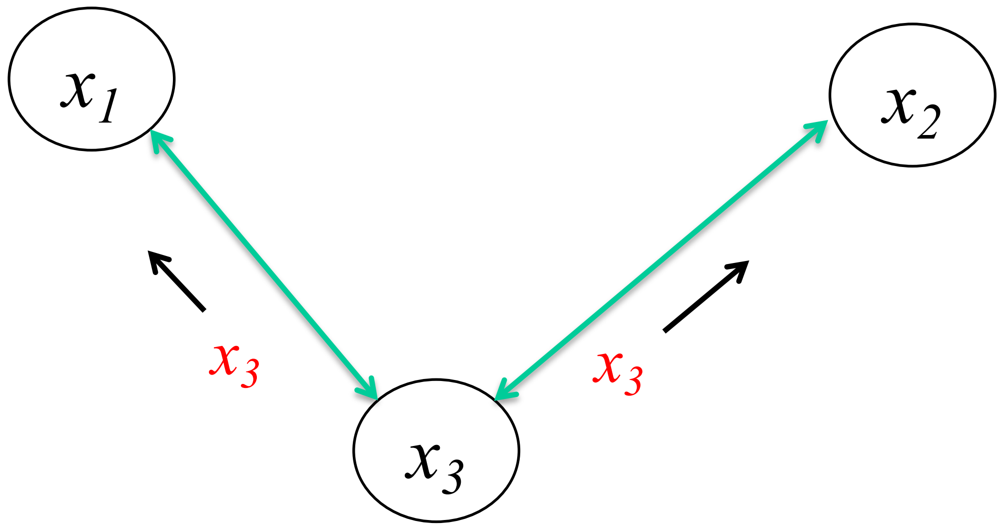
$$x_3[t+1] \longleftarrow \frac{1}{3}x_1[t] + \frac{1}{3}x_2[t] + \frac{1}{3}x_3[t] - \alpha \nabla f_3(x_3[t])$$

Distributed Optimization

As time → ∞

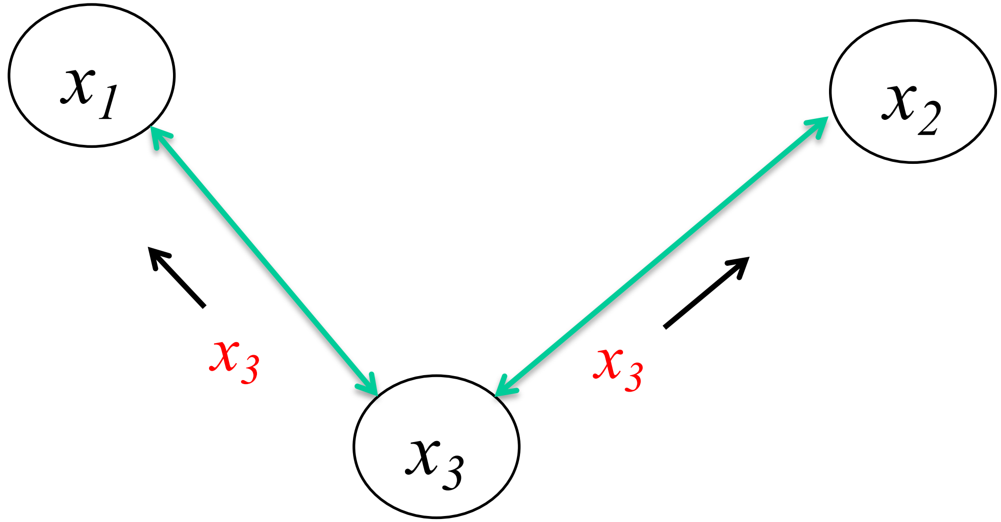
- Consensus: All agents converge to same estimate
- Optimality

$$\operatorname{argmin} \sum f_i(x)$$



$$x_1 \longleftarrow \frac{2}{3}x_1 + \frac{1}{3}x_3 - \alpha \nabla f_1(x_1)$$

$$x_3 \longleftarrow \frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3 - \alpha \nabla f_3(x_3)$$



$$x_1[t+1] \longleftarrow \frac{2}{3}x_1[t] + \frac{1}{3}x_3[t] - \alpha \nabla f_1(x_1[t])$$

$$x_3[t+1] \longleftarrow \frac{1}{3}x_1[t] + \frac{1}{3}x_2[t] + \frac{1}{3}x_3[t] - \alpha \nabla f_3(x_3[t])$$

$$\begin{bmatrix} x_3[t] \\ x_3[t] \\ x_3[t] \end{bmatrix} \qquad \begin{bmatrix} f_1(x_1[t]) \\ f_2(x_2[t]) \\ f_3(x_3[t]) \end{bmatrix}$$

$$x[t+1] \longleftarrow M\, x[t] - \alpha_t \nabla f(x[t])$$

$$x_1[t+1] \leftarrow \frac{2}{3}x_1[t] + \frac{1}{3}x_3[t] - \alpha_t \nabla f_1(x_1[t])$$

$$x_3[t+1] \leftarrow \frac{1}{3}x_1[t] + \frac{1}{3}x_2[t] + \frac{1}{3}x_3[t] - \alpha_t \nabla f_3(x_3[t])$$

$$x[t+1] \leftarrow M\,x[t] - \alpha_t \nabla f(x[t])$$

Doubly stochastic M

$$x[t+1] \longleftarrow M\, x[t] - \alpha_t \nabla f(x[t])$$

$$x[t+1] \longleftarrow \; M\,x[t] - \;\alpha_t\,\nabla f\,(x[t])$$

$$x[1] \longleftarrow \; M\,x[0] - \;\alpha_0\,\nabla f\,(x[0])$$

$$x[t+1] \longleftarrow M\,x[t] - \alpha_t \nabla f(x[t])$$

$$x[1] \longleftarrow M\,x[0] - \alpha_0 \nabla f(x[0])$$

$$x[2] \longleftarrow M\,x[1] - \alpha_1 \nabla f(x[1])$$

$$x[t+1] \longleftarrow \; M\,x[t] - \;\alpha_t\,\nabla f(x[t])$$

$$x[1] \longleftarrow \; M\,x[0] - \;\alpha_0\,\nabla f(x[0])$$

$$x[2] \longleftarrow \; M\,x[1] - \;\alpha_1\,\nabla f(x[1])$$

$$= M^2\,x[0] - \;\alpha_0 M\,\nabla f(x[0]) \; - \;\alpha_1\,\nabla f(x[1])$$

$$x[t+1] \longleftarrow \; M\, x[t] - \; \alpha_t \,\nabla f(x[t])$$

$$x[1] \longleftarrow \; M\, x[0] - \; \alpha_0 \,\nabla f(x[0])$$

$$x[2] \longleftarrow \; M\, x[1] - \; \alpha_1 \,\nabla f(x[1])$$

$$= M^2\, x[0] - \; \alpha_0 M \,\nabla f(x[0]) \; - \; \alpha_1 \,\nabla f(x[1])$$

$$x[3] \longleftarrow \; M\, x[2] - \; \alpha_2 \,\nabla f(x[2])$$

$$x[t+1] \longleftarrow M\, x[t] - \alpha_t \nabla f(x[t])$$

$$x[1] \longleftarrow M\, x[0] - \alpha_0 \nabla f(x[0])$$

$$x[2] \longleftarrow M\, x[1] - \alpha_1 \nabla f(x[1])$$

$$= M^2\, x[0] - \alpha_0 M \nabla f(x[0]) - \alpha_1 \nabla f(x[1])$$

$$x[3] \longleftarrow M\, x[2] - \alpha_2 \nabla f(x[2])$$

$$= M^3\, x[0]$$

$$- \alpha_0 M^2 \nabla f(x[0]) - \alpha_1 M \nabla f(x[1]) - \alpha_2 \nabla f(x[2])$$

$$x[t+1] \longleftarrow M\,x[t] - \alpha_t \nabla f(x[t])$$

$$x[1] \longleftarrow M\,x[0] - \alpha_0 \nabla f(x[0])$$

$$x[2] \longleftarrow M\,x[1] - \alpha_1 \nabla f(x[1])$$

$$= M^2\,x[0] - \alpha_0 M \nabla f(x[0]) - \alpha_1 \nabla f(x[1])$$

$$x[3] \longleftarrow M\,x[2] - \alpha_2 \nabla f(x[2])$$

$$= M^3\,x[0]$$

$$- \alpha_0 M^2 \nabla f(x[0]) - \alpha_1 M \nabla f(x[1]) - \alpha_2 \nabla f(x[2])$$

Claims

- Estimates at different nodes converge ➔ Consensus
- The estimates converges to $\text{argmin} \sum f_i(x)$