AI over networks: when decisions shall be taken together

Damiano Varagnolo

NTNU, Department of Engineering Cybernetics

Who am I?

Purposes of this seminar

- discuss about some technological problems and potential solutions
- connect with you

Purposes of this seminar

- discuss about some technological problems and potential solutions
- connect with you

today's cut: divulgation

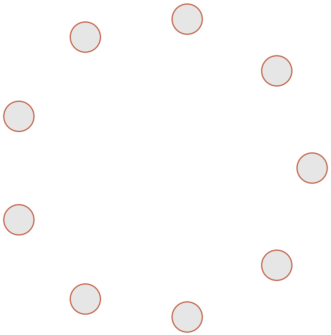
Roadmap

- centralized vs. distributed
- how shall we share information?
- consensus
- taking decisions over networks

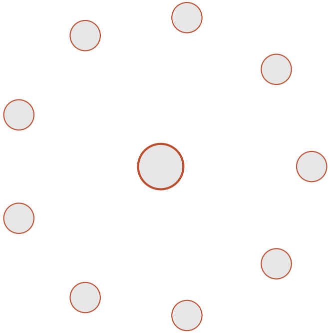
centralized vs. distributed

What does centralized mean?

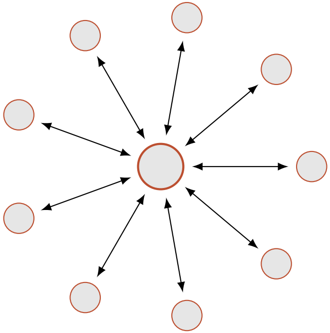
What does centralized mean?



What does centralized mean?

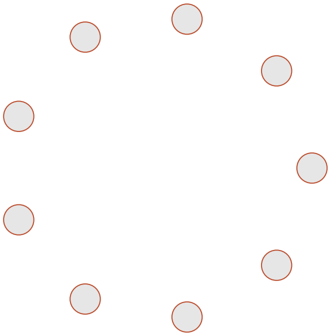


What does centralized mean?

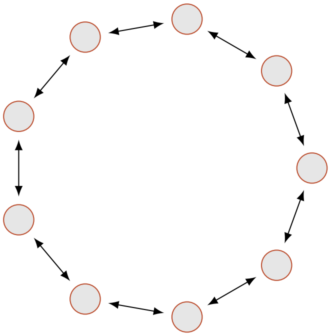


What does distributed mean?

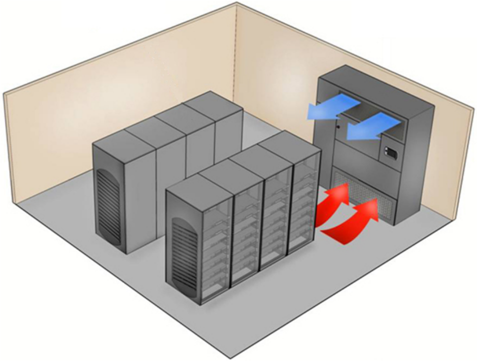
What does distributed mean?

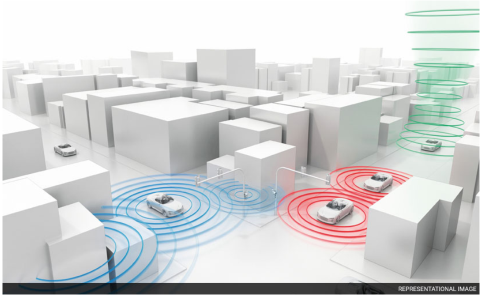


What does distributed mean?



Different paradigms for different applications





Summary of the pros and cons

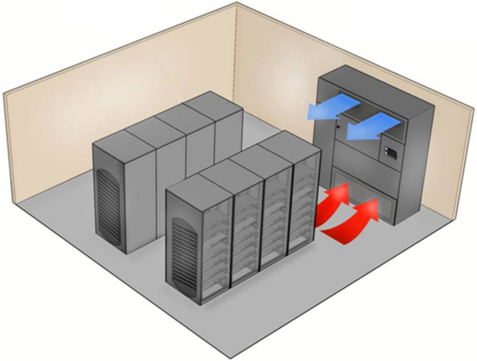
logical simplicity vs. practical feasibility

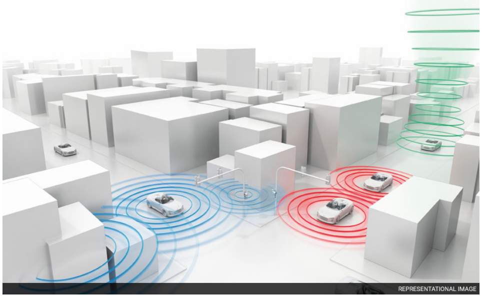
main problems: what to exchange, and how

how shall we share information?

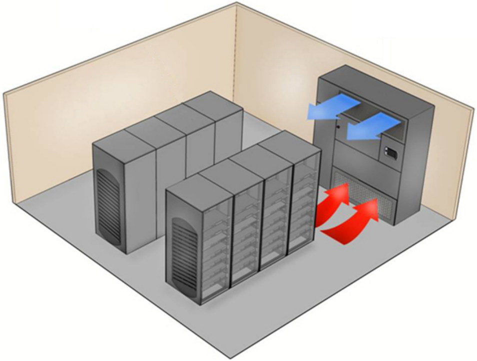
(in the distributed paradigm case)

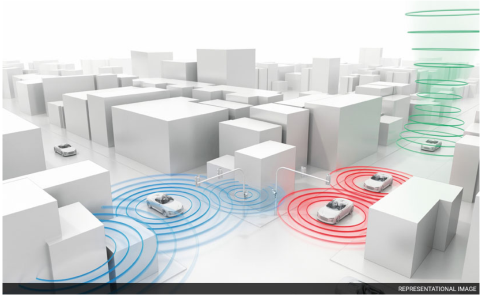
Fixed vs. dynamic topologies



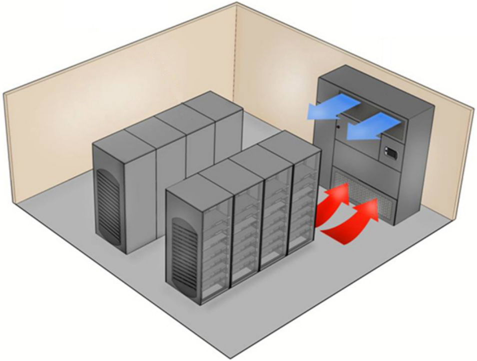


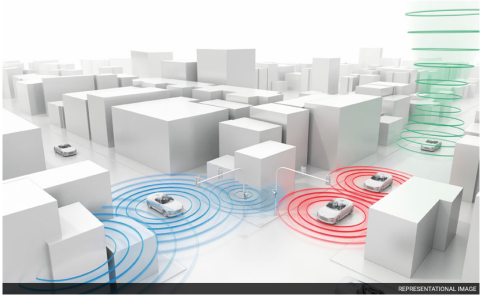
Wired vs. wireless communications



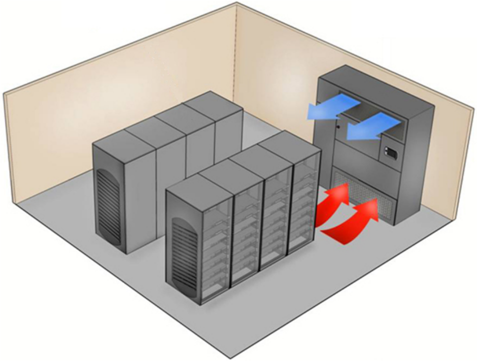


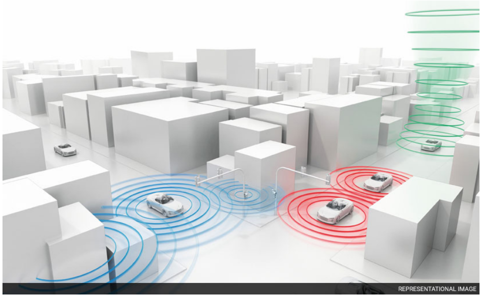
Synchronous vs. asynchronous communications





Lossless vs. lossy channels





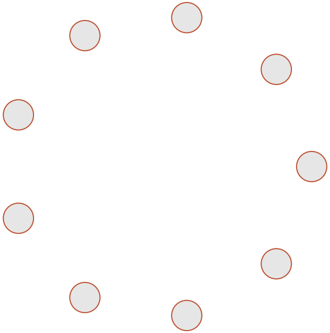
Summarizing, what are the main characteristics?

- do the connections change in time?
- what is the communication medium?
- is there a shared knowledge of time?
- may the communications fail?

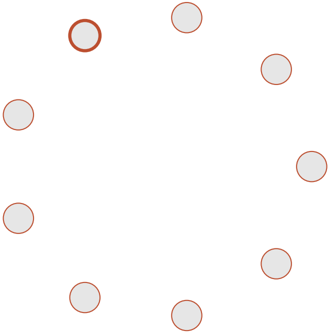
towards consensus:
the basic strategies for exchanging information

A fundamental strategy: gossip

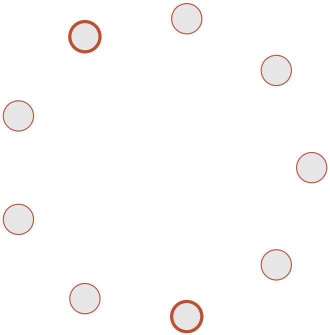
A fundamental strategy: gossip



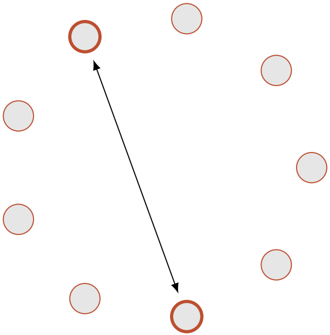
A fundamental strategy: gossip



A fundamental strategy: gossip

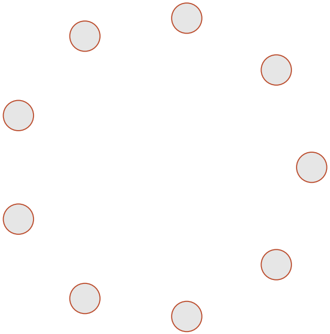


A fundamental strategy: gossip

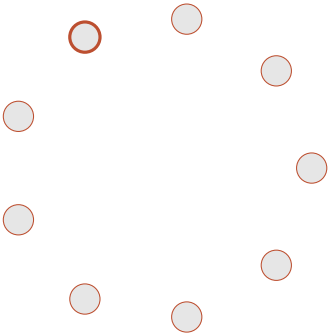


Byzantine generals, and their problems

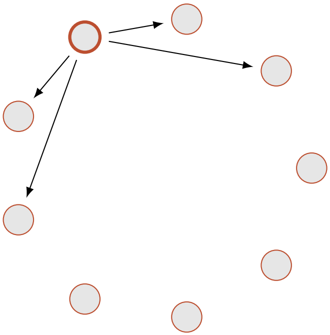
Broadcast communications



Broadcast communications



Broadcast communications



Summarizing, how can we be robust?

Summarizing, how can we be robust?

- broadcast
- asynchronous
- tolerating packet losses

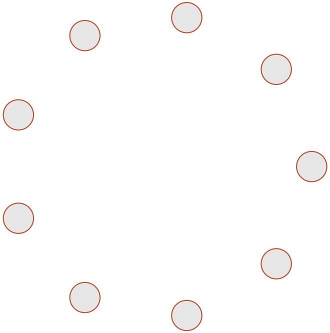
Summarizing, how can we be robust?

- broadcast
- asynchronous
- tolerating packet losses

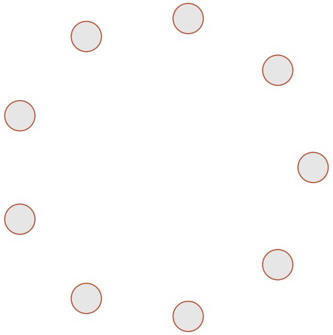
remember: there is no free lunch!

consensus

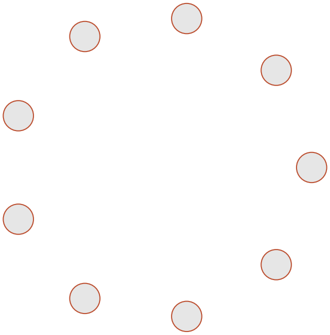
What does ***consensus*** mean?



Why is it important?



Why is it important?



example of a collective decision: *if in average colder than 18°, then turn the heaters on*

Average consensus

$$\begin{cases} \text{local state:} & \theta_i, \quad i = 1, \dots, n \\ \text{desired quantity:} & \dfrac{1}{n}\sum_{i=1}^{n} \theta_i \end{cases} \tag{1}$$

Max consensus

$$\begin{cases} \text{local state:} & \theta_i, \quad i = 1, \dots, n \\ \\ \text{desired quantity:} & \max\{\theta_1, \dots, \theta_n\} \end{cases} \tag{2}$$

Max consensus

$$\begin{cases} \text{local state:} & \theta_i, \quad i = 1, \ldots, n \\ \text{desired quantity:} & \max\{\theta_1, \ldots, \theta_n\} \end{cases} \tag{2}$$

example of a collective decision: *randomly select a leader*

Max consensus: how?

Order statistics consensus

$$\begin{cases} \text{local state:} & \theta_i, \quad i = 1, \ldots, n \\ \\ \text{desired quantities:} & \max\{\theta_1, \ldots, \theta_n\} \ \text{ and } \ \min\{\theta_1, \ldots, \theta_n\} \end{cases} \tag{3}$$

Order statistics consensus

$$\begin{cases} \text{local state:} & \theta_i, \quad i = 1, \ldots, n \\ \\ \text{desired quantities:} & \max\{\theta_1, \ldots, \theta_n\} \text{ and } \min\{\theta_1, \ldots, \theta_n\} \end{cases} \tag{3}$$

example of a collective decision: *rapid anonymous statistical counting*

Important remarks

max and order statistics consensus protocols can be broadcast, asynchronous, and using lossy media

Important remarks

max and order statistics consensus protocols can be broadcast, asynchronous, and using lossy media

. . . but what about average consensus?

small detour:

average consensus in practice

Gossip consensus

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \end{cases}$$

$$P(k) = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

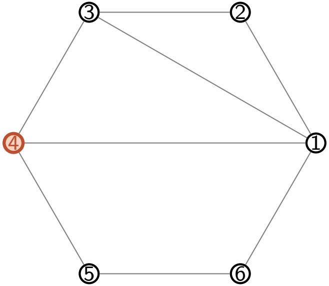
Ratio consensus, i.e., how to use broadcast communications

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

Ratio consensus, i.e., how to use broadcast communications

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

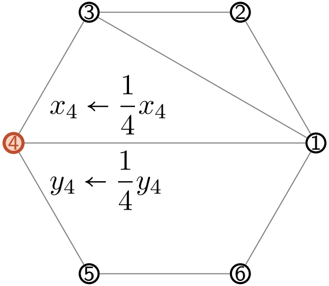
$$P(k) = \begin{bmatrix} 1 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Ratio consensus, i.e., how to use broadcast communications

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

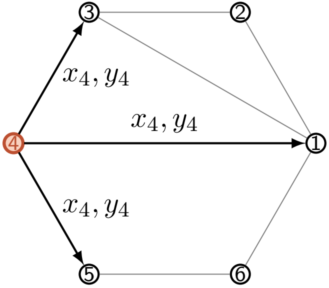
$$P(k) = \begin{bmatrix} 1 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Ratio consensus, i.e., how to use broadcast communications

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

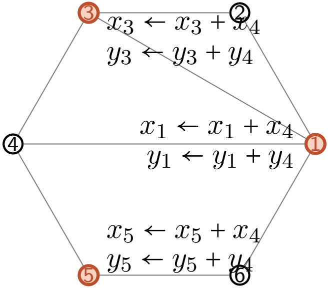
$$P(k) = \begin{bmatrix} 1 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Ratio consensus, i.e., how to use broadcast communications

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

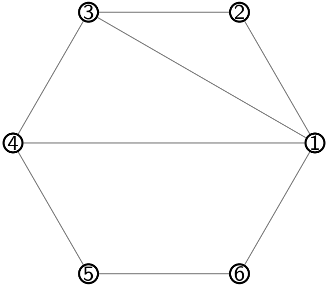
$$P(k) = \begin{bmatrix} 1 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Ratio consensus, i.e., how to use broadcast communications

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

$$P(k) = \begin{bmatrix} 1 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

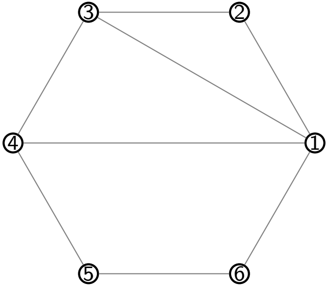


$$\begin{cases} x_i(k) \to \beta_i(k) \sum_j x_i(0) \\ \\ y_i(k) \to \beta_i(k) \sum_j y_i(0) \end{cases}$$

Ratio consensus, i.e., how to use broadcast communications

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

$$P(k) = \begin{bmatrix} 1 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



$$\begin{cases} x_i(k) \to \beta_i(k) \sum_j x_i(0) \\ \\ y_i(k) \to \beta_i(k) \sum_j y_i(0) \end{cases} \implies z_i(k) := \frac{x_i(k)}{y_i(k)} \to \frac{\sum_i x_i(0)}{\sum_i y_i(0)} = \frac{1}{N} \sum_i \theta_i$$

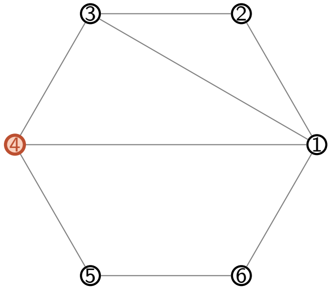
Robust ratio consensus, i.e., how to handle lossy channels

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

Robust ratio consensus, i.e., how to handle lossy channels

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

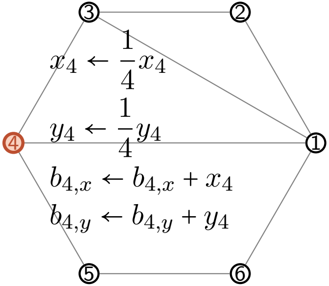
$$P(k) = \begin{bmatrix} 1 & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



Robust ratio consensus, i.e., how to handle lossy channels

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

$$P(k) = \begin{bmatrix} 1 & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

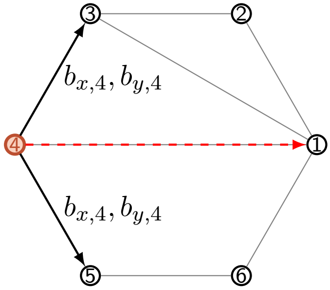


- $b_{i,x}$: total cumulative mass of x_i
- $\beta_{i,x}^{(j)}$: j's local estimate of $b_{i,x}$

Robust ratio consensus, i.e., how to handle lossy channels

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

$$P(k) = \begin{bmatrix} 1 & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

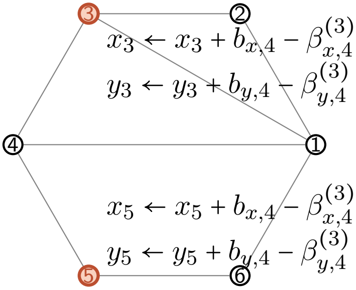


- $b_{i,x}$: total cumulative mass of x_i
- $\beta_{i,x}^{(j)}$: j's local estimate of $b_{i,x}$

Robust ratio consensus, i.e., how to handle lossy channels

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

$$P(k) = \begin{bmatrix} 1 & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

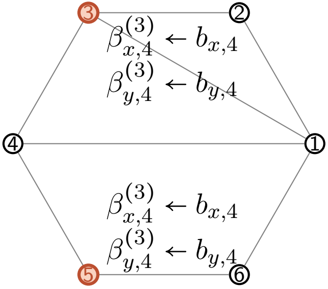


- $b_{i,x}$: total cumulative mass of x_i
- $\beta^{(j)}_{i,x}$: j's local estimate of $b_{i,x}$

Robust ratio consensus, i.e., how to handle lossy channels

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

$$P(k) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

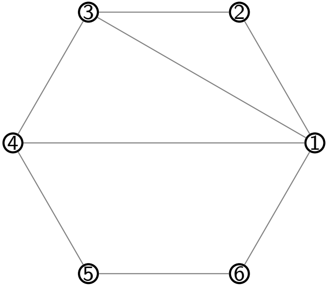


- $b_{i,x}$: total cumulative mass of x_i
- $\beta^{(j)}_{i,x}$: j's local estimate of $b_{i,x}$

Robust ratio consensus, i.e., how to handle lossy channels

$$\begin{cases} \boldsymbol{x}(k+1) = P(k)\boldsymbol{x}(k) \\ x_i(0) = \theta_i \\ \\ \boldsymbol{y}(k+1) = P(k)\boldsymbol{y}(k) \\ y_i(0) = 1 \end{cases}$$

$$P(k) = \begin{bmatrix} 1 & 0 & 0 & \mathbf{0} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



$$z_i(k) = \frac{x_i(k)}{y_i(k)} \to \frac{1}{N}\sum_j \theta_i$$

- $b_{i,x}$: total cumulative mass of x_i
- $\beta_{i,x}^{(j)}$: j's local estimate of $b_{i,x}$

taking decisions over networks

Numerical optimization: a hidden technology

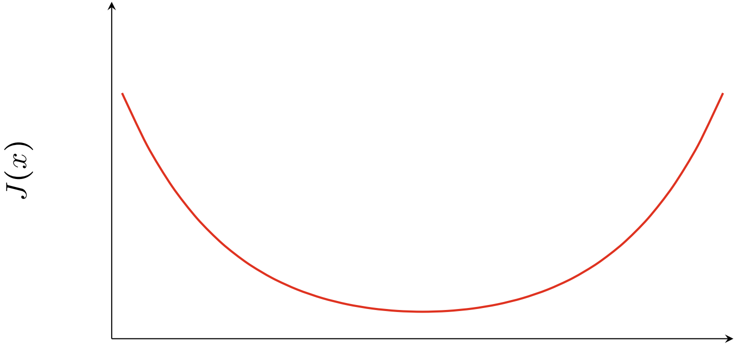
$$
\begin{aligned}
x^* = \quad & \arg\min_{x \in \mathcal{X}} J(x) \\
\text{s.t.} \quad & Ax = b \\
& g(x) \geq 0
\end{aligned} \tag{4}
$$

Numerical optimization: a hidden technology

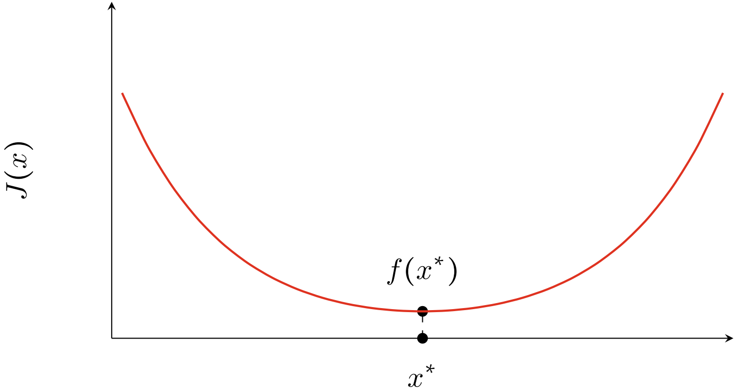
$$
\begin{aligned}
x^* = \quad & \arg\min_{x \in \mathcal{X}} J(x) \\
\text{s.t.} \quad & Ax = b \\
& g(x) \geq 0
\end{aligned}
\tag{4}
$$

every decision is an optimization!

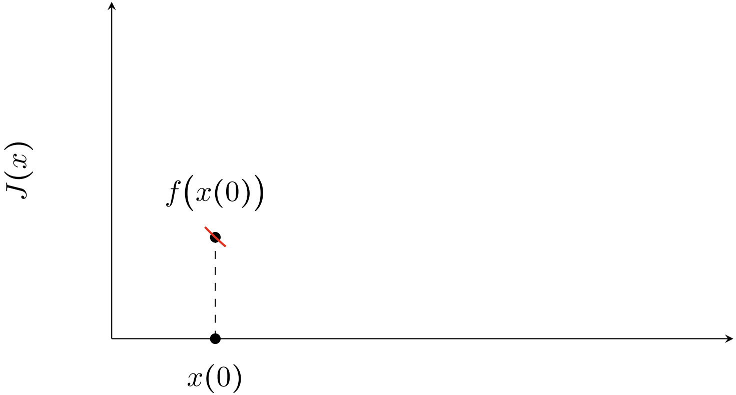
Centralized Newton Raphson



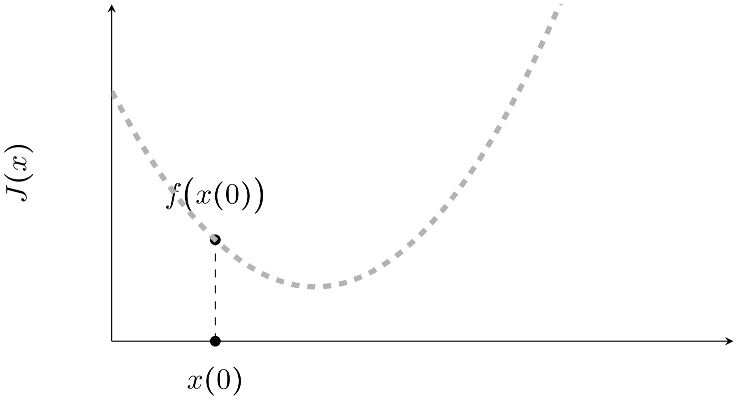
Centralized Newton Raphson



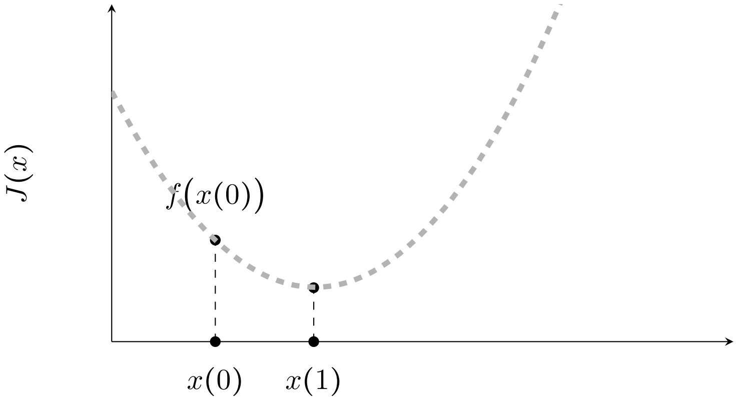
Centralized Newton Raphson



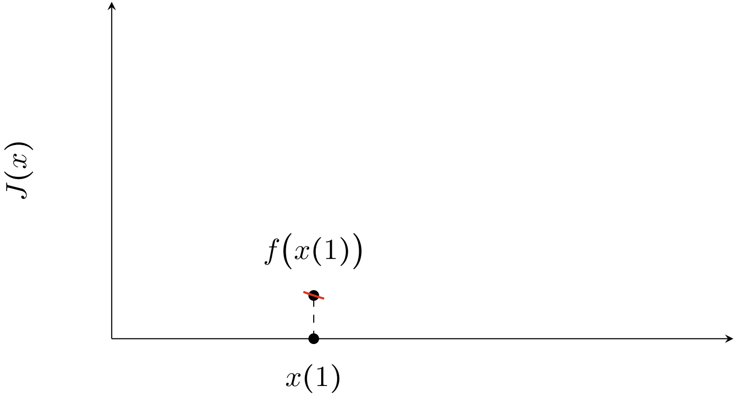
Centralized Newton Raphson



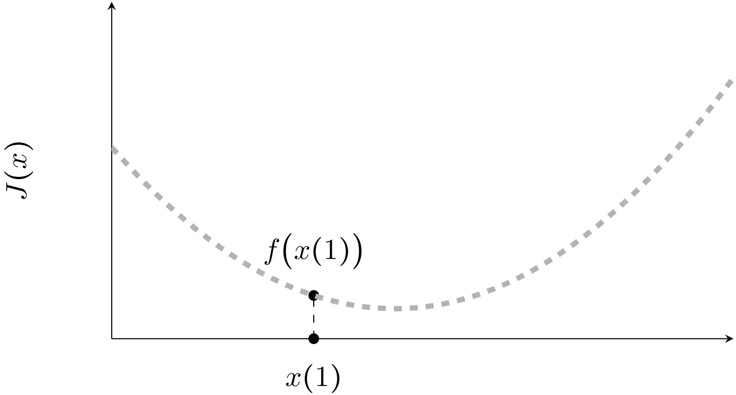
Centralized Newton Raphson



Centralized Newton Raphson



Centralized Newton Raphson



From the centralized NR to its distributed version

Newton update:

$$x^+ = x - \frac{f'(x)}{f''(x)}$$

thus

$$f(x) = \sum_{i=1}^{N} f_i(x) \implies x^+ = x - \frac{\sum_{i=1}^{N} f_i'(x)}{\sum_{i=1}^{N} f_i''(x)}$$

From the centralized NR to its distributed version

Newton update:

$$x^+ = x - \frac{f'(x)}{f''(x)}$$

thus

$$f(x) = \sum_{i=1}^{N} f_i(x) \implies x^+ = x - \frac{\sum_{i=1}^{N} f_i'(x)}{\sum_{i=1}^{N} f_i''(x)} = \frac{\sum_{i=1}^{N} \Big(f_i''(x)x - f_i'(x) \Big)}{\sum_{i=1}^{N} f_i''(x)}$$

From the centralized NR to its distributed version

Newton update:

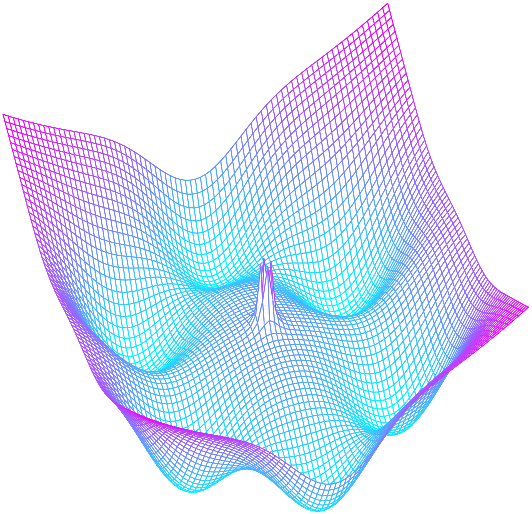
$$x^+ = x - \frac{f'(x)}{f''(x)}$$

thus

$$f(x) = \sum_{i=1}^{N} f_i(x) \implies x^+ = x - \frac{\sum_{i=1}^{N} f_i'(x)}{\sum_{i=1}^{N} f_i''(x)} = \frac{\frac{1}{N}\sum_{i=1}^{N}\Big(f_i''(x)x - f_i'(x)\Big)}{\frac{1}{N}\sum_{i=1}^{N} f_i''(x)}$$

i.e., parallel of two average consensi

Example: distributed formation control



Example: distributed formation control

video

a look into the future

TODOs

- more general & robust distributed optimization procedures
- develop self-tuning procedures

TODOs

- more general & robust distributed optimization procedures
- develop self-tuning procedures

- relieve humans from automatable burdens

AI over networks: when decisions shall be taken together

Damiano Varagnolo

NTNU, Department of Engineering Cybernetics