# Sparsity-Promoting Optimal Control of Distributed Systems

**Mihailo Jovanović**
www.umn.edu/∼mihailo

joint work with:
**Makan Fardad**
**Fu Lin**

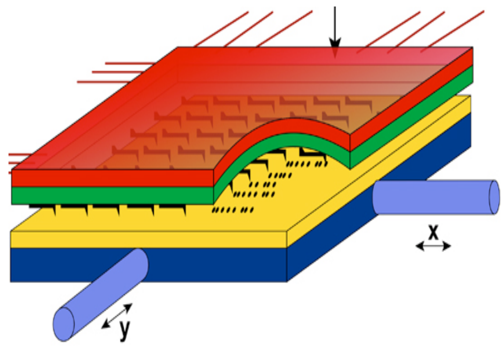
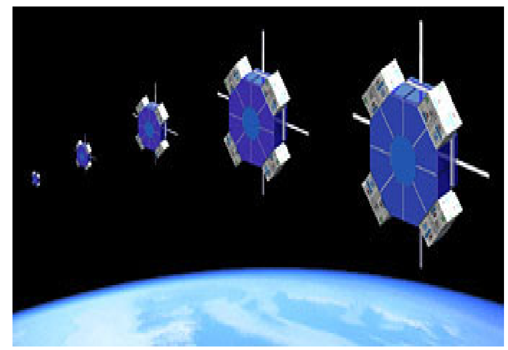
University of Minnesota

**LIDS Seminar, MIT; Nov 6, 2012**

# Large dynamic networks

- OF INCREASING IMPORTANCE IN MODERN TECHNOLOGY

APPLICATIONS:

| wind farms | micro-cantilevers | aircraft formations<br>satellite constellations |
| --- | --- | --- |
|  |  |  |

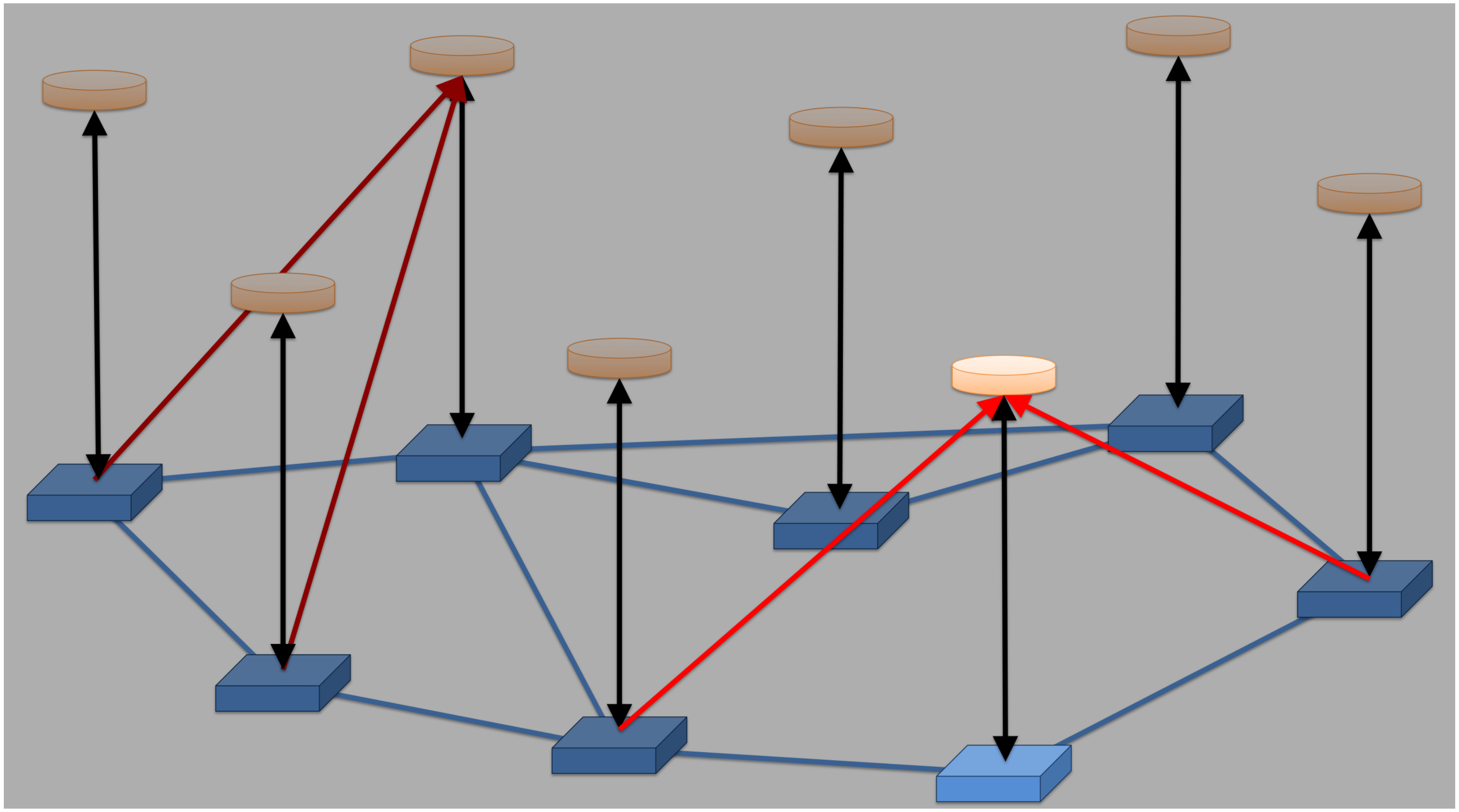
- INTERACTIONS CAUSE COMPLEX BEHAVIOR
  - ⋆ **cannot be predicted by analyzing isolated subsystems**

- SPECIAL STRUCTURE
  - ⋆ **every subsystem has sensors and actuators**

# Structured distributed control

- **Blue layer:** **distributed plant and its interaction links**

**memoryless structured controller**

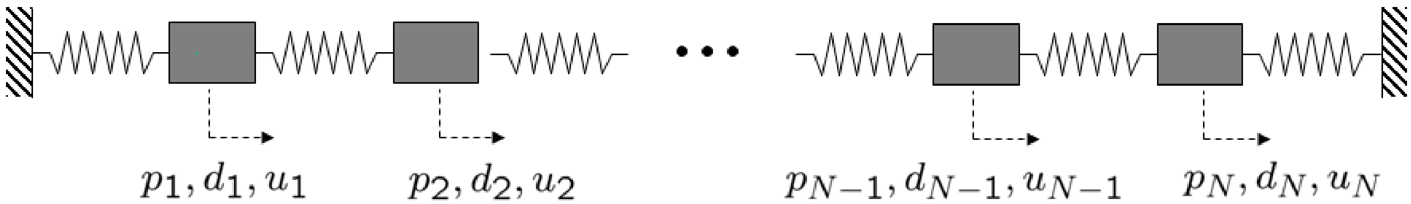


KEY CHALLENGE:

**identification of a signal exchange network**

performance vs. sparsity

# Example: Mass-spring system




MEMORYLESS CONTROLLER: $u(t) \;=\; -\begin{bmatrix} F_p & F_v \end{bmatrix} \begin{bmatrix} p(t) \\ v(t) \end{bmatrix}$

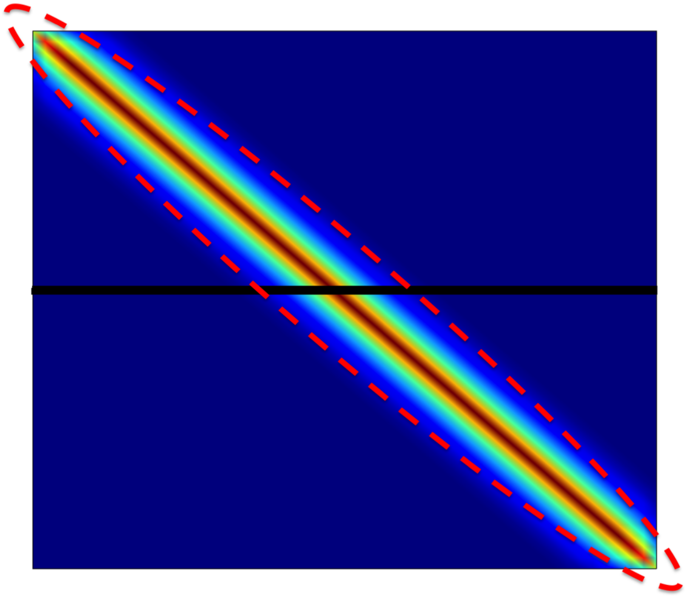
- **Objective: design** $\begin{bmatrix} F_p & F_v \end{bmatrix}$ **to minimize steady-state variance of** $p$, $v$, $u$

OPTIMAL CONTROLLER – LINEAR QUADRATIC REGULATOR

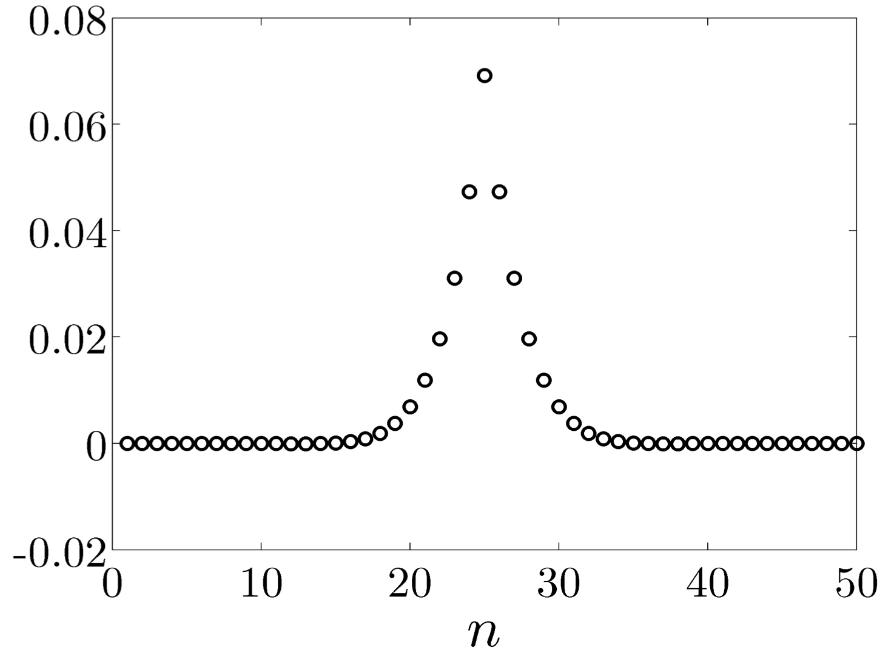
$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \end{bmatrix} = -\underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}}_{F_p} \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \end{bmatrix} - \underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}}_{F_v} \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \\ v_4(t) \end{bmatrix}$$

# Structure of optimal controller

**position feedback matrix:**



**position gains for middle mass:**



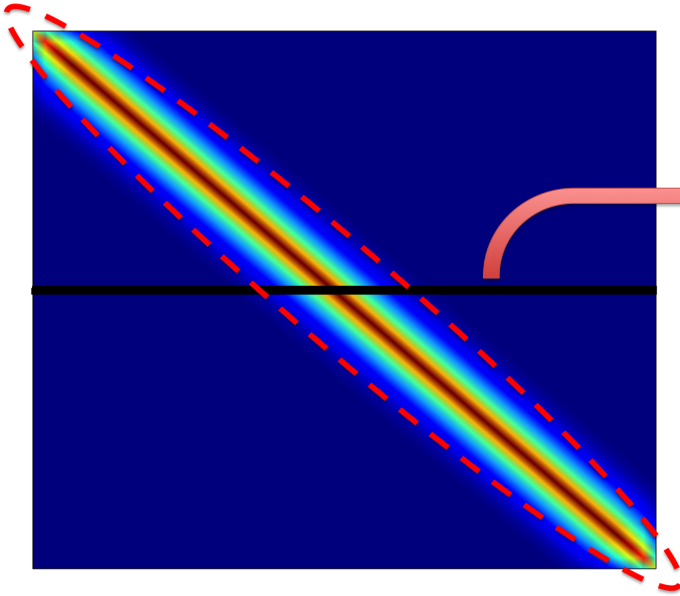

- OBSERVATIONS
  - ⋆ Diagonals almost constant (modulo edges)
  - ⋆ Off-diagonal decay of a centralized gain

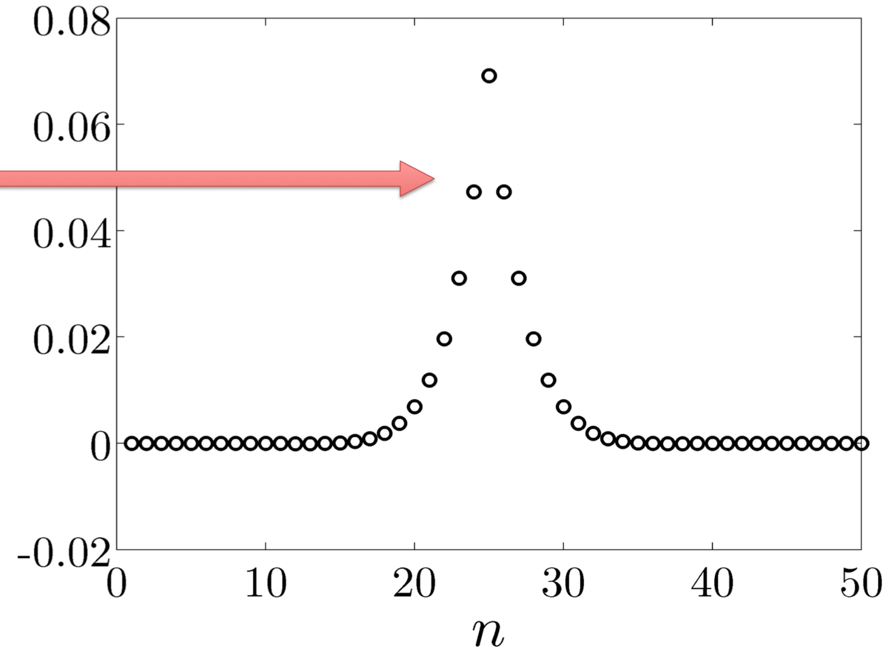
*Bamieh, Paganini, Dahleh, IEEE TAC '02*

*Motee & Jadbabaie, IEEE TAC '08*

# Structure of optimal controller

**position feedback matrix:** **position gains for middle mass:**





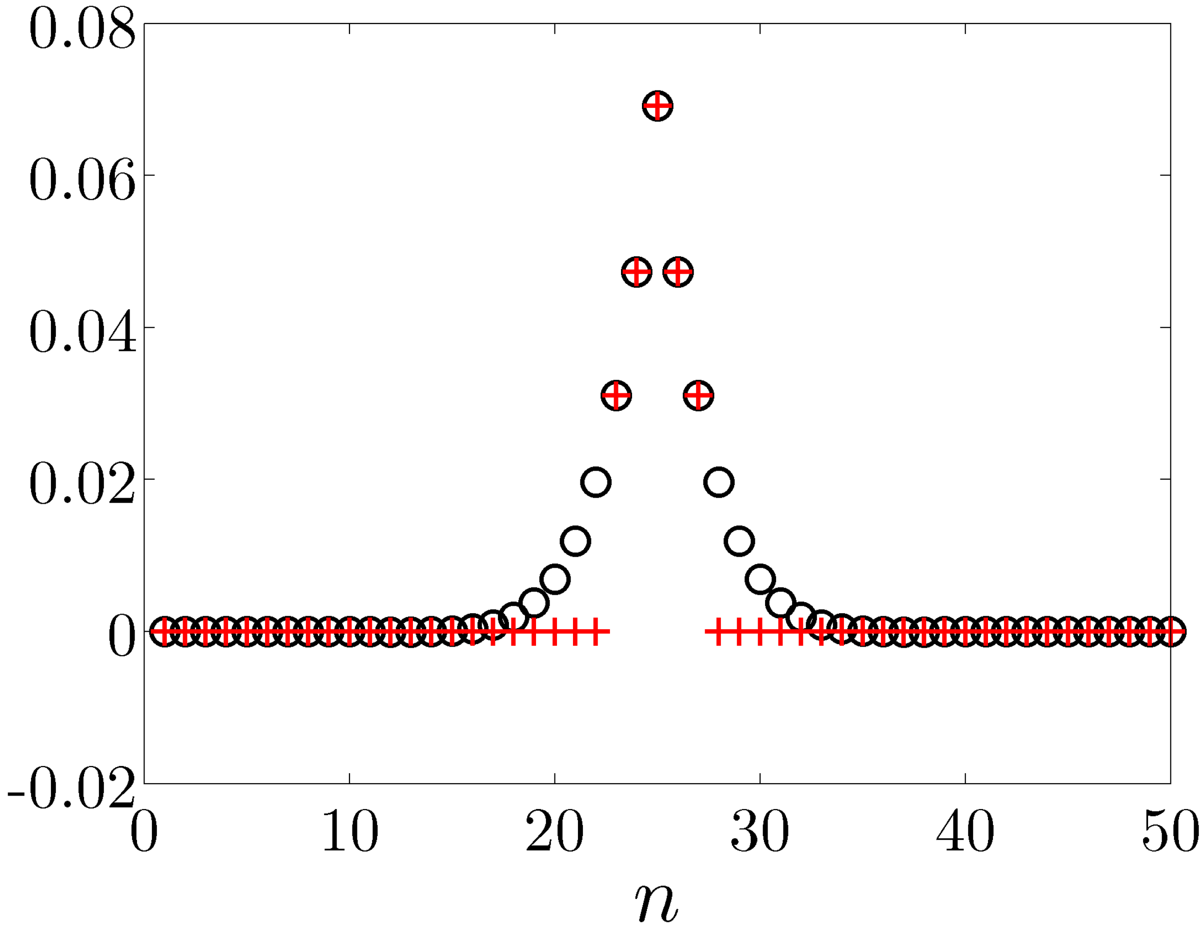

- OBSERVATIONS
  - ⋆ Diagonals almost constant (modulo edges)
  - ⋆ Off-diagonal decay of a centralized gain

*Bamieh, Paganini, Dahleh, IEEE TAC '02*

*Motee & Jadbabaie, IEEE TAC '08*

# Enforcing localization?

- **One approach: truncating centralized controller**




- POSSIBLE DANGERS
  - ⋆ **Performance degradation**
  - ⋆ **Instability**

# Outline

❶ Sparsity-promoting optimal control

- ⋆ **Design of sparse and block sparse feedback gains**
- ⋆ **Tools from control theory, optimization, and compressive sensing**

❷ Algorithm

- ⋆ **Alternating direction method of multipliers**

❸ Examples

❹ Summary and outlook

# SPARSITY-PROMOTING OPTIMAL CONTROL

# State-feedback $H_2$ problem

$$\begin{aligned} \dot{x} &= A\,x \;+\; B_1\,d \;+\; B_2\,u \\ z &= \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix} x \;+\; \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} u \\ u &= -F\,x \end{aligned}$$

- CLOSED-LOOP $H_2$ NORM

$$J(F) \;=\; \operatorname{trace}\left( \int_0^\infty \mathrm{e}^{(A-B_2F)^T t} \left(Q \;+\; F^T R F\right) \mathrm{e}^{(A-B_2F)t} \,\mathrm{d}t\, B_1 B_1^T \right)$$

⋆ **no structural constraints**

**globally optimal controller:**

$$\boxed{\begin{aligned} A^T P \;+\; P\,A \;-\; P\,B_2\,R^{-1}B_2^T P \;+\; Q &= 0 \\ F_c &= R^{-1}B_2^T P \end{aligned}}$$

# Sparsity-promoting $H_2$ problem

$$\text{minimize} \qquad \underbrace{J(F)}_{\text{variance amplification}} \qquad + \qquad \underbrace{\gamma\, \mathbf{card}\,(F)}_{\text{sparsity-promoting penalty function}}$$

⋆ $\mathbf{card}\,(F)$ – number of non-zero elements of $F$

$$F \;=\; \begin{bmatrix} 5.1 & -2.3 & 0 & 1.5 \\ 0 & 3.2 & 1.6 & 0 \\ 0 & -4.3 & 1.8 & 5.2 \end{bmatrix} \;\Rightarrow\; \mathbf{card}\,(F) \;=\; 8$$

⋆ $\gamma > 0$ – performance vs. sparsity tradeoff

*Lin, Fardad, Jovanović, IEEE TAC '12* (submitted; arXiv:1111.6188v2)

# Convex relaxations of $\mathbf{card}\,(F)$

$\ell_1$ **norm:** $\displaystyle\sum_{i,\,j} |F_{ij}|$

**weighted** $\ell_1$ **norm:** $\displaystyle\sum_{i,\,j} W_{ij}\,|F_{ij}|, \quad W_{ij} \;\geq\; 0$

**separable: sum of element-wise functions**

- CARDINALITY VS. WEIGHTED $\ell_1$ NORM

$$\{W_{ij} \;=\; 1/|F_{ij}|, \;\; F_{ij} \;\neq\; 0\} \quad \Rightarrow \quad \mathbf{card}\,(F) \;=\; \sum_{i,\,j} W_{ij}\,|F_{ij}|$$

RE-WEIGHTED SCHEME

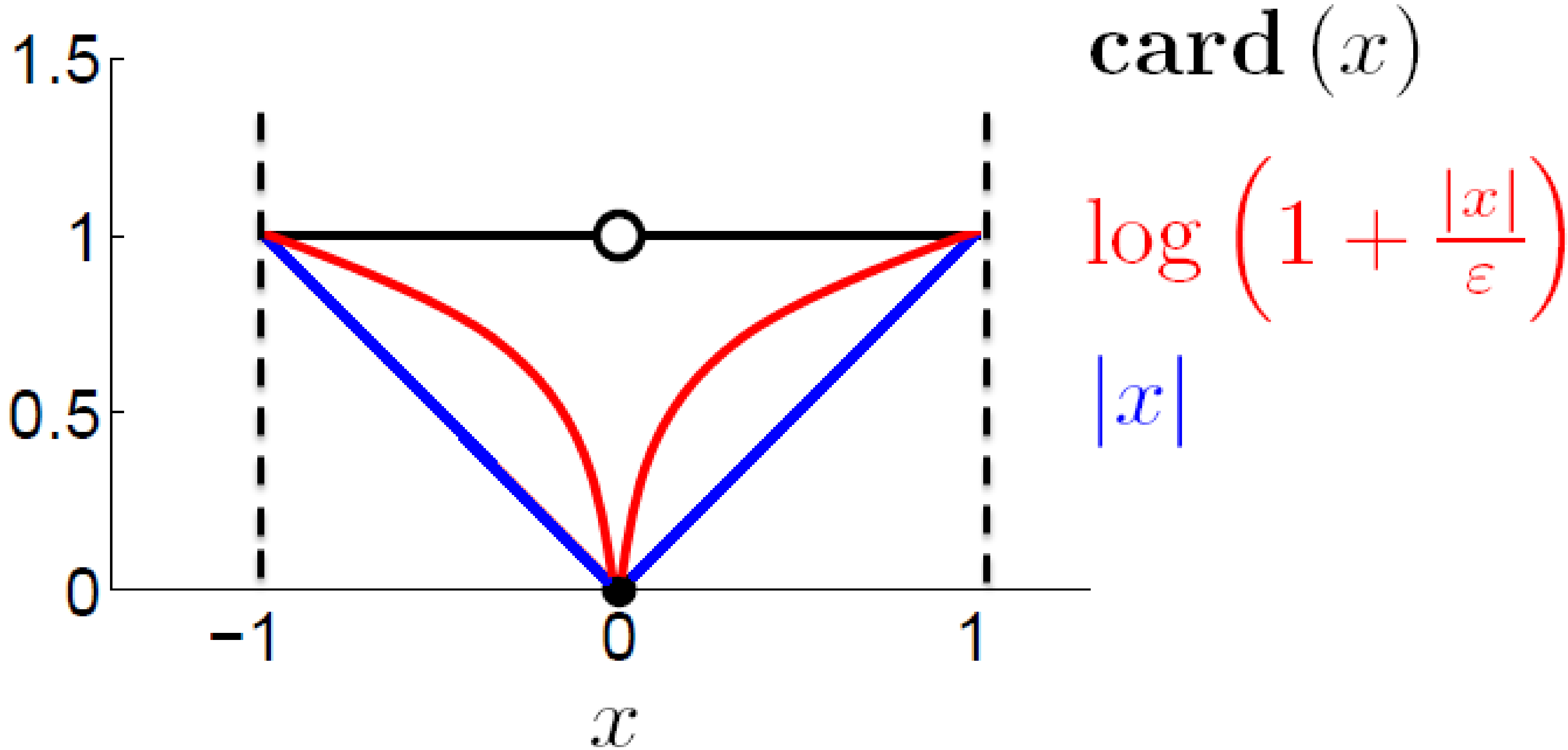
⋆ **Use feedback gains from previous iteration to form weights**

$$W_{ij}^{+} \;=\; \frac{1}{|F_{ij}| \;+\; \varepsilon}$$

*Candès, Wakin, Boyd, J. Fourier Anal. Appl. '08*

# A non-convex relaxation of $\mathbf{card}\,(F)$

**sum-of-logs:** $\displaystyle\sum_{i,j} \log\left(1 + \frac{|F_{ij}|}{\varepsilon}\right), \qquad 0 < \varepsilon \ll 1$




*Candès, Wakin, Boyd, J. Fourier Anal. Appl. '08*

# Sparsity-promoting penalty functions

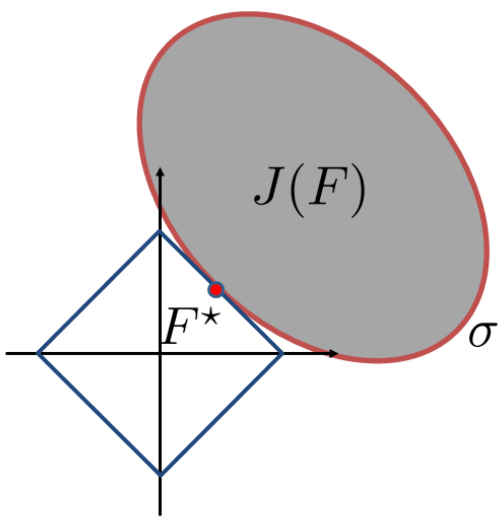
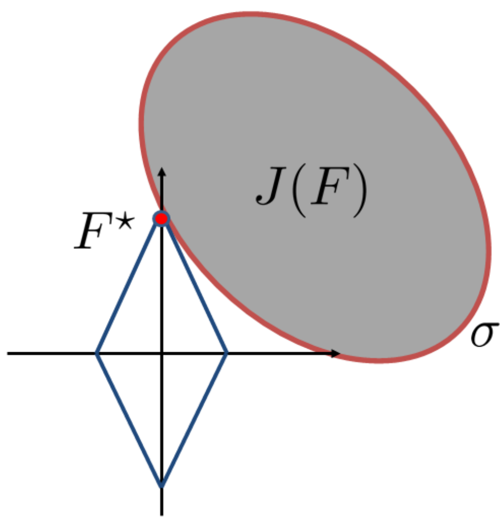
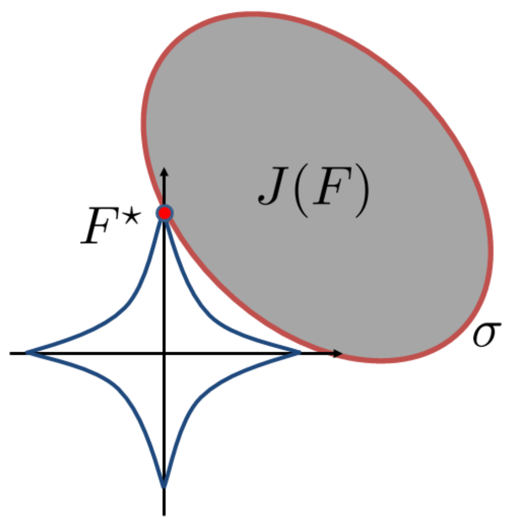
**original problem:**

$$\begin{array}{ll} \text{minimize} & \textbf{card}\,(F) \\ \text{subject to} & J(F) \,\leq\, \sigma \end{array}$$

$\Rightarrow$

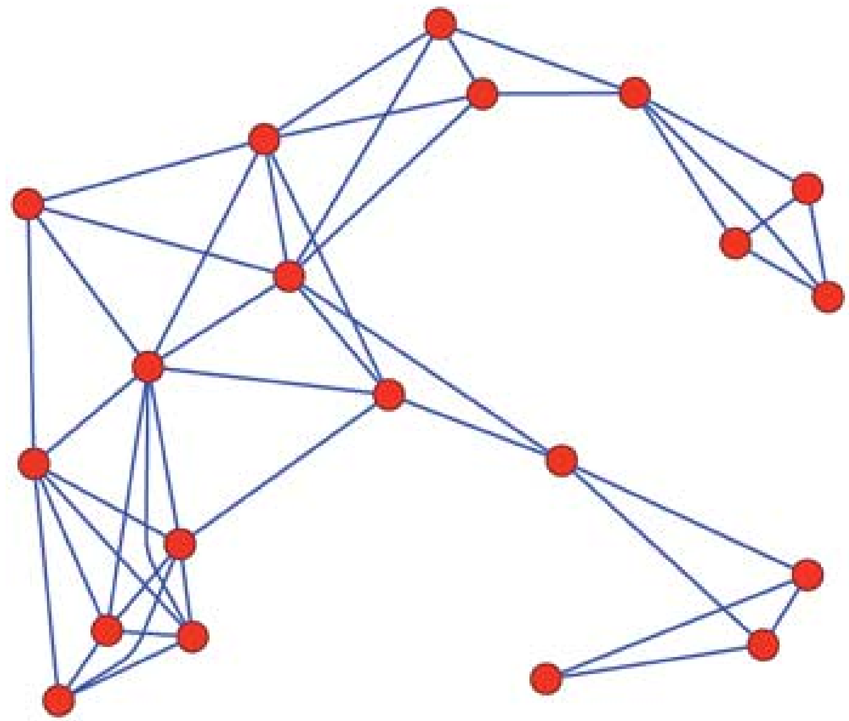
**relaxation:**

$$\begin{array}{ll} \text{minimize} & g(F) \\ \text{subject to} & J(F) \,\leq\, \sigma \end{array}$$

| $\ell_1$ | **weighted $\ell_1$** | **sum-of-logs** |
|---|---|---|
|  |  |  |

# A CLASS OF CONVEX PROBLEMS

# Consensus by distributed computation



- RELATIVE INFORMATION EXCHANGE WITH NEIGHBORS
  - ⋆ **simple distributed averaging algorithm**

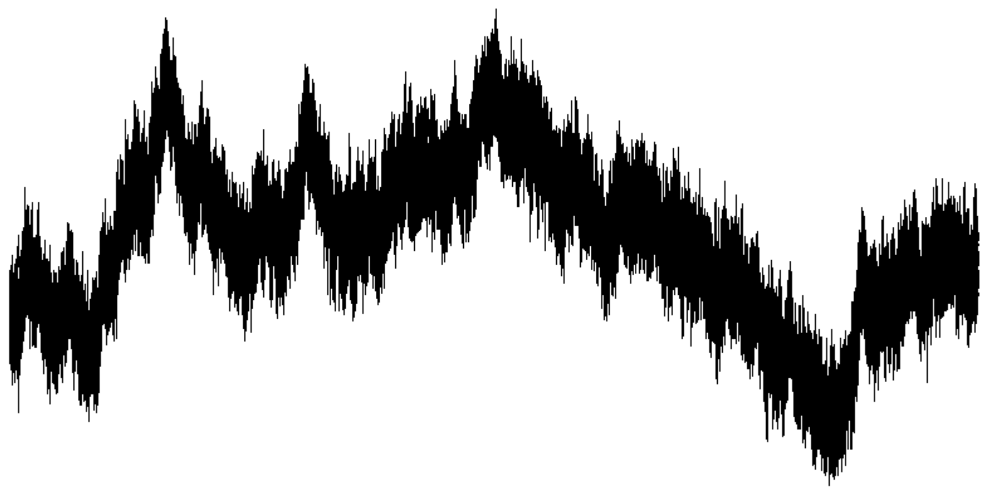
$$\dot{x}_i(t) = -\sum_{j \in \mathcal{N}_i} \Big(x_i(t) - x_j(t)\Big)$$

**connected network** $\Rightarrow$ **convergence to the average value**

# Consensus with stochastic disturbances

$$\dot{x}_i(t) \;=\; -\sum_{j\,\in\,\mathcal{N}_i}\Big(x_i(t) \,-\, x_j(t)\Big) \;+\; d_i(t)$$

- **Average mode:** $\bar{x}(t) \;=\; \dfrac{1}{N}\displaystyle\sum_{i\,=\,1}^{N} x_i(t)$ : undergoes random walk



If other modes are stable, $x_i(t)$ fluctuates around $\bar{x}(t)$

deviation from average: $\tilde{x}_i(t) \;=\; x_i(t) \;-\; \bar{x}(t)$

steady-state variance: $\displaystyle\lim_{t\,\to\,\infty} \mathcal{E}\left(\tilde{x}^T(t)\,\tilde{x}(t)\right)$

# Design of undirected networks of single integrators

- **Convex optimization problem**

$$\begin{array}{ll} \text{minimize} & \operatorname{trace}\left(Q^{1/2}\,(F + \mathbb{1}\mathbb{1}^T/N)^{-1}\,Q^{1/2} \;+\; R\,F\right) \;+\; \gamma\,\|W \circ F\|_{\ell_1} \\ \text{subject to} & F\,\mathbb{1} \;=\; 0 \\ & F \;+\; \mathbb{1}\mathbb{1}^T/N \;\succ\; 0 \end{array}$$

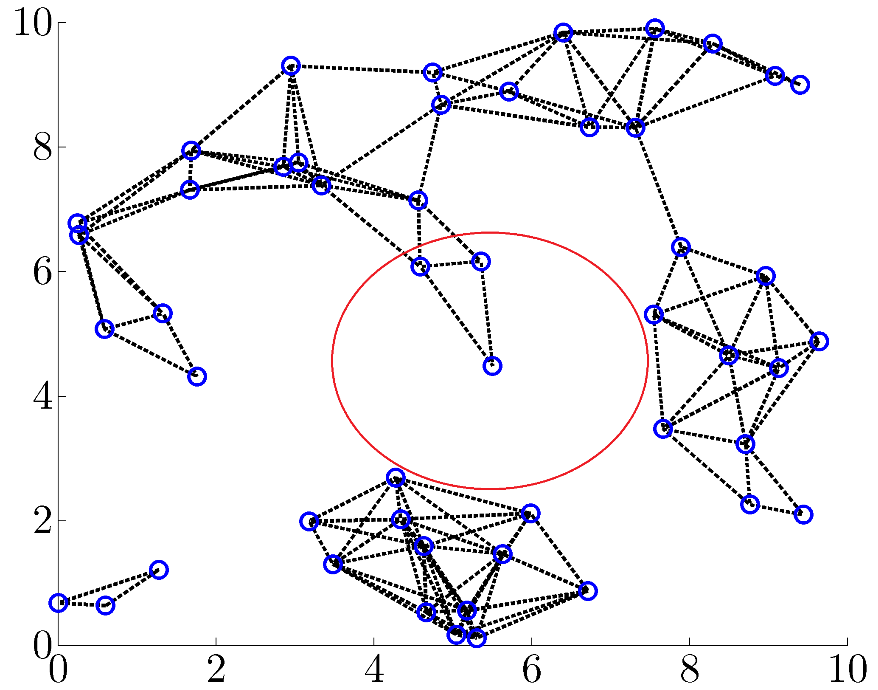
**can be formulated as SDP:**

$$\boxed{\begin{array}{ll} \text{minimize} & \operatorname{trace}\,(X \;+\; R\,F) \;+\; \gamma\,\mathbb{1}^T\,Y\,\mathbb{1} \\ \text{subject to} & \left[\begin{array}{cc} X & Q^{1/2} \\ Q^{1/2} & F + \mathbb{1}\mathbb{1}^T/N \end{array}\right] \succeq\; 0 \\ & -Y \;\leq\; W \circ F \;\leq\; Y \\ & F\,\mathbb{1} \;=\; 0 \end{array}}$$

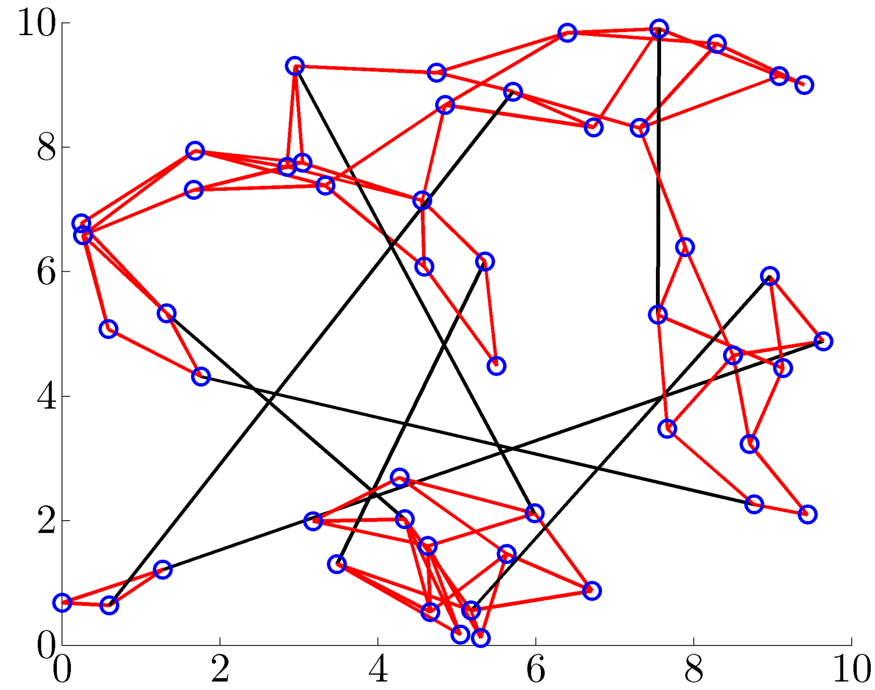
# Sparsity-promoting consensus algorithm

- UNDIRECTED NETWORK WITH STOCHASTIC DISTURBANCES

**local performance graph:**




**identified communication graph:**

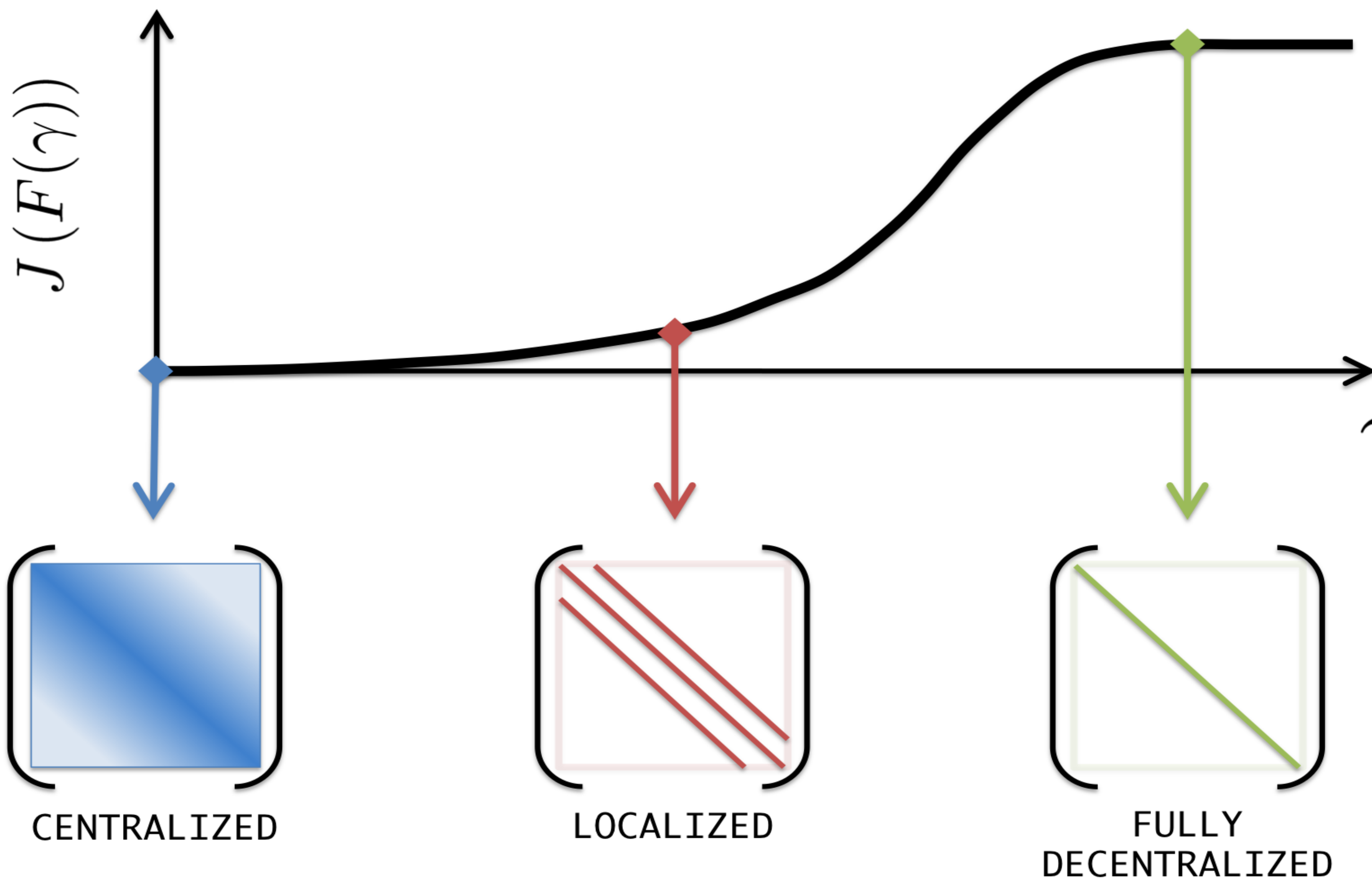



$$z = \begin{bmatrix} Q_{\rm loc}^{1/2}\, x \\ \left(I - \frac{1}{N}\mathbb{1}\mathbb{1}^T\right) x \\ u \end{bmatrix}$$

$$\begin{aligned} \textbf{card}\,(F)\,/\textbf{card}\,(F_{\rm c}) &= 7\% \\ (J - J_{\rm c})\,/J_{\rm c} &= 14\% \end{aligned}$$

# Parameterized family of feedback gains

$$F(\gamma) \;:=\; \arg\min_{F} \left( J(F) \;+\; \gamma\, g(F) \right)$$



# Alternating direction method of multipliers

$$\text{minimize} \quad J(F) \;+\; \gamma\, g(F)$$

- **Step 1: introduce additional variable/constraint**

$$\boxed{\begin{array}{ll} \text{minimize} & J(F) \;+\; \gamma\, g(G) \\ \text{subject to} & F \;-\; G \;=\; 0 \end{array}}$$

**benefit: decouples $J$ and $g$**

- **Step 2: introduce augmented Lagrangian**

$$\mathcal{L}_\rho(F, G, \Lambda) \;=\; J(F) \;+\; \gamma\, g(G) \;+\; \operatorname{trace}\left(\Lambda^T (F \;-\; G)\right) \;+\; \frac{\rho}{2}\, \|F \;-\; G\|_F^2$$

- **Step 3: use ADMM for augmented Lagrangian minimization**

$$\mathcal{L}_\rho(F,G,\Lambda) \;=\; J(F) \,+\, \gamma\, g(G) \,+\, \mathrm{trace}\left(\Lambda^T(F-G)\right) \,+\, \frac{\rho}{2}\,\|F-G\|_F^2$$

ADMM:

$$\boxed{\begin{array}{lcl} F^{k+1} & := & \underset{F}{\arg\min}\ \mathcal{L}_\rho(F, G^k, \Lambda^k) \\[1em] G^{k+1} & := & \underset{G}{\arg\min}\ \mathcal{L}_\rho(F^{k+1}, G, \Lambda^k) \\[1em] \Lambda^{k+1} & := & \Lambda^k \,+\, \rho\left(F^{k+1} \,-\, G^{k+1}\right) \end{array}}$$

MANY MODERN APPLICATIONS

- ⋆ distributed computing
- ⋆ distributed signal processing
- ⋆ image denoising
- ⋆ machine learning

*Eckstein & Bertsekas, '92; Boyd et al., '11*

- **Step 4: Polishing – back to structured optimal design**

$\star$ ADMM $\begin{cases} \textbf{identifies sparsity patterns} \\ \textbf{provides good initial condition for structured design} \end{cases}$

$\star$ NECESSARY CONDITIONS FOR OPTIMALITY OF THE STRUCTURED PROBLEM

$$\boxed{\begin{aligned} (A - B_2 F)^T P + P (A - B_2 F) &= -(Q + F^T R F) \\ (A - B_2 F) L + L (A - B_2 F)^T &= -B_1 B_1^T \\ \left[ (R F - B_2^T P) L \right] \circ I_{\mathcal{S}} &= 0 \end{aligned}}$$

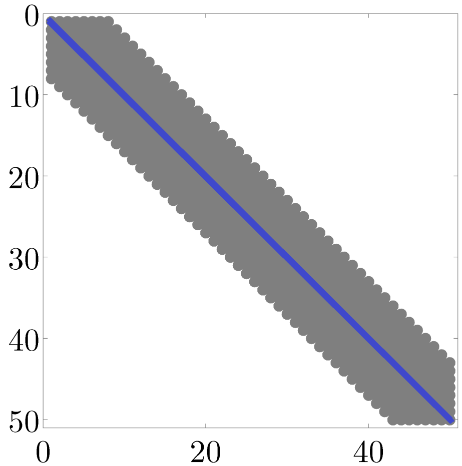
**Newton's method** $+$ **conjugate gradient**

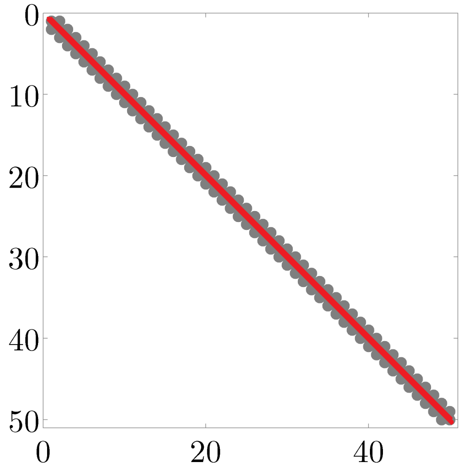
$I_{\mathcal{S}}$ - **structural identity**

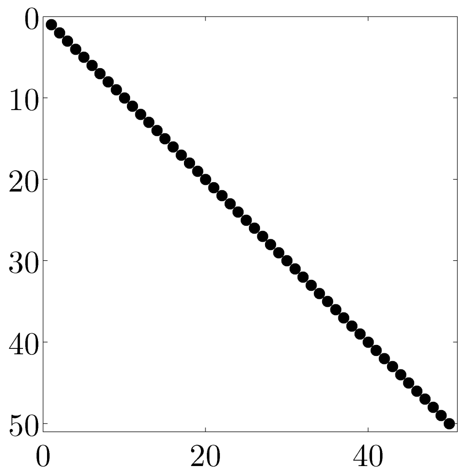
$$F = \begin{bmatrix} * & * & & \\ * & * & * & \\ & * & * & * \\ & & * & * \end{bmatrix} \quad \Rightarrow \quad I_{\mathcal{S}} = \begin{bmatrix} 1 & 1 & & \\ 1 & 1 & 1 & \\ & 1 & 1 & 1 \\ & & 1 & 1 \end{bmatrix}$$

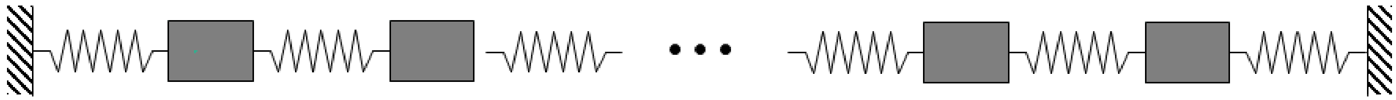
# EXAMPLES

# Mass-spring system

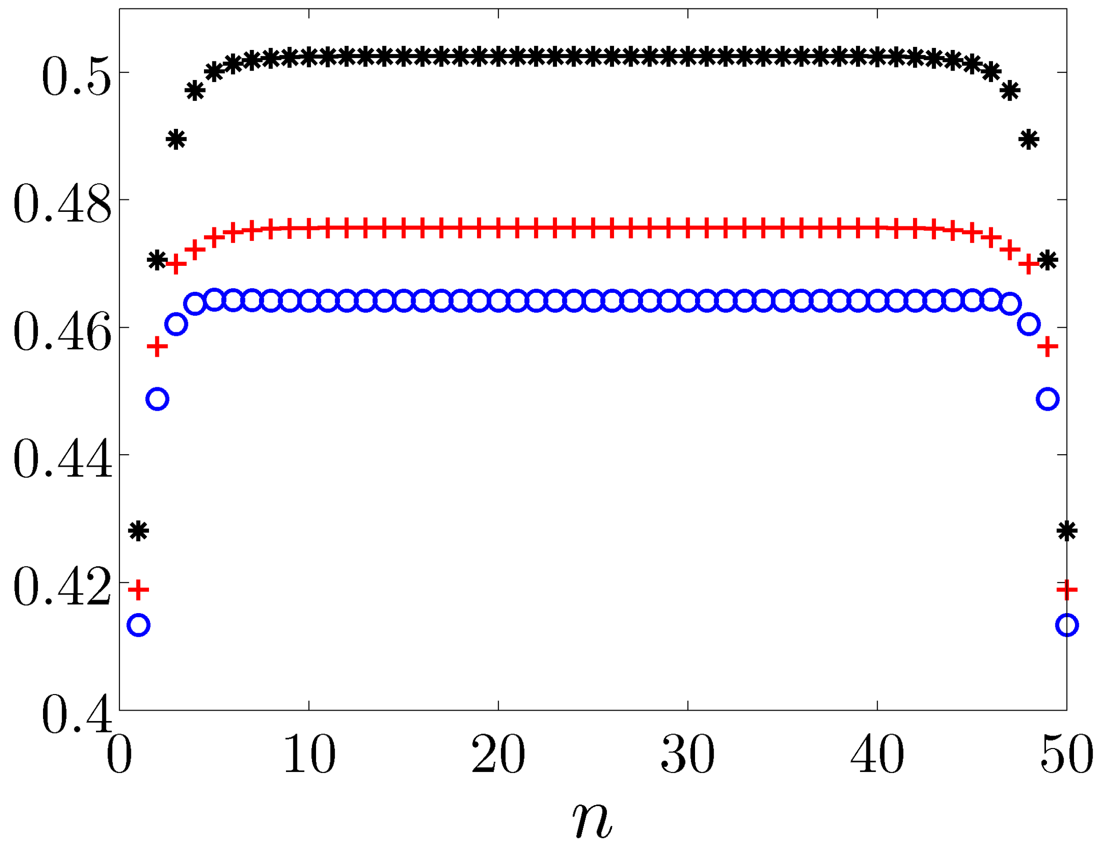


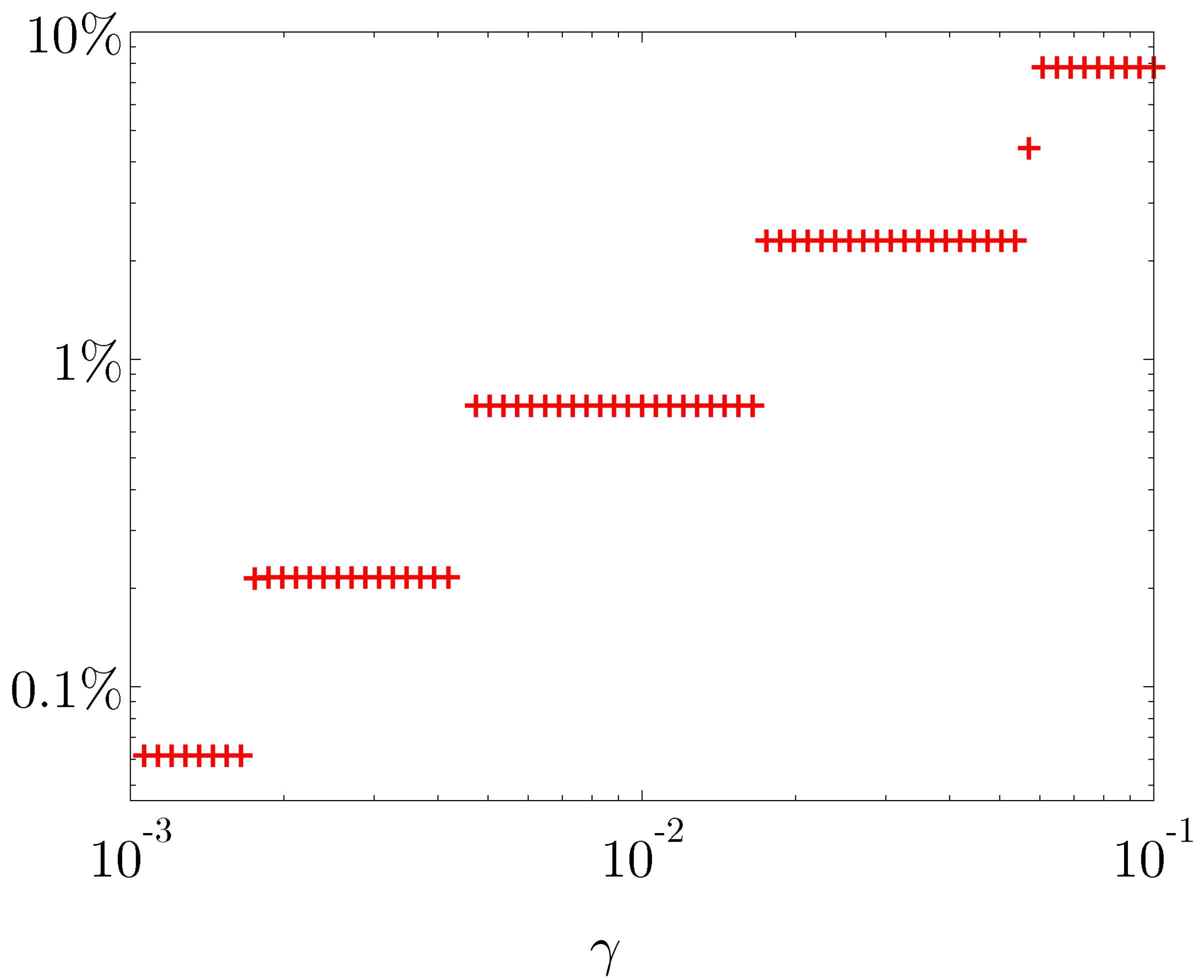
$\mathrm{diag}\,(F_v)$:




$\gamma = 10^{-4}$ $\gamma = 0.03$ $\gamma = 0.1$

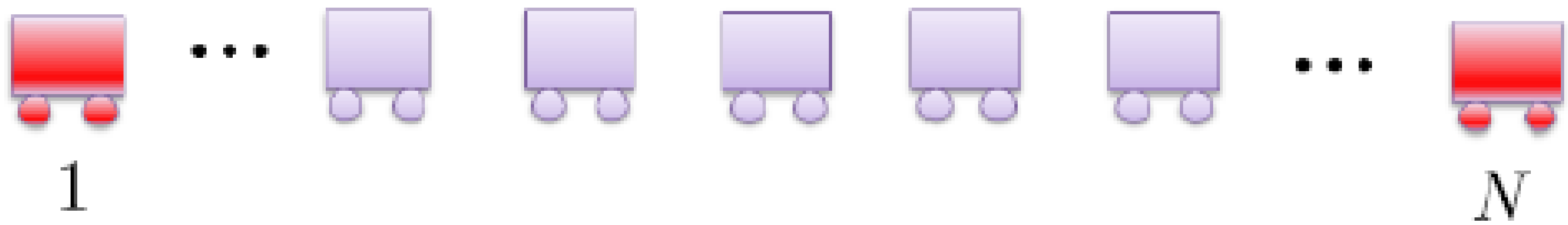
- **Performance comparison: sparse vs. centralized**

$(J - J_c)/J_c$:




| $\mathbf{card}\,(F)\,/\mathbf{card}\,(F_c)$ | $(J - J_c)/J_c$ |
|---|---|
| 10% | 0.75% |
| 6% | 2.4% |
| 2% | 7.8% |

# Formation of vehicles



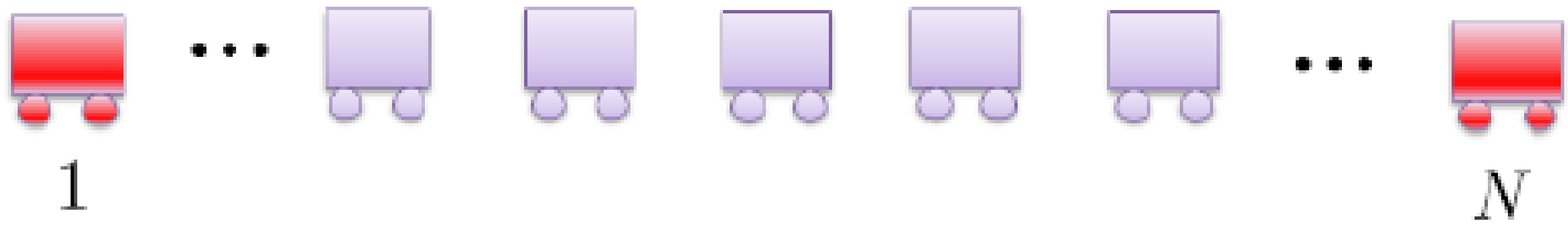

⋆ **Each vehicle: relative information exchange**

$$u_i = -\sum_{j \neq i} F_{ij} \left( x_i - x_j \right), \qquad i \in \{2, \dots, N-1\}$$

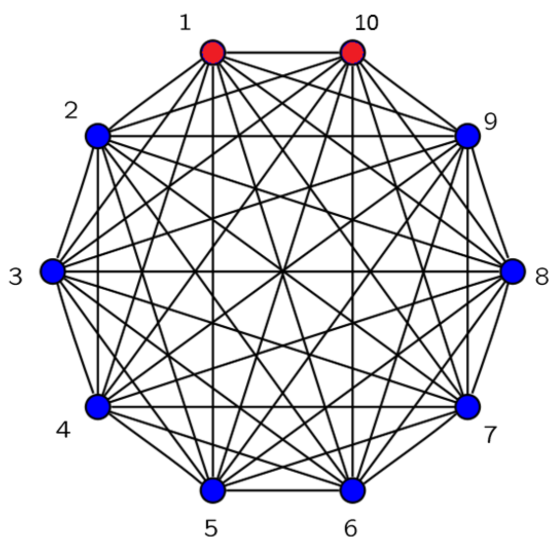
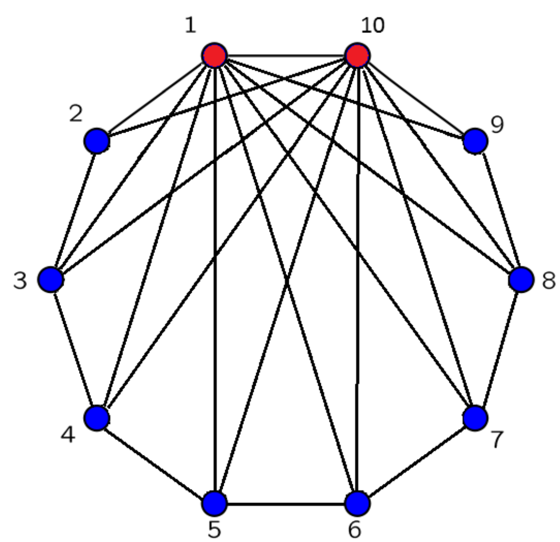
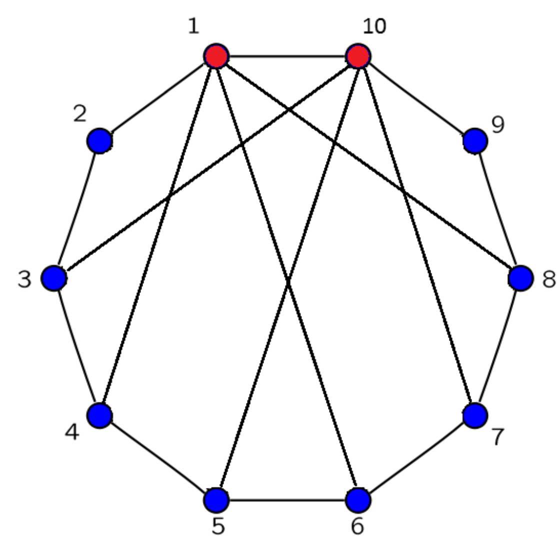
⋆ **Leaders: equipped with GPS devices**

$$u_1 = -\sum_{j \neq 1} F_{1j} \left( x_1 - x_j \right) - F_{11} \, x_1$$

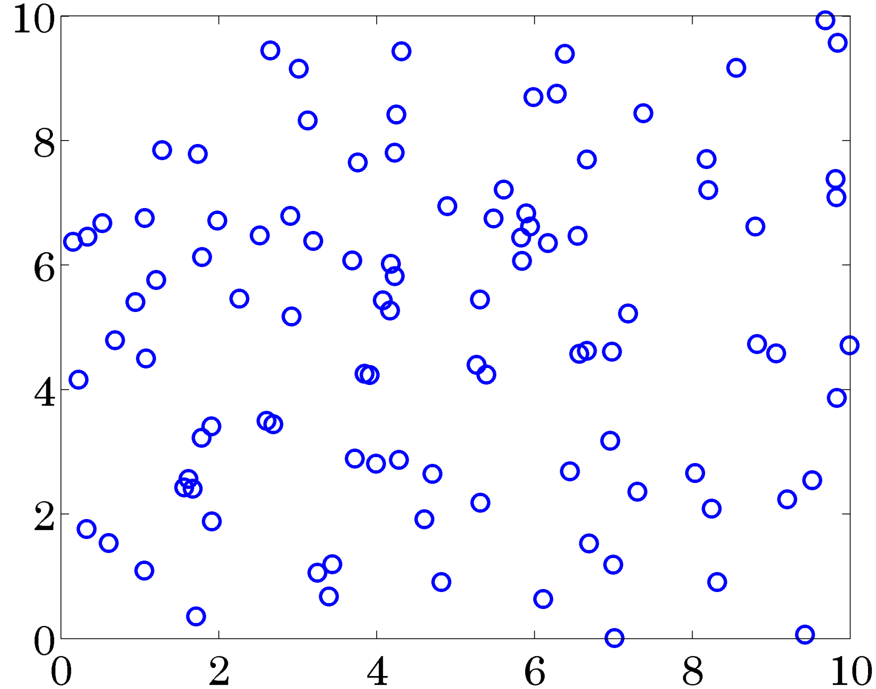
$$u_N = -\sum_{j \neq N} F_{Nj} \left( x_N - x_j \right) - F_{NN} \, x_N$$




## IDENTIFIED COMMUNICATION ARCHITECTURES:

| $\gamma = 0$ | $\gamma = 0.01$ | $\gamma = 0.05$ |
|---|---|---|
|   |   |   |
| **all-to-all** | **nearest neighbors** + **leaders-to-all** | **nearest neighbors** + **leaders-to-some** |

# Network with $100$ nodes

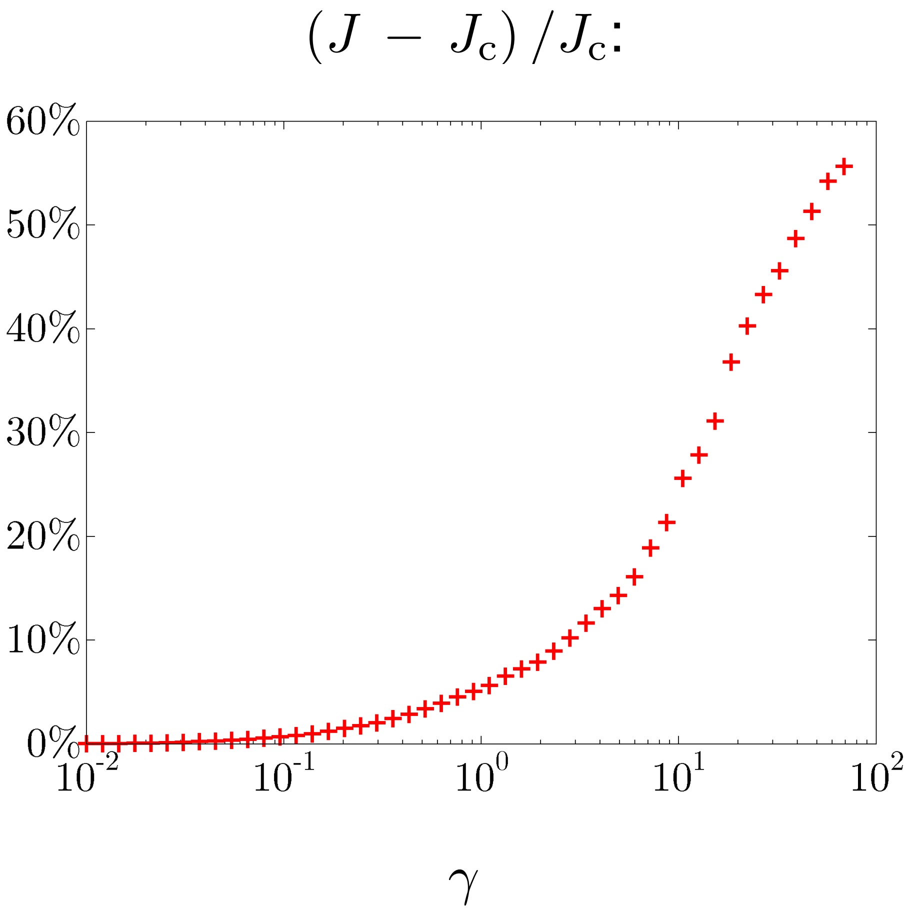


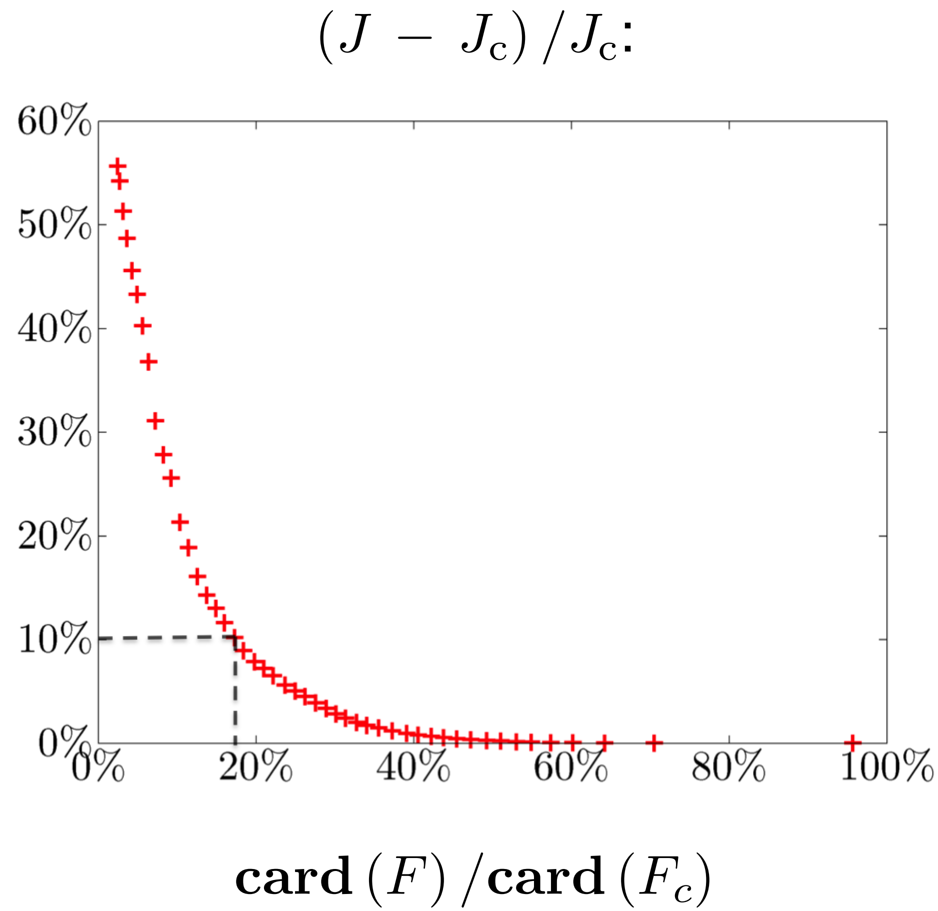

$$\begin{bmatrix} \dot{p}_i \\ \dot{v}_i \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} p_i \\ v_i \end{bmatrix}}_{\textbf{unstable dynamics}} + \underbrace{\sum_{j \neq i} \mathrm{e}^{-\alpha(i,j)} \begin{bmatrix} p_j \\ v_j \end{bmatrix}}_{\textbf{coupling}} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (d_i + u_i)$$

$\alpha(i,j)$: Euclidean distance between nodes $i$ and $j$

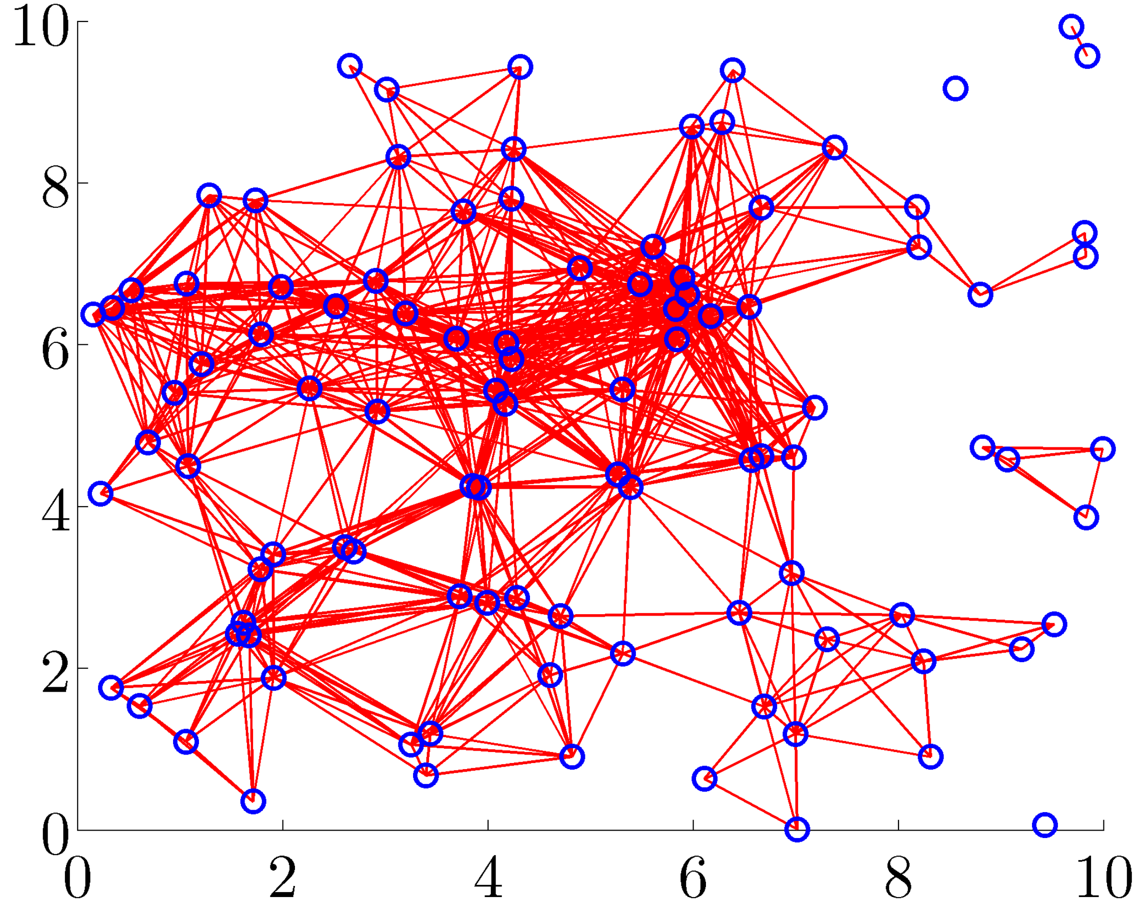
*Motee & Jadbabaie, IEEE TAC '08*

- **Performance comparison: sparse vs. centralized**



## identified communication graph:

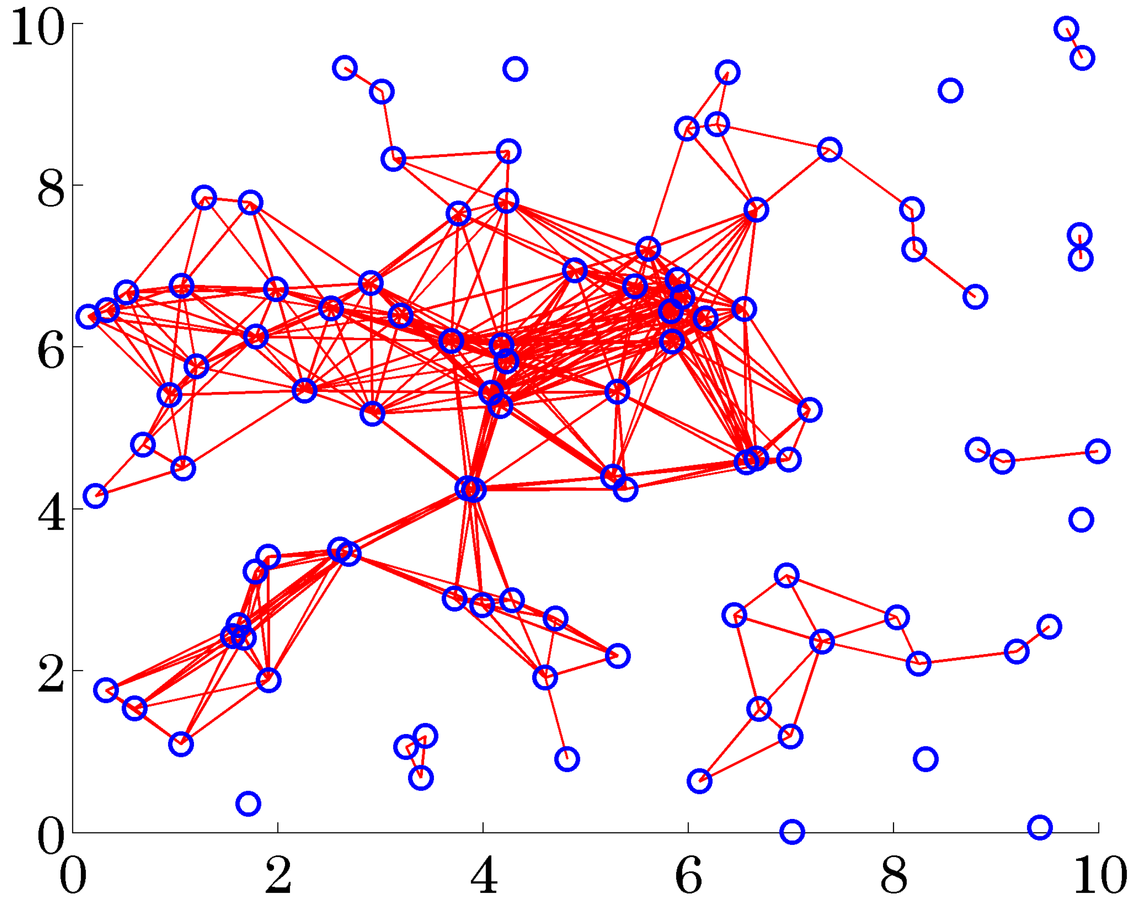



$$\gamma = 5$$

$$\mathbf{card}\,(F)\,/\,\mathbf{card}\,(F_{\mathrm{c}}) = 8.8\%$$

$$(J - J_{\mathrm{c}})\,/\,J_{\mathrm{c}} = 24.6\%$$

## identified communication graph:

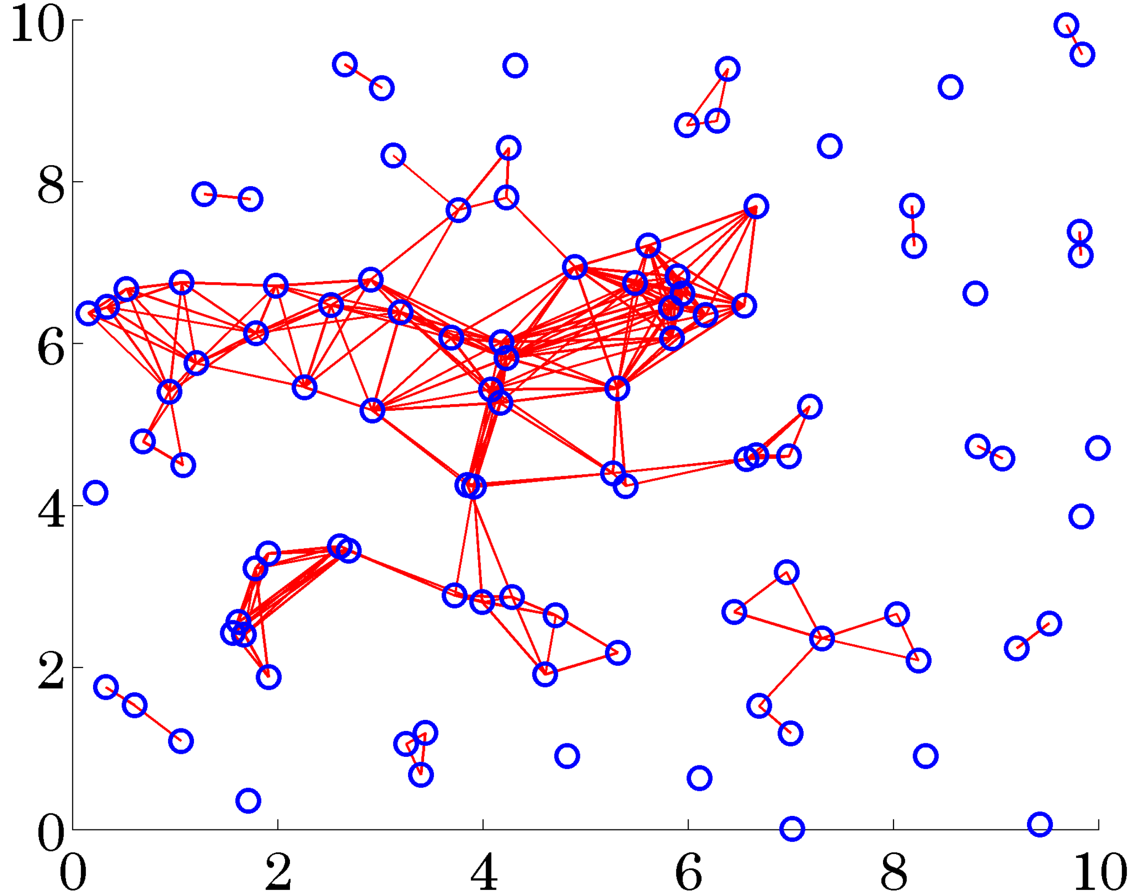



$$\gamma = 11$$

$$\mathbf{card}\,(F)\,/\mathbf{card}\,(F_{\mathrm{c}}) = 5.1\%$$

$$(J - J_{\mathrm{c}})\,/J_{\mathrm{c}} = 40.9\%$$

## identified communication graph:

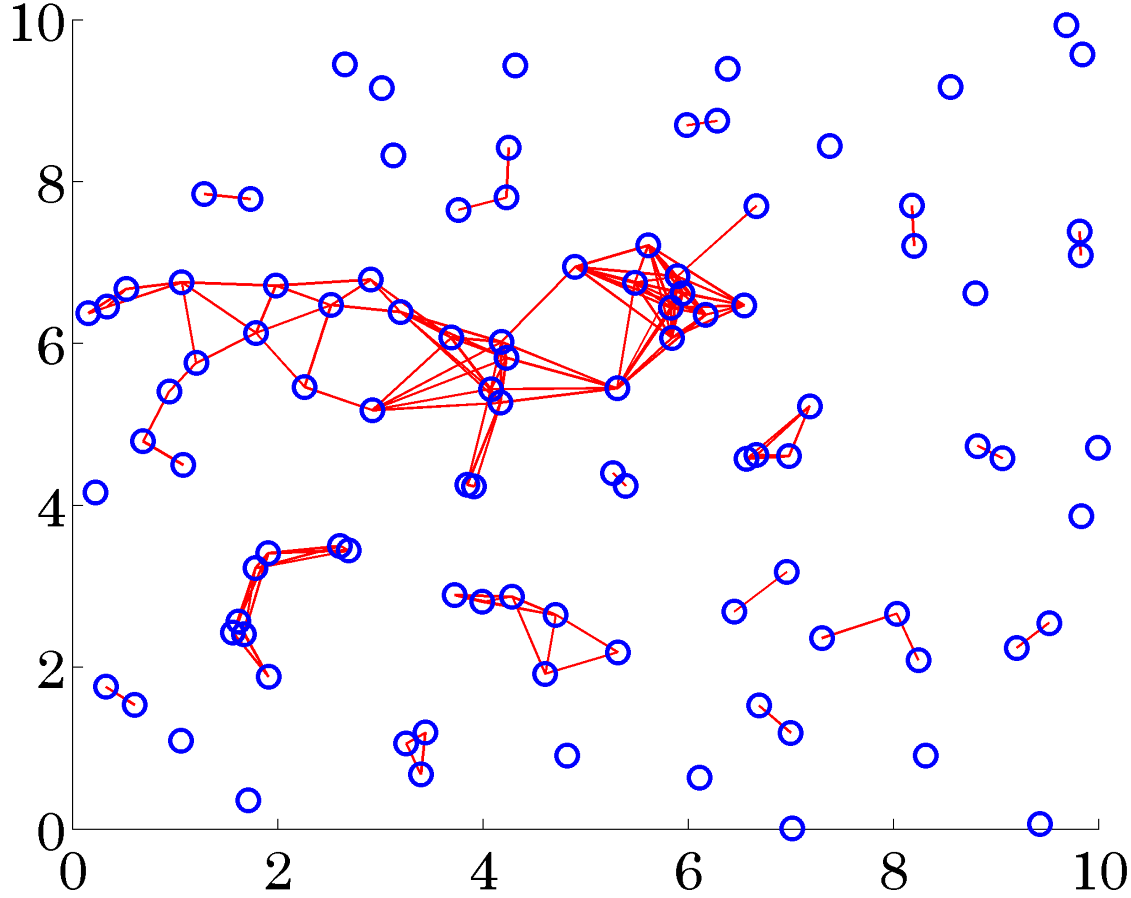


$$\gamma = 18$$

$$\mathbf{card}\,(F)\,/\mathbf{card}\,(F_{\mathrm{c}}) = 3.4\%$$

$$(J - J_{\mathrm{c}})\,/J_{\mathrm{c}} = 48.7\%$$

## identified communication graph:

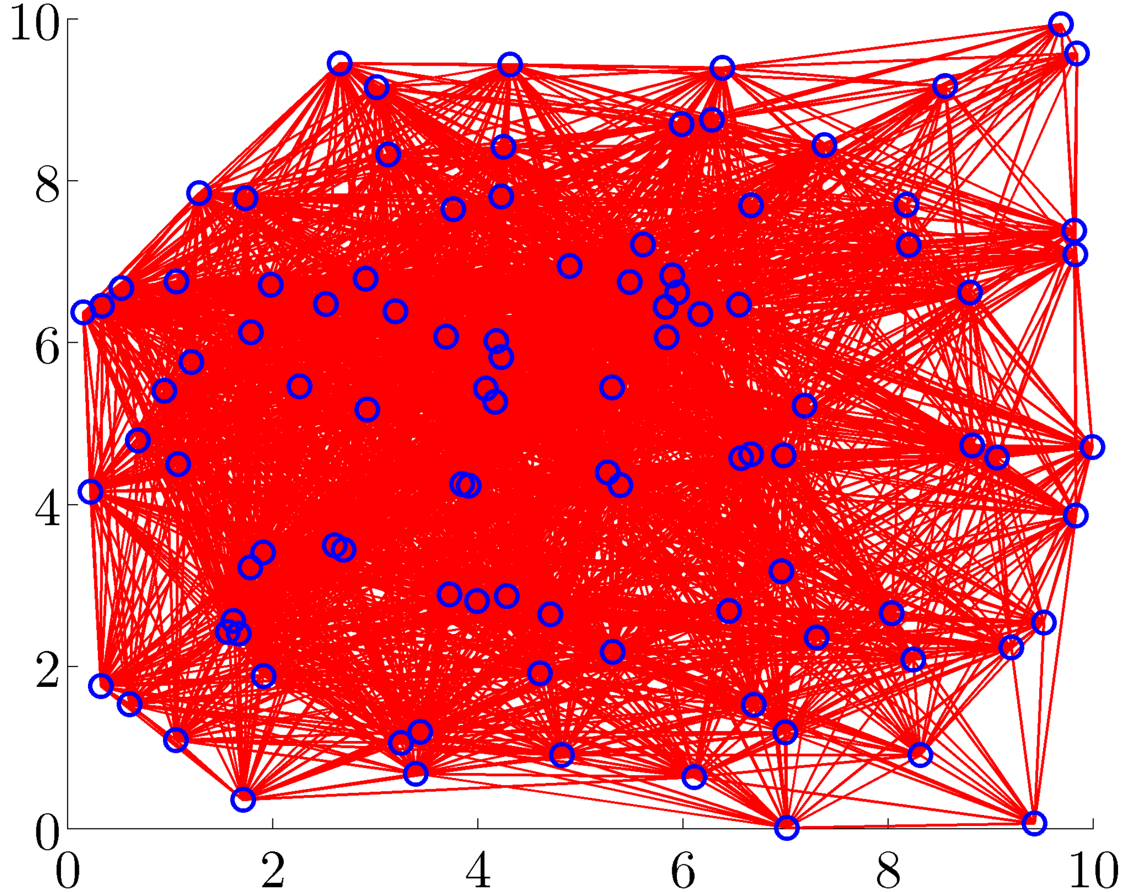



$$\gamma = 30$$

$$\mathbf{card}(F)/\mathbf{card}(F_c) = 2.4\%$$

$$(J - J_c)/J_c = 54.8\%$$

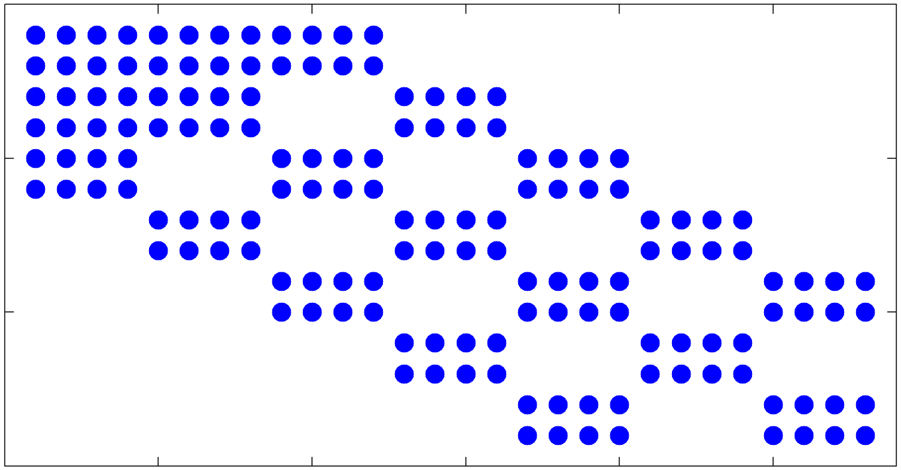
## communication graph of a truncated centralized gain:




$\mathbf{card}\,(F) \;=\; 7380 \;\; (36.9\%)$

**non-stabilizing**

# Extension: Block sparsity



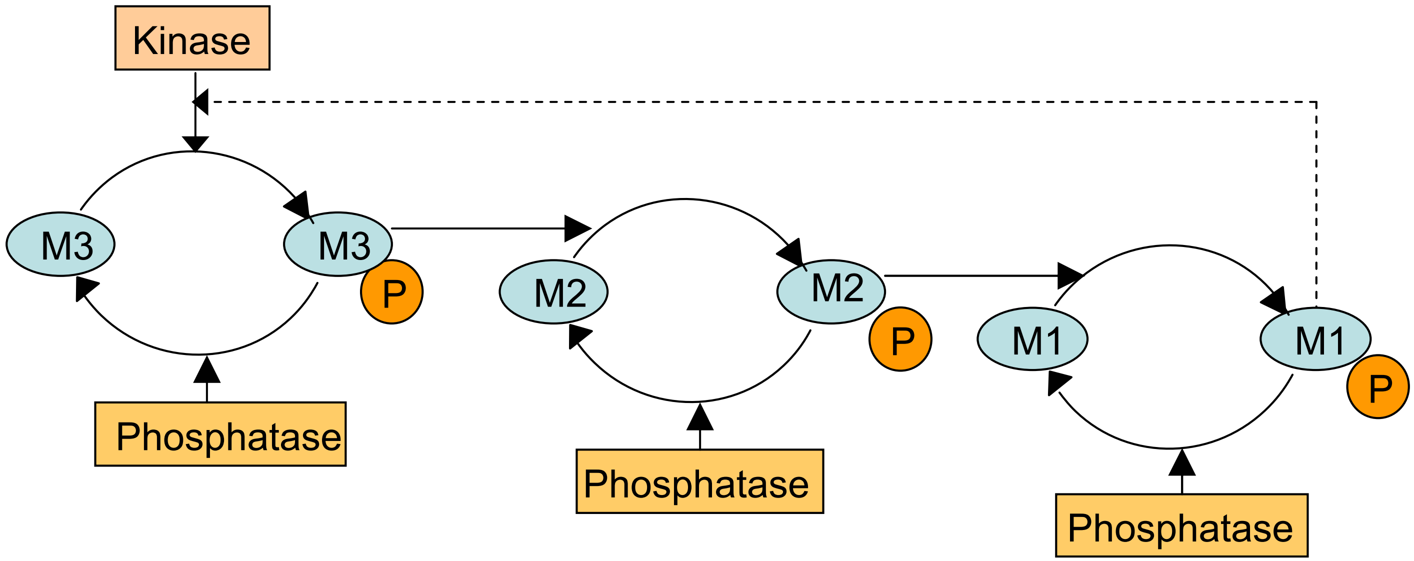
- $\mathbf{card}_{\mathrm{b}}\,(F)$ – number of non-zero blocks of $F$

$$\mathbf{card}_{\mathrm{b}}\,(F) \;=\; \sum_{i,j} \mathbf{card}\,(\|F_{ij}\|_F)$$

- PENALTY FUNCTIONS THAT PROMOTE BLOCK SPARSITY
  - ⋆ generalized $\ell_1$, weighted $\ell_1$, sum-of-logs

# An example from biochemical reactions

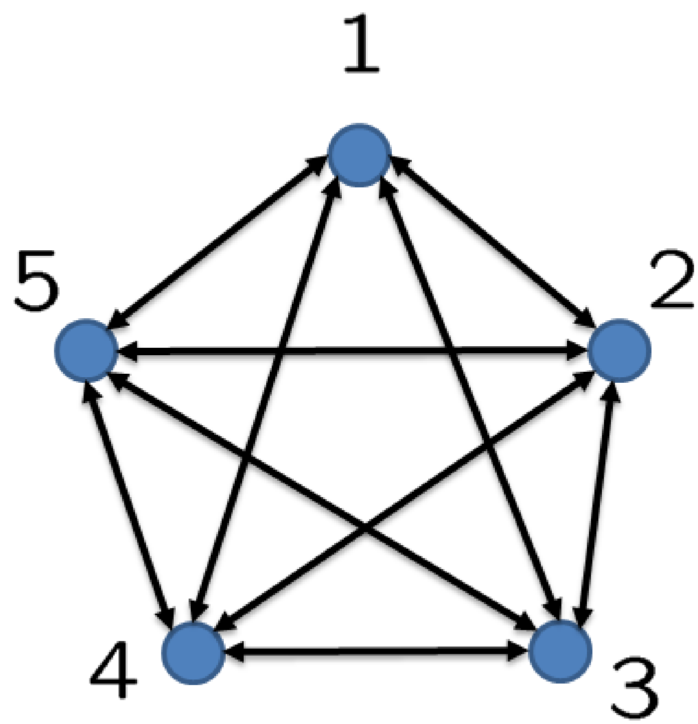
- CYCLIC INTERCONNECTION STRUCTURE




$$\dot{x}_i = [A]_{ii}\, x_i + [B_1]_{ii}\, d_i + [B_2]_{ii}\, u_i$$

$$[A]_{ii} = \begin{bmatrix} -1 & 0 & -3 \\ 3 & -1 & 0 \\ 0 & 3 & -1 \end{bmatrix}, \quad [B_1]_{ii} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad [B_2]_{ii} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

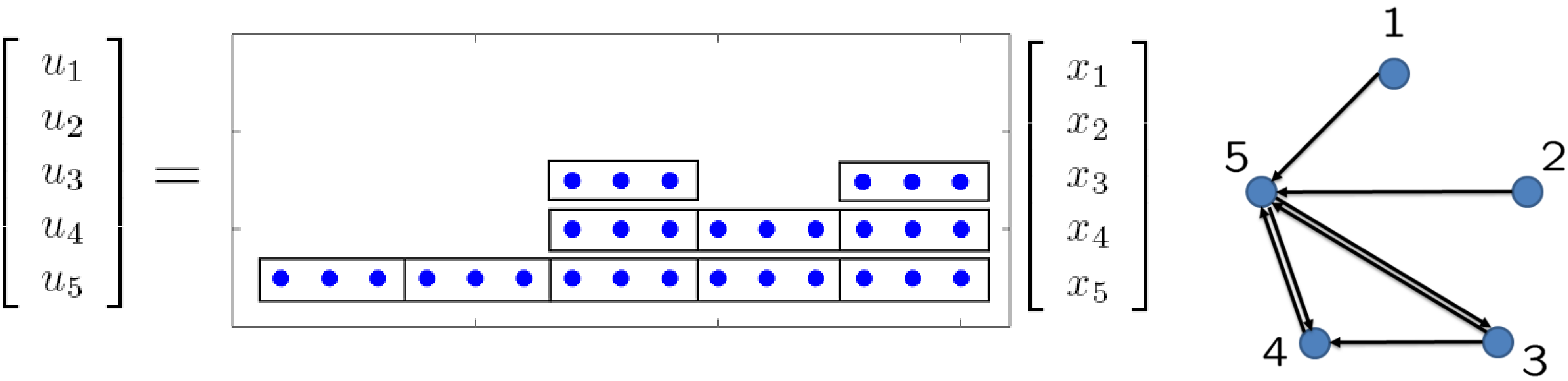
- NONSYMMETRIC BIDIRECTIONAL COUPLING



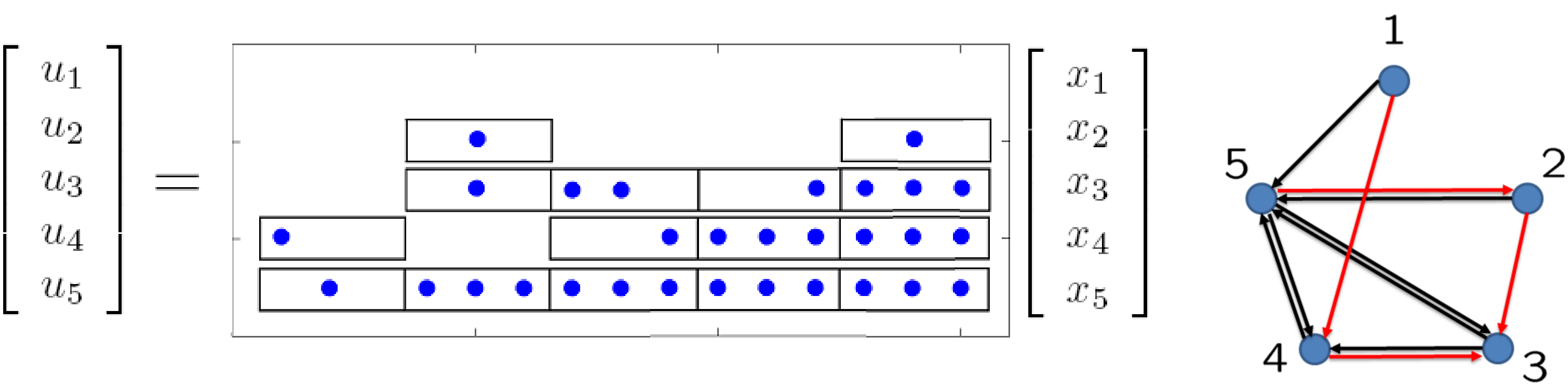

$$\dot{x}_i = [A]_{ii}\, x_i - \sum_{j=1}^{N} (i-j)\,(x_i - x_j) + [B_1]_{ii}\, d_i + [B_2]_{ii}\, u_i$$

# Block sparse vs. sparse

## block sparse design:




## sparse design:




**small difference in performance:** $\dfrac{J_{\mathrm{bs}} - J_{\mathrm{s}}}{J_{\mathrm{s}}} < 1\%$

# ALGORITHM: DETAILS

# Separability of $G$-minimization problem

$$\underset{G}{\text{minimize}} \quad \gamma\, g(G) \;+\; \frac{\rho}{2}\,\|G - V\|_F^2$$

$$V \;:=\; F^{k+1} \;+\; (1/\rho)\Lambda^k$$

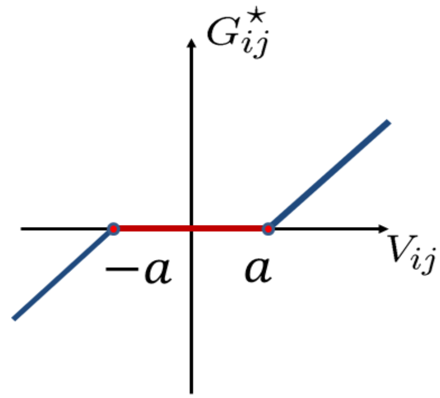
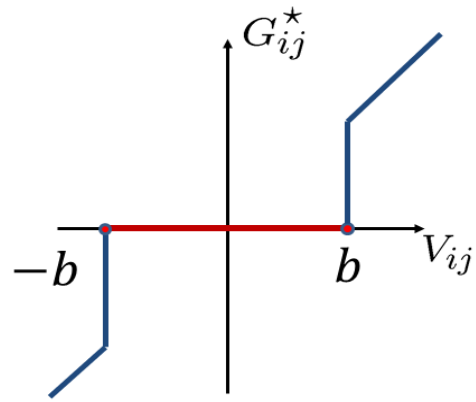
**weighted $\ell_1$:** $\underset{G_{ij}}{\text{minimize}} \;\; \sum_{i,\,j} \left(\gamma\, W_{ij}\, |G_{ij}| \;+\; \frac{\rho}{2}\,(G_{ij} \;-\; V_{ij})^2\right)$

**sum-of-logs:** $\underset{G_{ij}}{\text{minimize}} \;\; \sum_{i,\,j} \left(\gamma \log\left(1 \;+\; \frac{|G_{ij}|}{\varepsilon}\right) \;+\; \frac{\rho}{2}\,(G_{ij} \;-\; V_{ij})^2\right)$

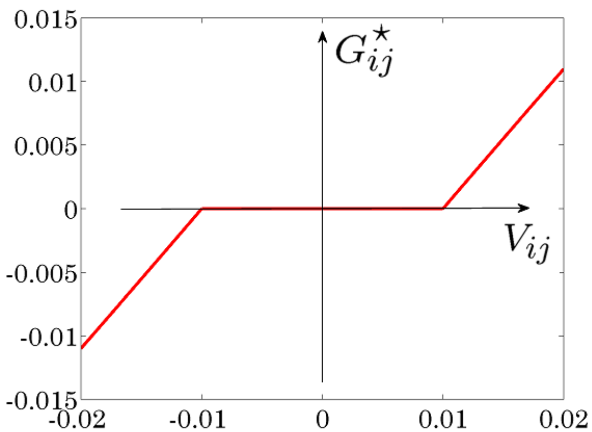
**cardinality:** $\underset{G_{ij}}{\text{minimize}} \;\; \sum_{i,\,j} \left(\gamma\, \mathbf{card}\,(G_{ij}) \;+\; \frac{\rho}{2}\,(G_{ij} \;-\; V_{ij})^2\right)$

**separability** $\Rightarrow$ **element-wise analytical solution**

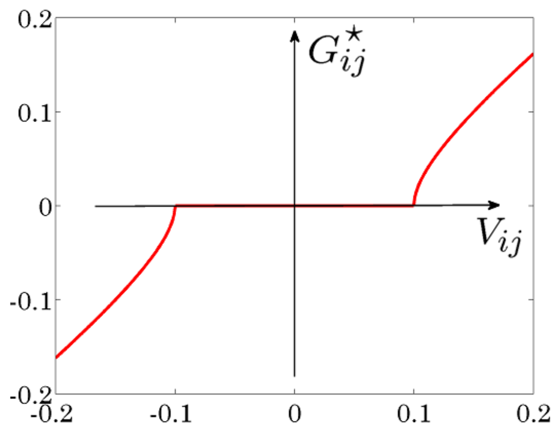
# Solution to $G$-minimization problem

| weighted $\ell_1$: shrinkage | cardinality: truncation |
|---|---|
|  |  |
| $a = (\gamma/\rho) W_{ij}$ | $b = \sqrt{2\gamma/\rho}$ |

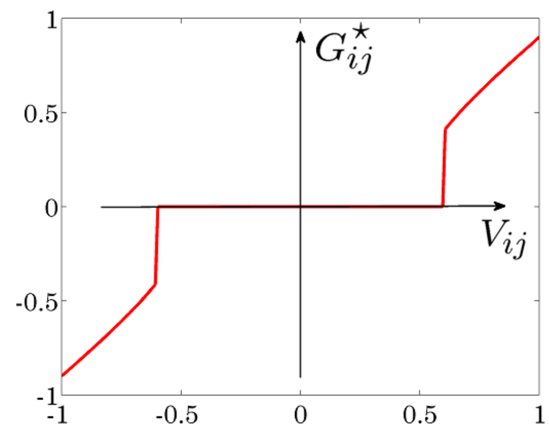
**sum-of-logs** (with $\rho = 100$, $\varepsilon = 0.1$):




$\gamma = 0.1$




$\gamma = 1$




$\gamma = 10$

# Solution to $F$-minimization problem

$$\underset{F}{\text{minimize}} \quad J(F) \,+\, \frac{\rho}{2}\,\|F - U\|_F^2$$

$$U \;:=\; G^k \;-\; (1/\rho)\Lambda^k$$

NECESSARY CONDITIONS FOR OPTIMALITY:

$$\boxed{\begin{aligned} (A - B_2F)L + L(A - B_2F)^T &= -B_1B_1^T \\ (A - B_2F)^TP + P(A - B_2F) &= -(Q + F^TRF) \\ FL + \rho(2R)^{-1}F &= R^{-1}B_2^TPL + \rho(2R)^{-1}U \end{aligned}}$$

- ITERATIVE SCHEME

$$\boxed{\text{Given} \quad F_0 \quad \text{solve for} \quad \{L_1, P_1\} \;\rightarrow\; F_1 \;\rightarrow\; \{L_2, P_2\} \;\rightarrow\; F_2 \;\cdots}$$

**descent direction** $+$ **line search** $\Rightarrow$ **convergence**

# Summary and outlook

- SPARSITY-PROMOTING OPTIMAL CONTROL
  - ⋆ Performance vs. sparsity tradeoff

    *Lin, Fardad, Jovanović, IEEE TAC '12* (submitted; arXiv:1111.6188v2)

- RELATED WORK
  - ⋆ Leader selection in large dynamic networks

    *Lin, Fardad, Jovanović, IEEE TAC '12* (submitted)
  - ⋆ Optimal synchronization of sparse oscillator networks

    *Fardad, Lin, Jovanović, ACC '12*
  - ⋆ Optimal dissemination of information in social networks

    *Fardad, Zhang, Lin, Jovanović, CDC '12* (to appear)
  - ⋆ Wide-area control of power systems

    *Dörfler, Jovanović, Chertkov, Bullo, ACC '13* (submitted)

- SOFTWARE
  - ⋆ www.umn.edu/∼mihailo/software/lqrsp/

```
>> solpath = lqrsp(A,B1,B2,Q,R,options);
```

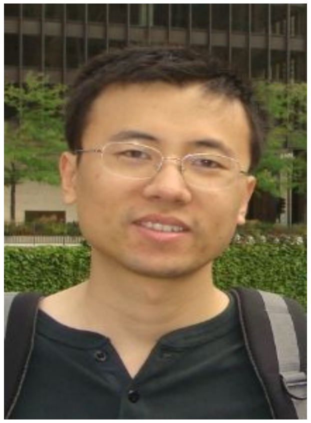
- ONGOING RESEARCH
  - ⋆ Design of sparse optimal estimators
  - ⋆ Observer-based sparse optimal design
  - ⋆ Identification of sparse dynamics
  - ⋆ Finite horizon problems

- WISH LIST
  - ⋆ Performance bounds on structured feedback design
  - ⋆ Distributed implementation of ADMM
  - ⋆ Identification of convex problems

# Acknowledgments

TEAM:

Makan Fardad
(Syracuse University)



Fu Lin
(U of M)

SUPPORT:

NSF CAREER Award CMMI-06-44793

NSF Award CMMI-09-27720

DISCUSSIONS:

Stephen Boyd

Roland Glowinski