Sparsity-Promoting Optimal Control of Distributed Systems

Mihailo Jovanović
www.umn.edu/∼mihailo

joint work with:
Makan Fardad
Fu Lin

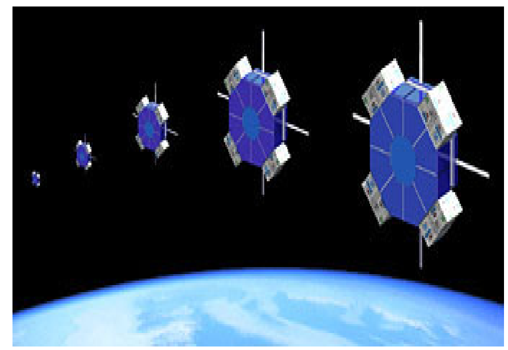
UNIVERSITY
OF MINNESOTA

2013 European Control Conference

Large dynamic networks

- OF INCREASING IMPORTANCE IN MODERN TECHNOLOGY

APPLICATIONS:

wind farms	power networks	aircraft formations satellite constellations
		

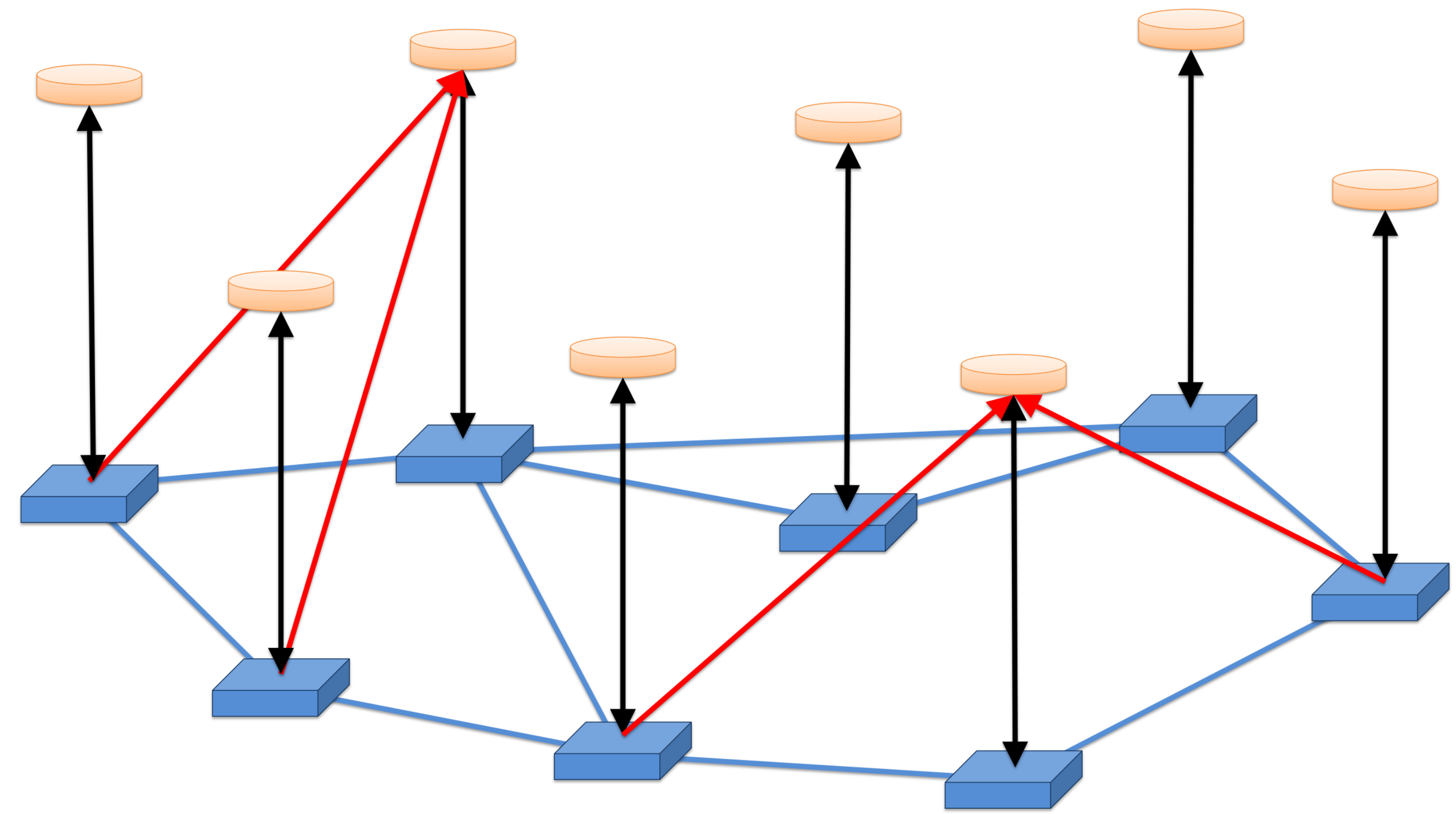
- INTERACTIONS CAUSE COMPLEX BEHAVIOR
  - ⋆ **cannot be predicted by analyzing isolated subsystems**

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

Structured distributed control

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

structured memoryless controller



KEY CHALLENGE:

identification of a signal exchange network

performance vs sparsity

State-feedback H_2 problem

dynamics: $$\dot{x} \;=\; A\,x \;+\; B_1\,d \;+\; B_2\,u$$

objective function: $$J \;=\; \lim_{t\,\to\,\infty} \mathcal{E}\left(x^T(t)\,Q\,x(t) \;+\; u^T(t)\,R\,u(t)\right)$$

memoryless controller: $$u \;=\; -F\,x$$

- Closed-loop H_2 norm

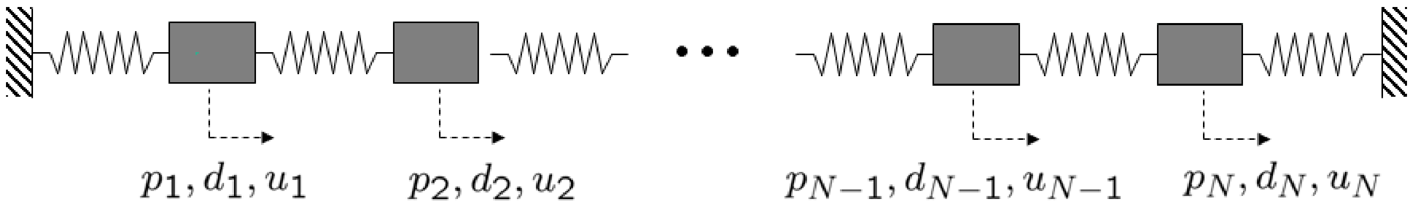
$$J(F) \;=\; \mathrm{trace}\left(\int_0^\infty \mathrm{e}^{(A-B_2F)^T t}\left(Q \;+\; F^T R F\right)\mathrm{e}^{(A-B_2F)t}\,\mathrm{d}t\,B_1 B_1^T\right)$$

⋆ **no structural constraints**

globally optimal controller:

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

An example: Mass-spring system



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

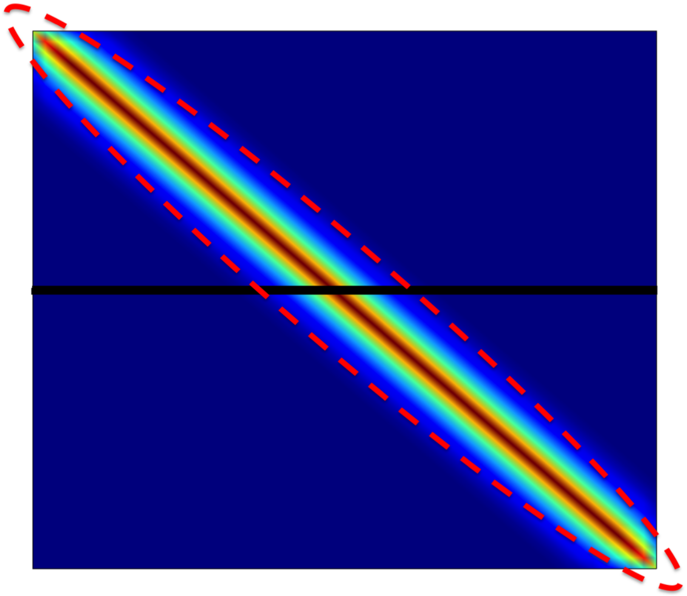
- **Objective: design** $\begin{bmatrix} F_p & F_v \end{bmatrix}$ **to minimize steady-state variance of** p, v, u

Optimal controller – Linear Quadratic Regulator

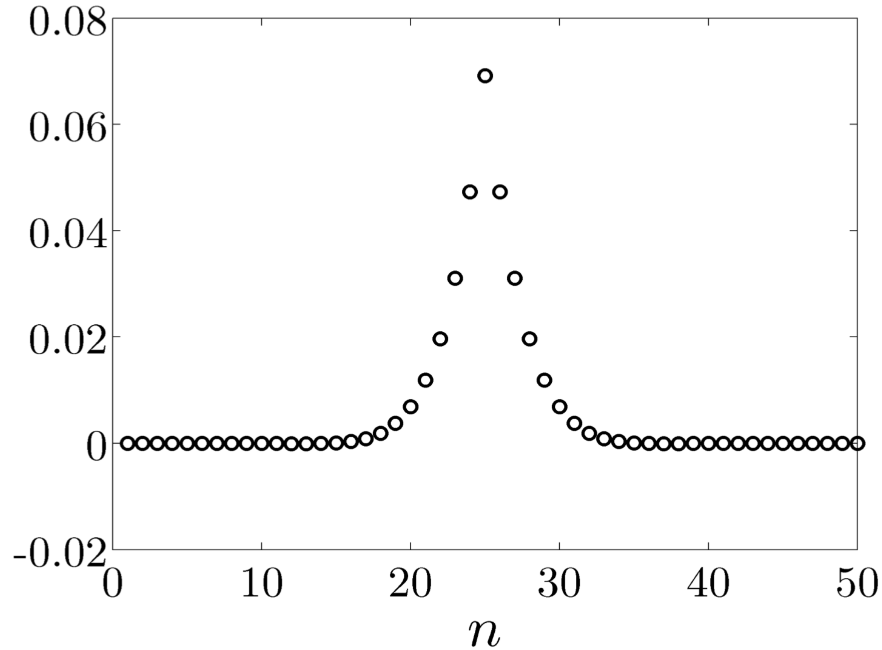
$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \end{bmatrix} = -\underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}}_{F_p} \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \end{bmatrix} - \underbrace{\begin{bmatrix} * & * & * & * \\ *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

Structure of optimal controller

position feedback matrix:



position gains for middle mass:



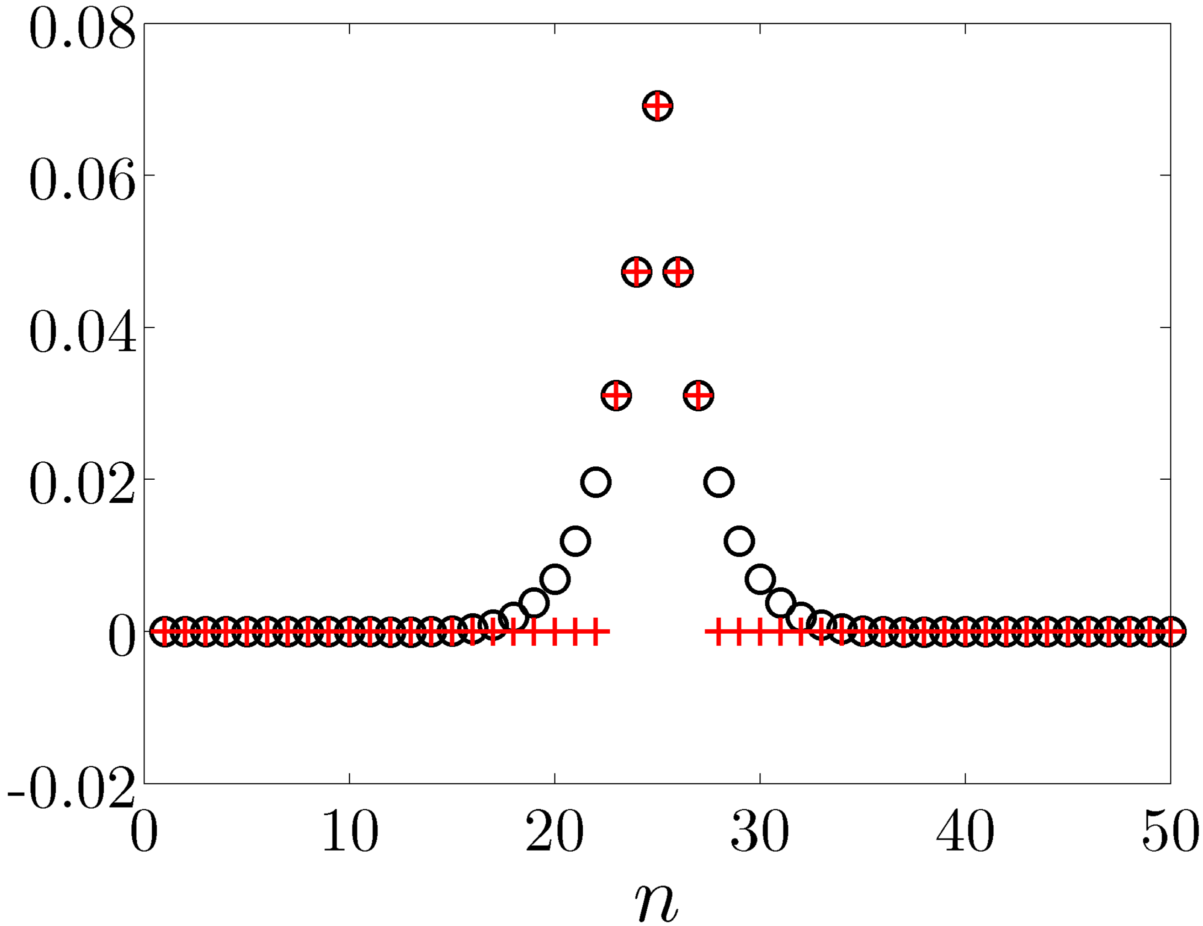
- OBSERVATIONS
  - ⋆ Diagonals almost constant (modulo edges)
  - ⋆ Off-diagonal decay of a centralized gain

Bamieh, Paganini, Dahleh, IEEE TAC '02

Motee & Jadbabaie, IEEE TAC '08

Enforcing localization?

- **One approach: truncating centralized controller**



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

Outline

❶ SPARSITY-PROMOTING OPTIMAL CONTROL

⋆ **Design of 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

$$\text{minimize} \quad \underbrace{J(F)}_{\substack{\downarrow \\ \textbf{variance amplification}}} \quad + \quad \underbrace{\gamma \sum_{i,\,j} W_{ij}\,|F_{ij}|}_{\substack{\downarrow \\ \textbf{sparsity-promoting penalty function}}}$$

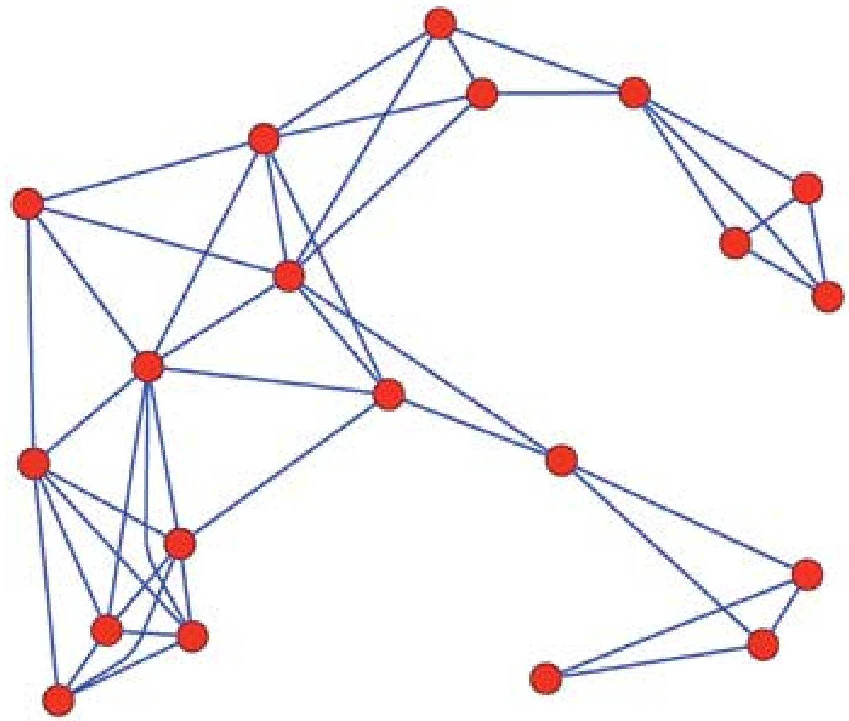
$\star\ \gamma > 0$ – performance vs sparsity tradeoff

$\star\ W_{ij} \geq 0$ – weights (for additional flexibility)

Lin, Fardad, Jovanović, IEEE TAC '13 (in press; arXiv:1111.6188)

A CLASS OF CONVEX PROBLEMS

Consensus by distributed computation



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

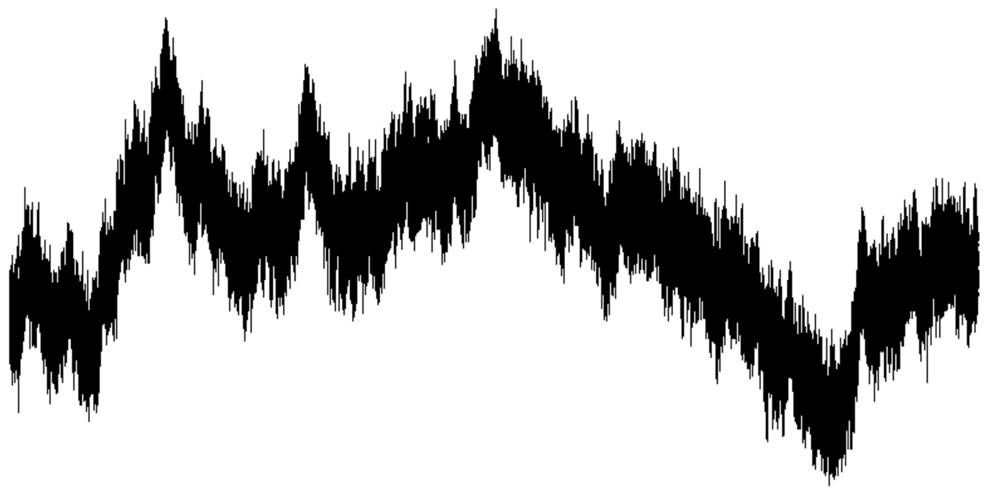
$$\dot{x}_i(t) \;=\; -\sum_{j\,\in\mathcal{N}_i}\Big(x_i(t)\;-\;x_j(t)\Big)$$

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:** $\bar{x}(t) \;=\; \dfrac{1}{N}\sum_{i\,=\,1}^{N} x_i(t)$: 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: $\lim_{t\,\to\,\infty} \mathcal{E}\left(\tilde{x}^T(t)\,\tilde{x}(t)\right)$

Design of undirected consensus networks

dynamics: $\dot{x} \;=\; d \;+\; u$

objective function: $J \;=\; \lim_{t \,\to\, \infty} \mathcal{E}\left(x^T(t)\, Q\, x(t) \;+\; u^T(t)\, R\, u(t)\right)$

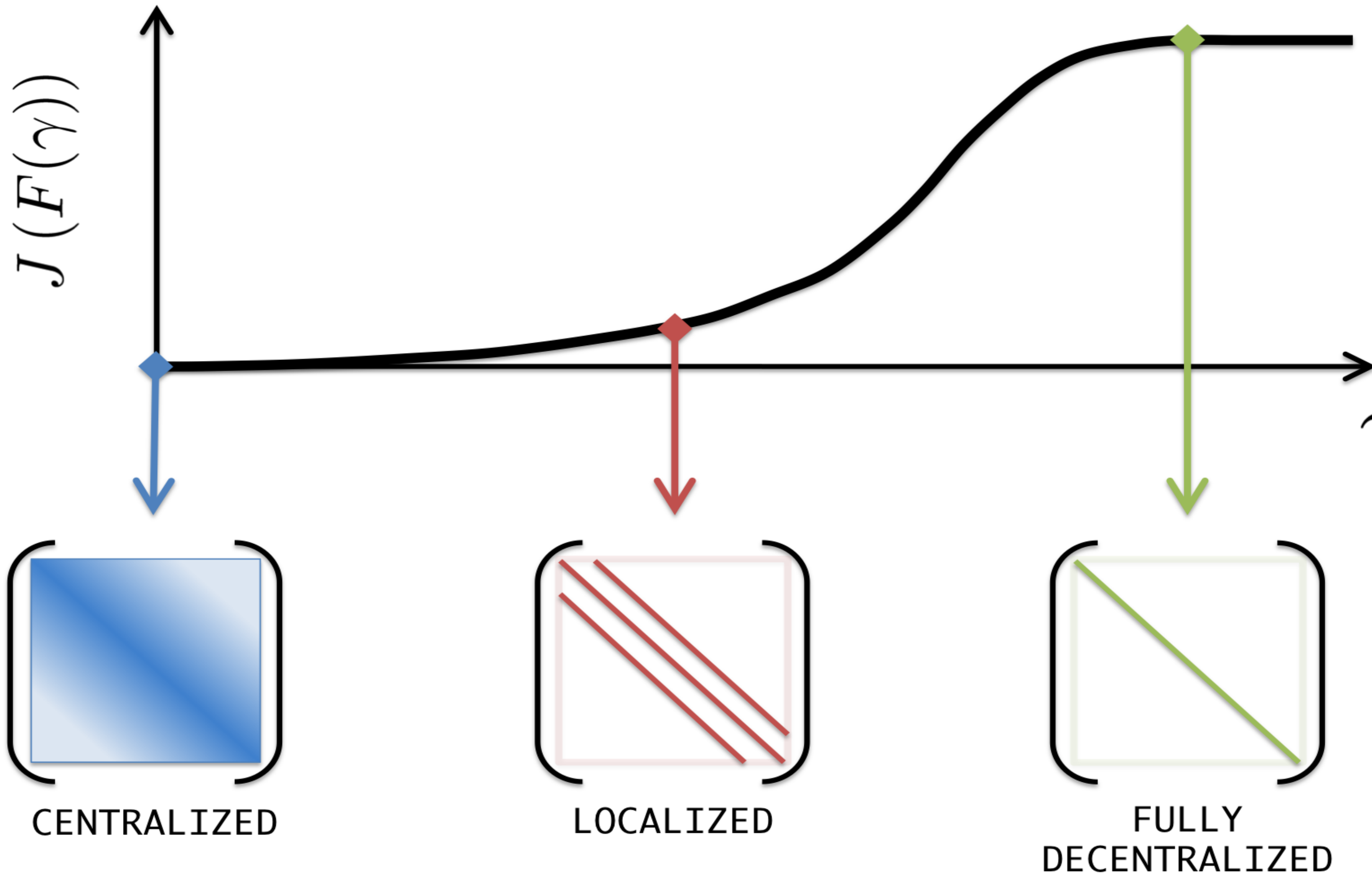
performance weights: $Q \;\succeq\; 0, \;\; R \;\succ\; 0$

can be formulated as an SDP:

$$
\begin{array}{ll}
\text{minimize} & \operatorname{trace}\,(X \;+\; R\,F) \;+\; \gamma\,\mathbb{1}^T\,Y\,\mathbb{1} \\[1em]
\text{subject to} & \begin{bmatrix} X & Q^{1/2} \\ Q^{1/2} & F + \mathbb{1}\mathbb{1}^T/N \end{bmatrix} \;\succeq\; 0 \\[1em]
& -Y_{ij} \;\leq\; W_{ij}\,F_{ij} \;\leq\; Y_{ij} \\[1em]
& F\,\mathbb{1} \;=\; 0
\end{array}
$$

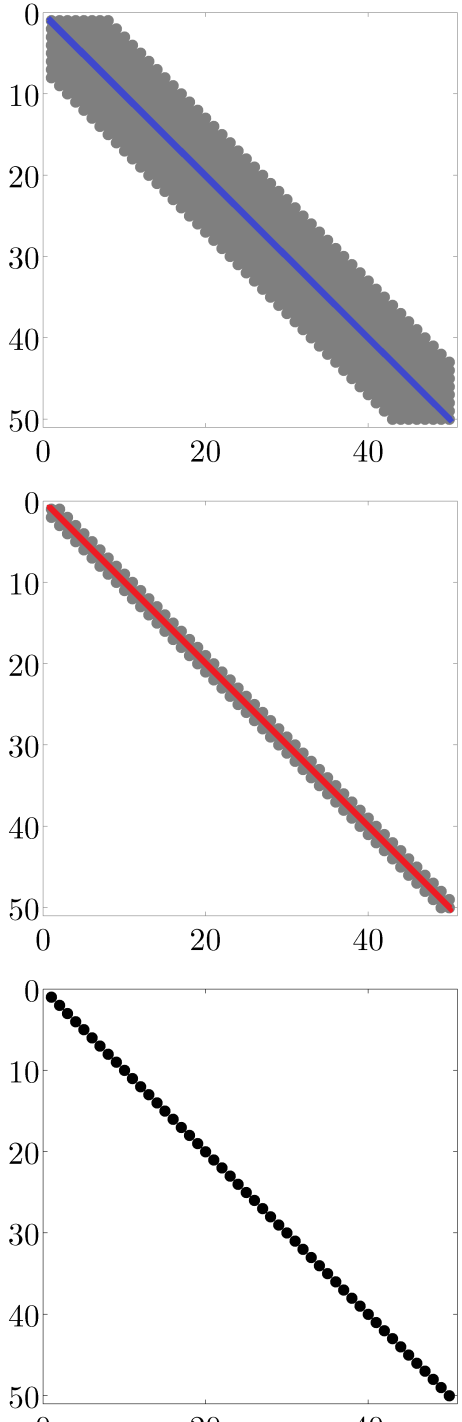
Parameterized family of feedback gains

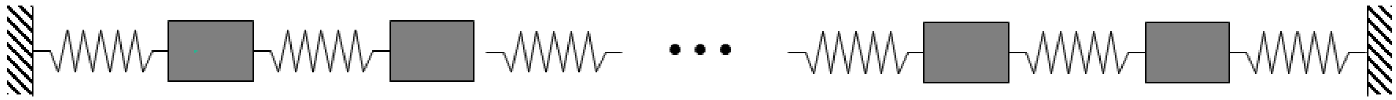
$$F(\gamma) \; := \; \underset{F}{\arg\min} \, (J(F) \; + \; \gamma \, g(F))$$



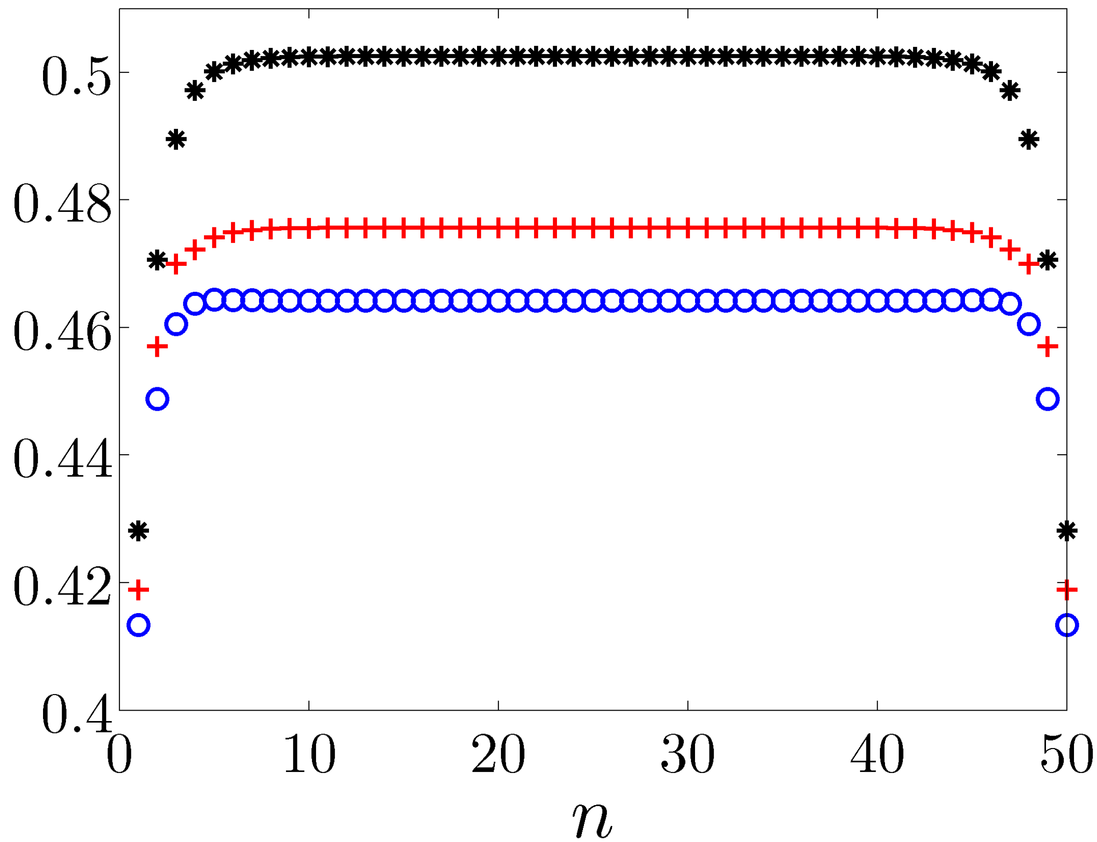
Examples

Mass-spring system



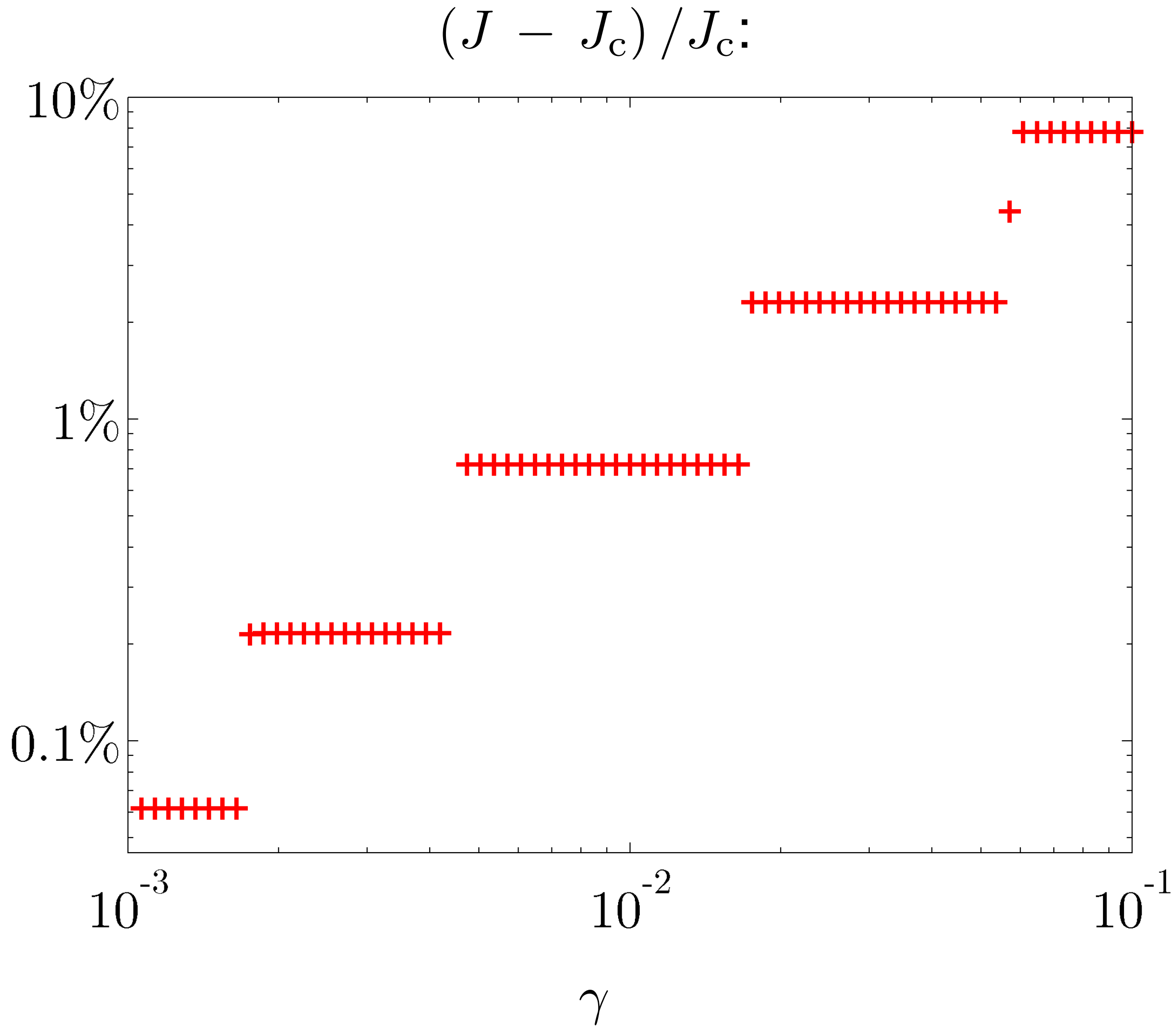


$\mathrm{diag}\,(F_v)$:



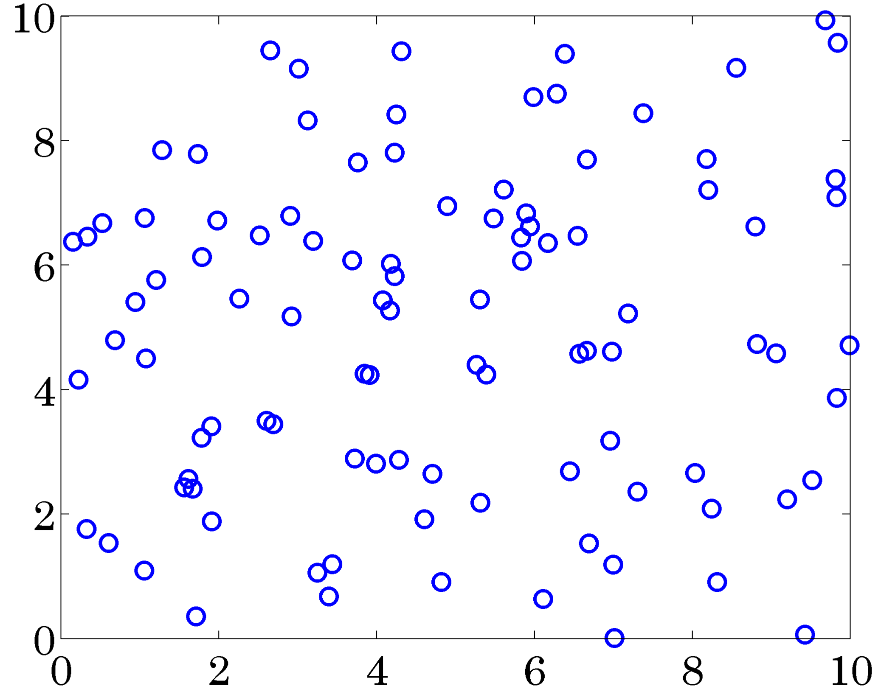
$\gamma = 10^{-4}$ $\gamma = 0.03$ $\gamma = 0.1$

- **Performance comparison: sparse vs centralized**



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

Network with 100 nodes

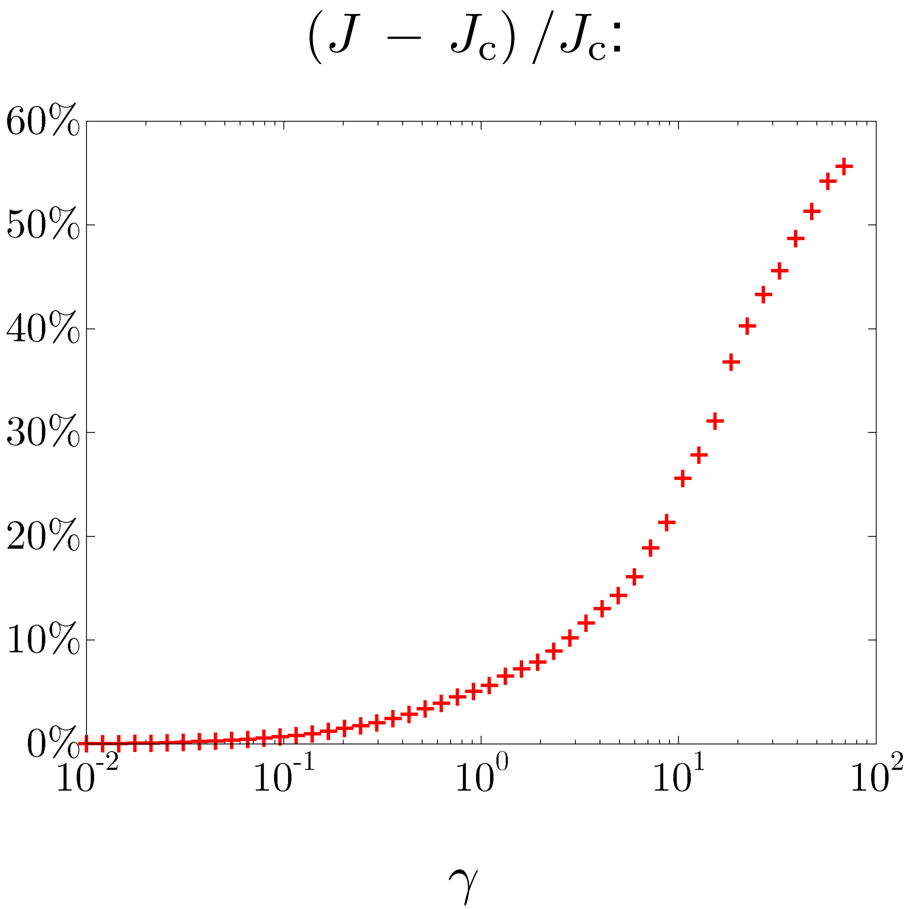


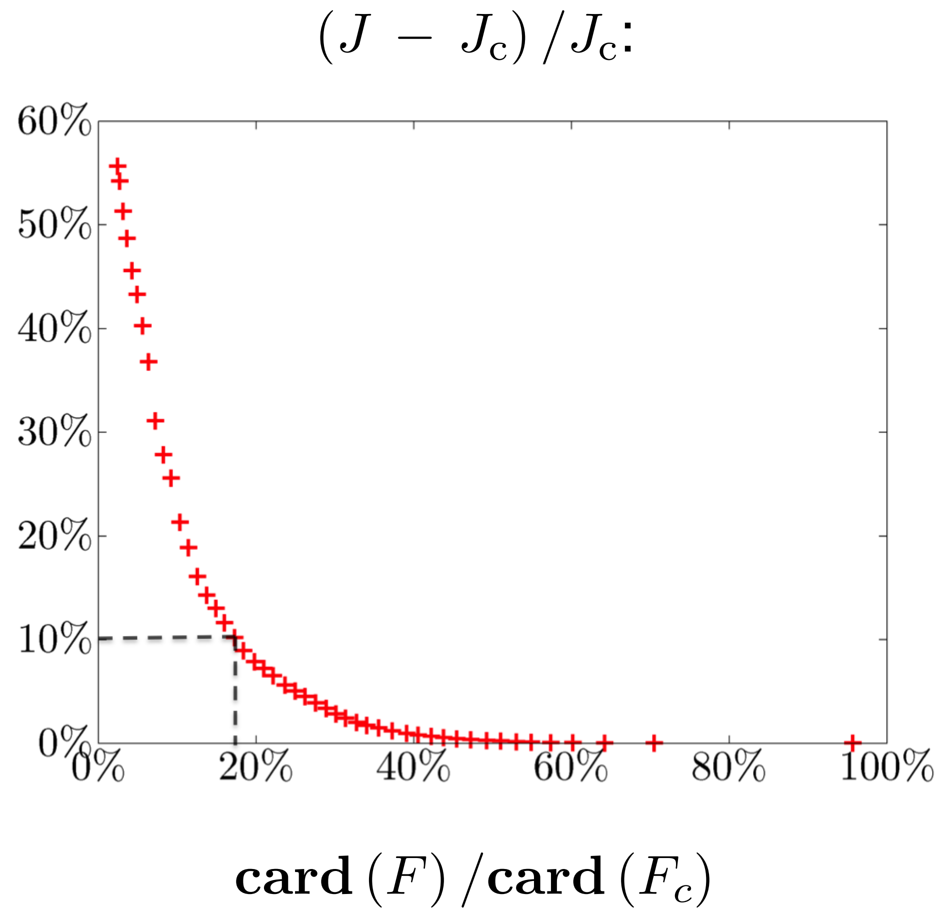
$$\begin{bmatrix} \dot{p}_i \\ \dot{v}_i \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} p_i \\ v_i \end{bmatrix}}_{\textbf{unstable dynamics}} + \underbrace{\sum_{j \neq i} \mathrm{e}^{-\alpha(i,j)} \begin{bmatrix} p_j \\ v_j \end{bmatrix}}_{\textbf{coupling}}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

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

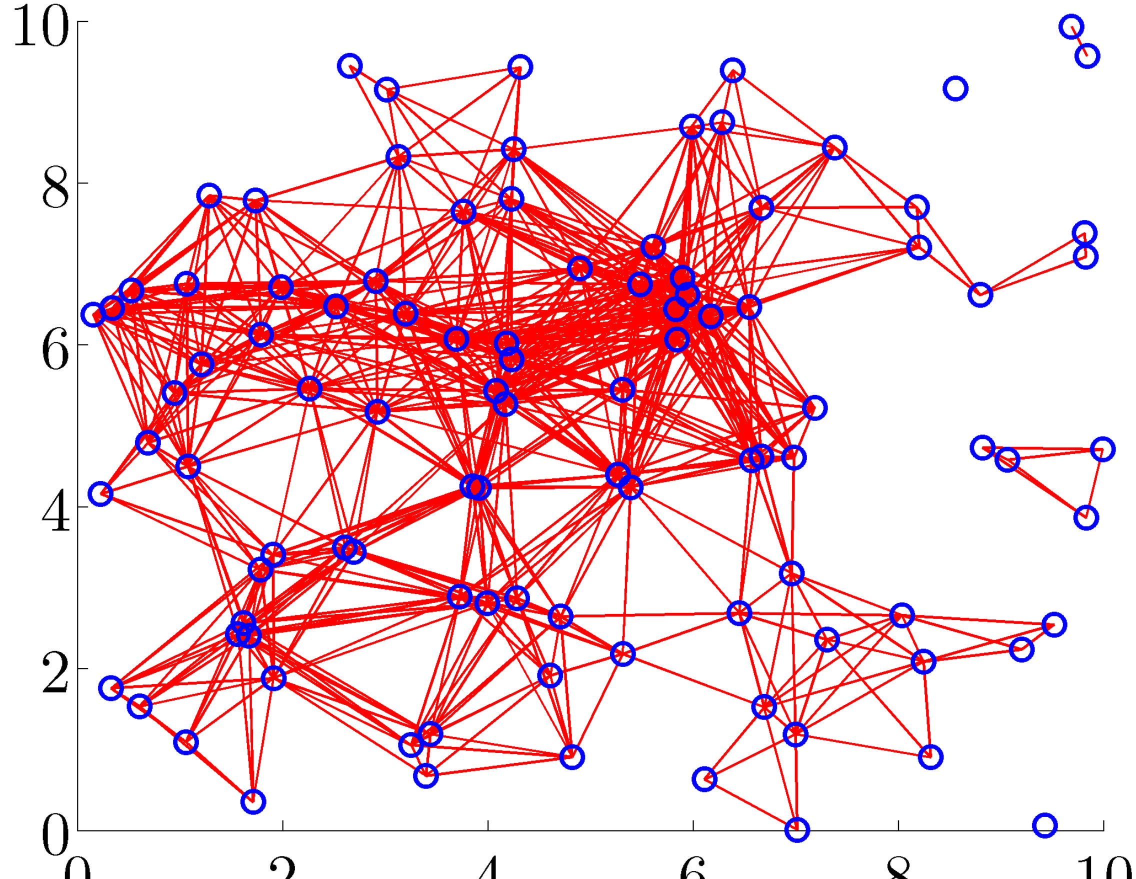
Motee & Jadbabaie, IEEE TAC '08

- **Performance comparison: sparse vs centralized**



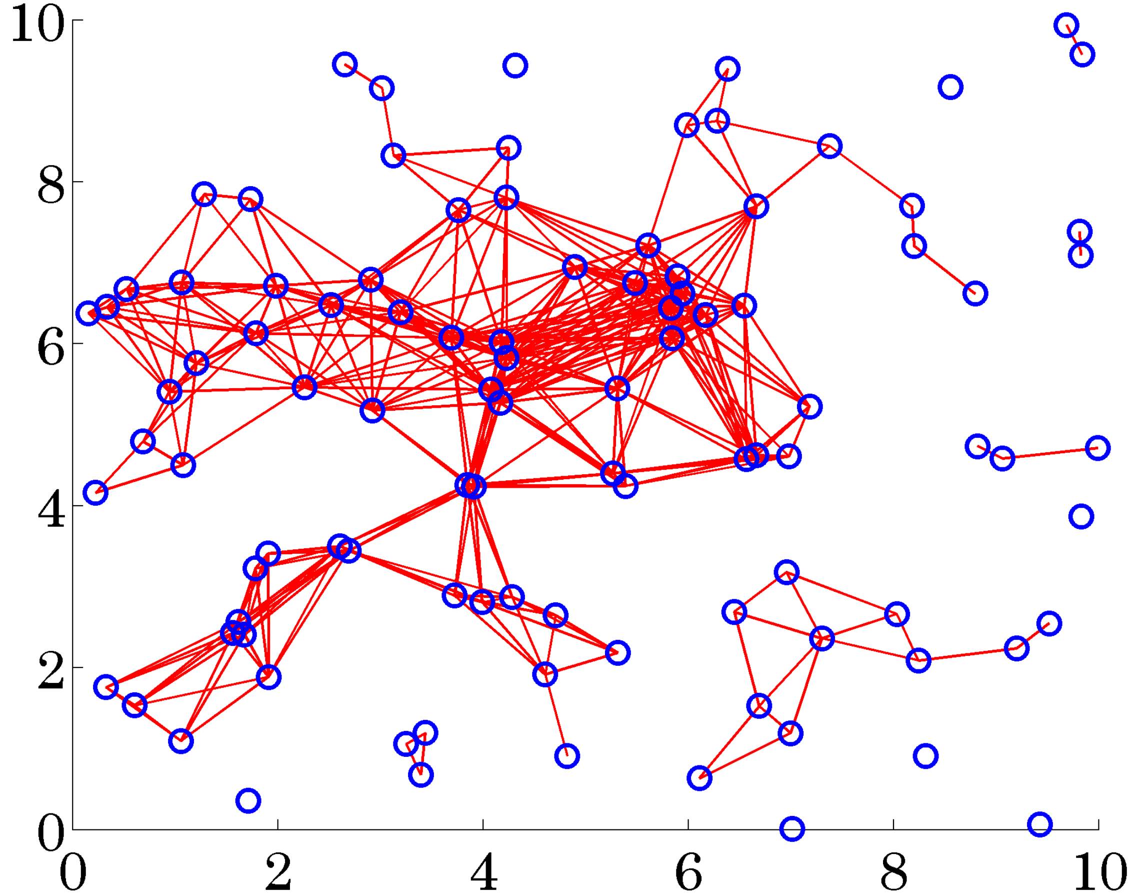


identified communication graph:



$$\begin{aligned} \gamma &= 5 \\ \mathbf{card}\,(F)\,/\mathbf{card}\,(F_{\mathrm{c}}) &= 8.8\% \\ (J - J_{\mathrm{c}})\,/J_{\mathrm{c}} &= 24.6\% \end{aligned}$$

identified communication graph:

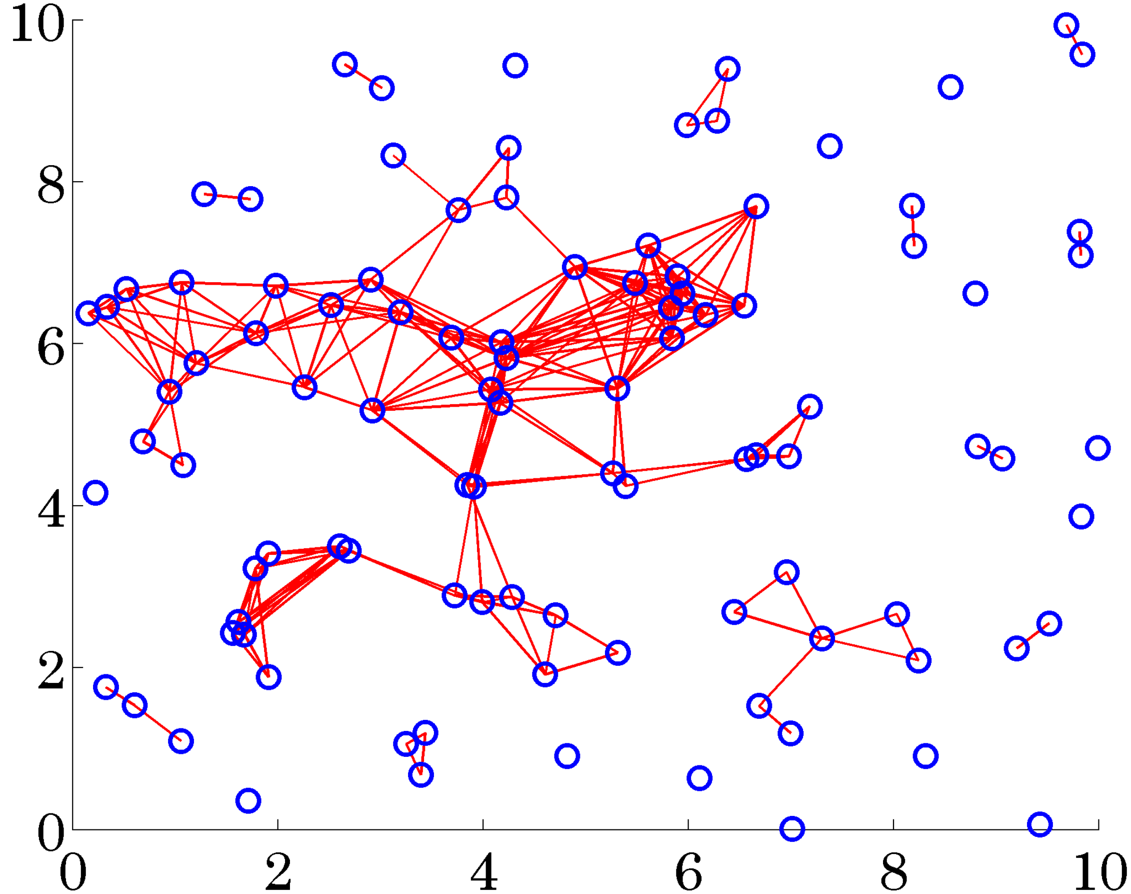


$$\gamma = 11$$

$$\mathbf{card}\,(F)\,/\mathbf{card}\,(F_c) = 5.1\%$$

$$(J - J_c)\,/J_c = 40.9\%$$

identified communication graph:

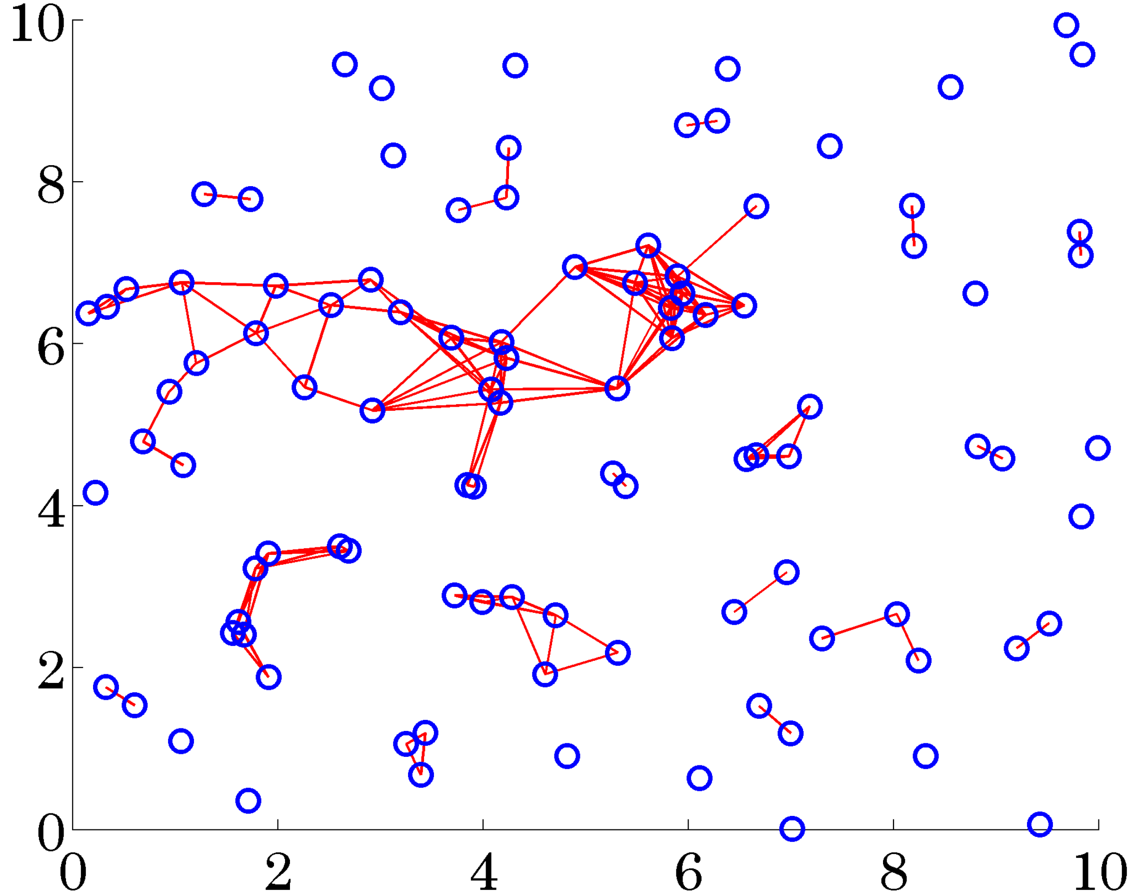


$$\gamma = 18$$

$$\textbf{card}\,(F)\,/\textbf{card}\,(F_{\rm c}) = 3.4\%$$

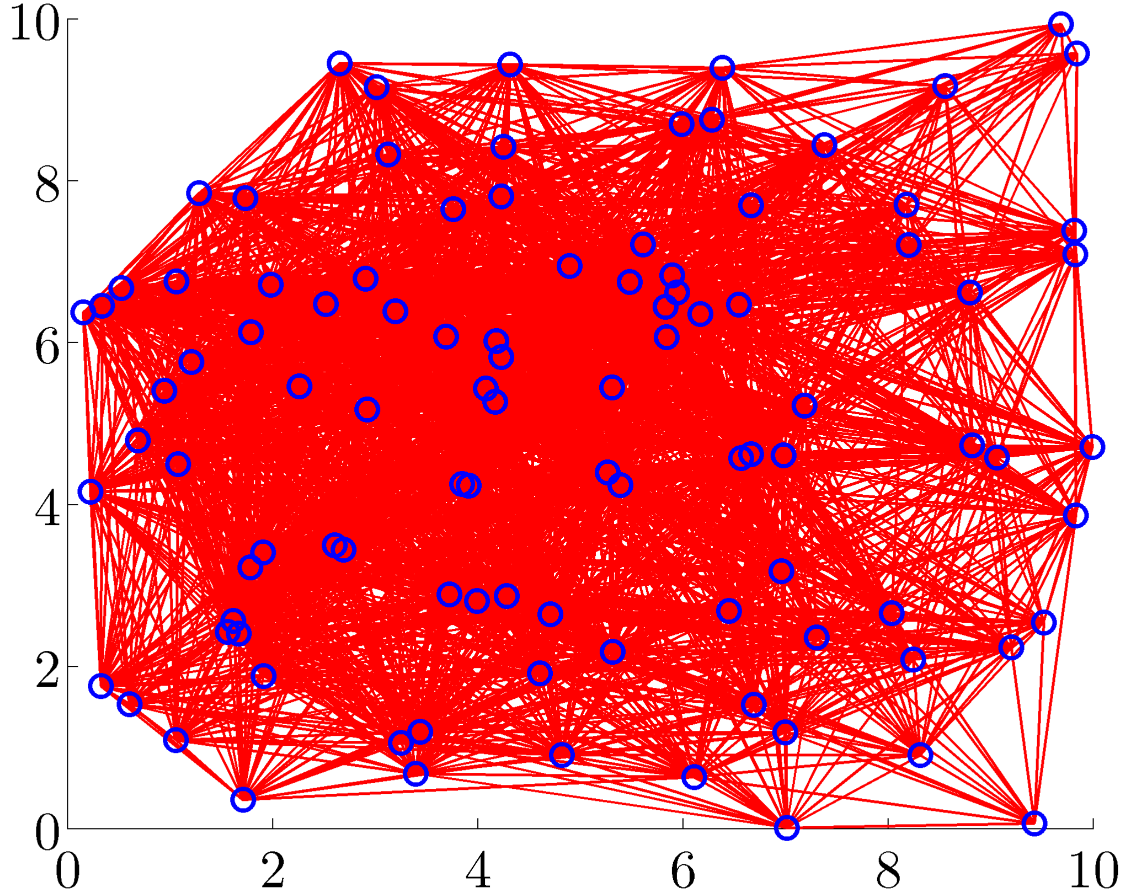
$$(J - J_{\rm c})\,/J_{\rm c} = 48.7\%$$

identified communication graph:



$$\begin{aligned} \gamma &= 30 \\ \mathbf{card}\,(F)\,/\mathbf{card}\,(F_{\mathrm{c}}) &= 2.4\% \\ (J - J_{\mathrm{c}})\,/J_{\mathrm{c}} &= 54.8\% \end{aligned}$$

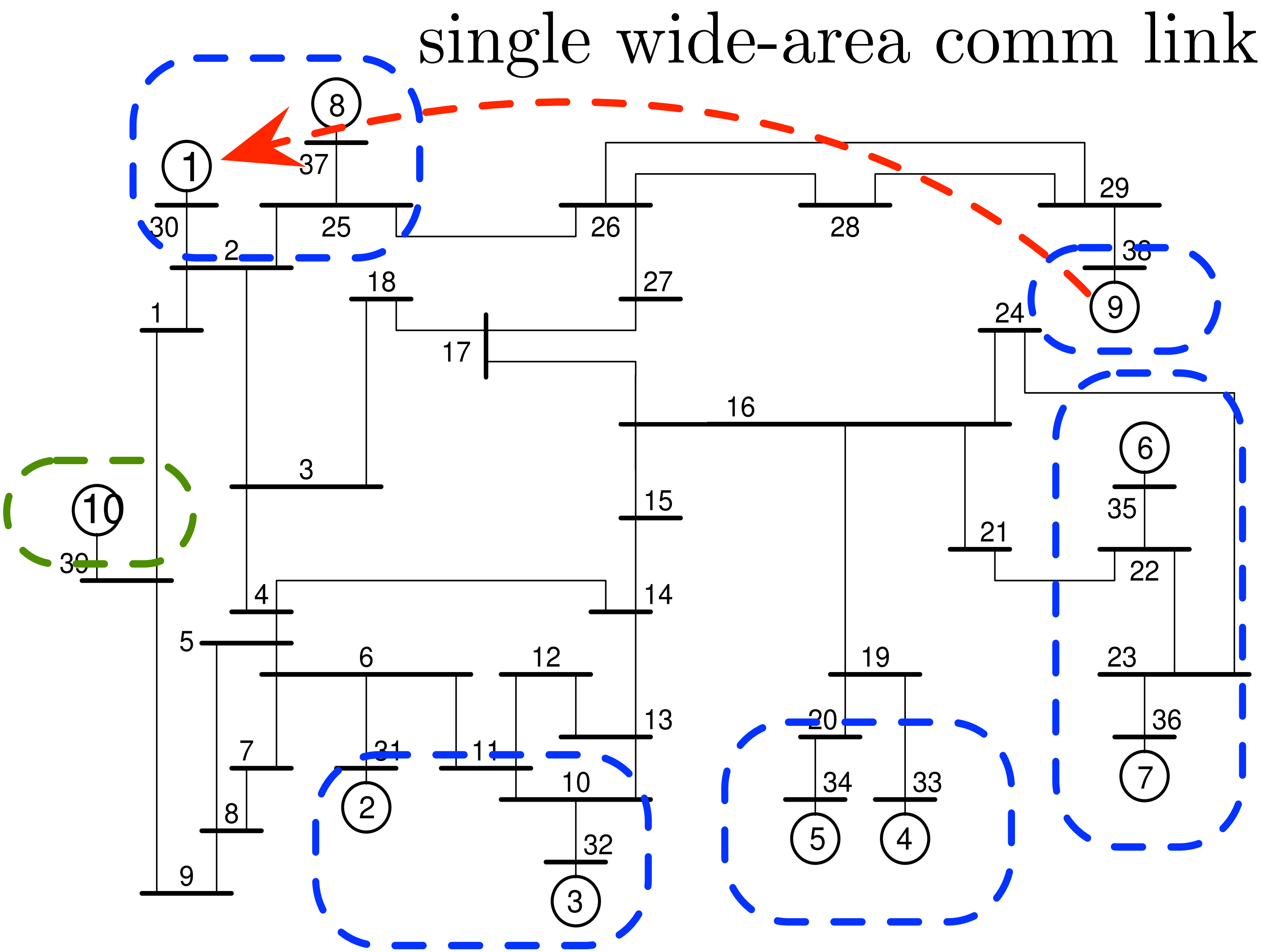
communication graph of a truncated centralized gain:



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

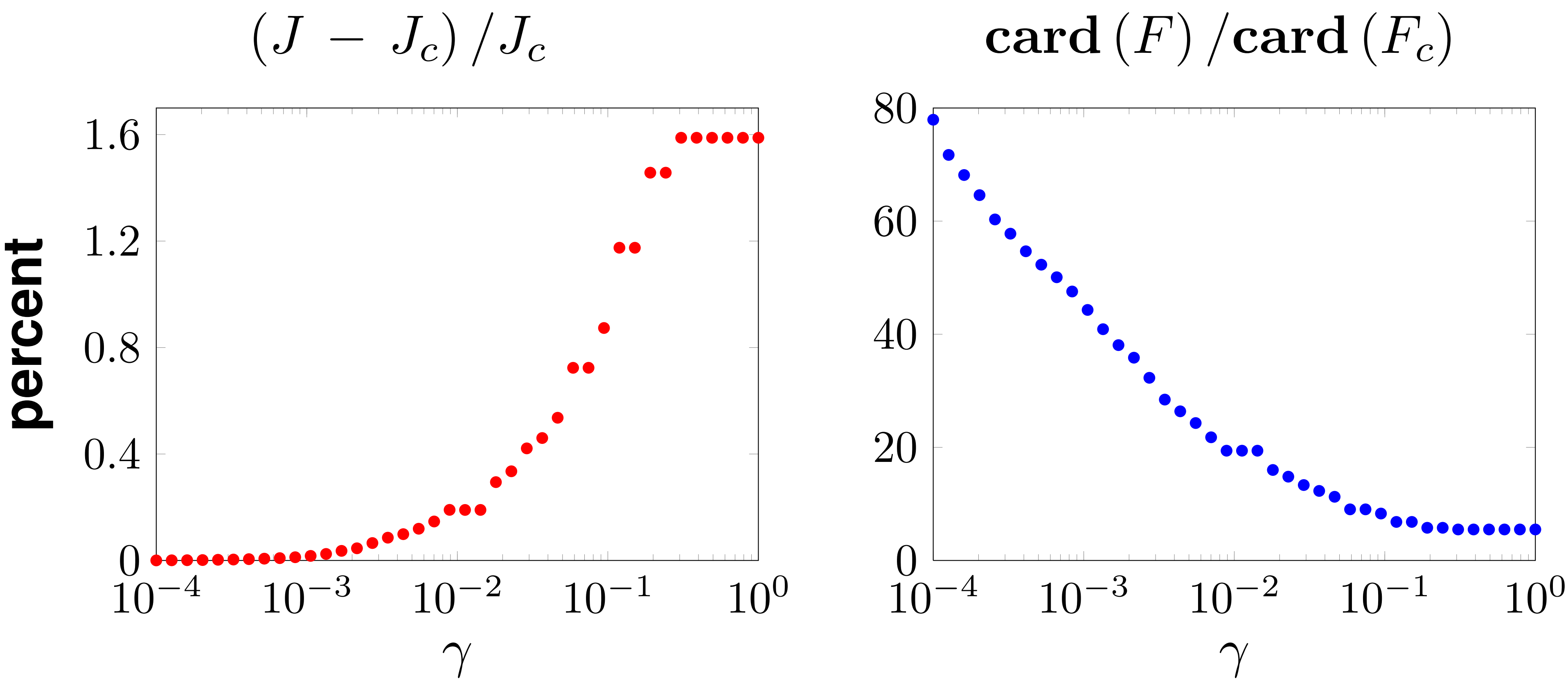
non-stabilizing

Wide area control of power networks



single long range interaction $\Rightarrow$ **nearly centralized performance**

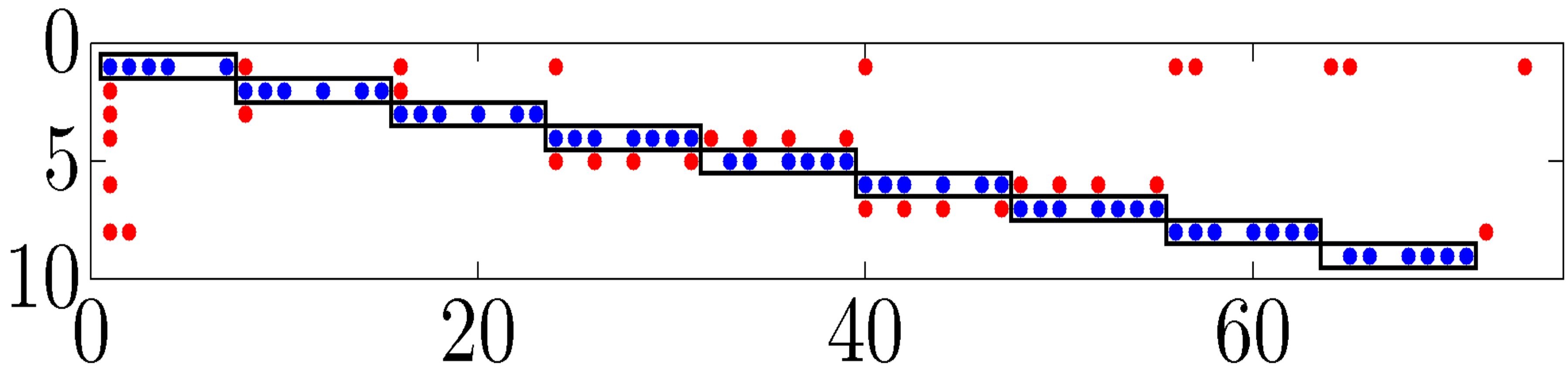
Performance vs sparsity



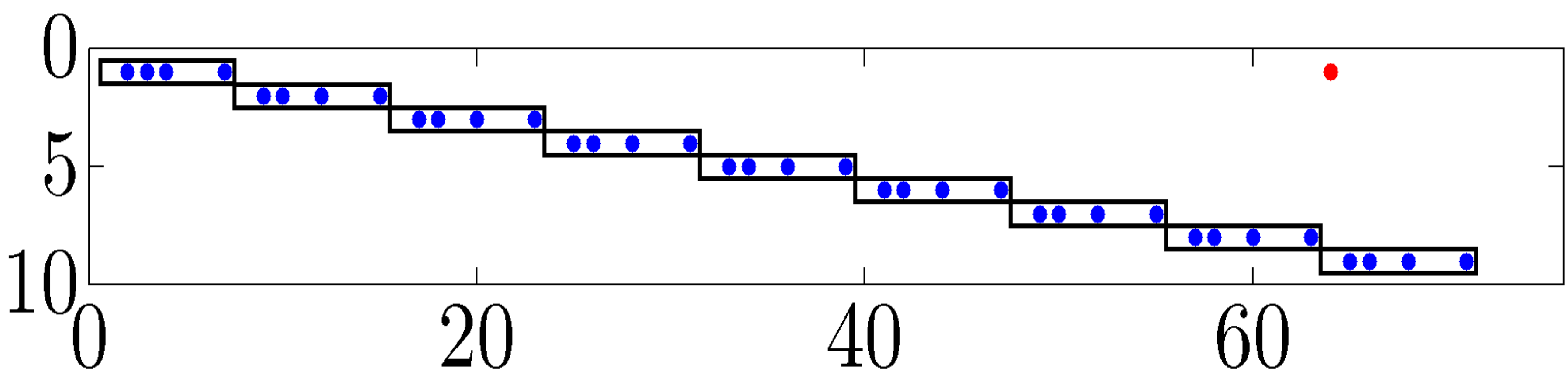
$$\gamma = 1 \xrightarrow{\text{relative to } F_c} \begin{cases} 1.6\,\% & \text{performance loss} \\ 5.5\,\% & \text{non-zero elements in } F \end{cases}$$

- **Signal exchange network**

$$\gamma = 0.0289, \textbf{card}(F) = 90$$



$$\gamma = 1, \textbf{card}(F) = 37$$

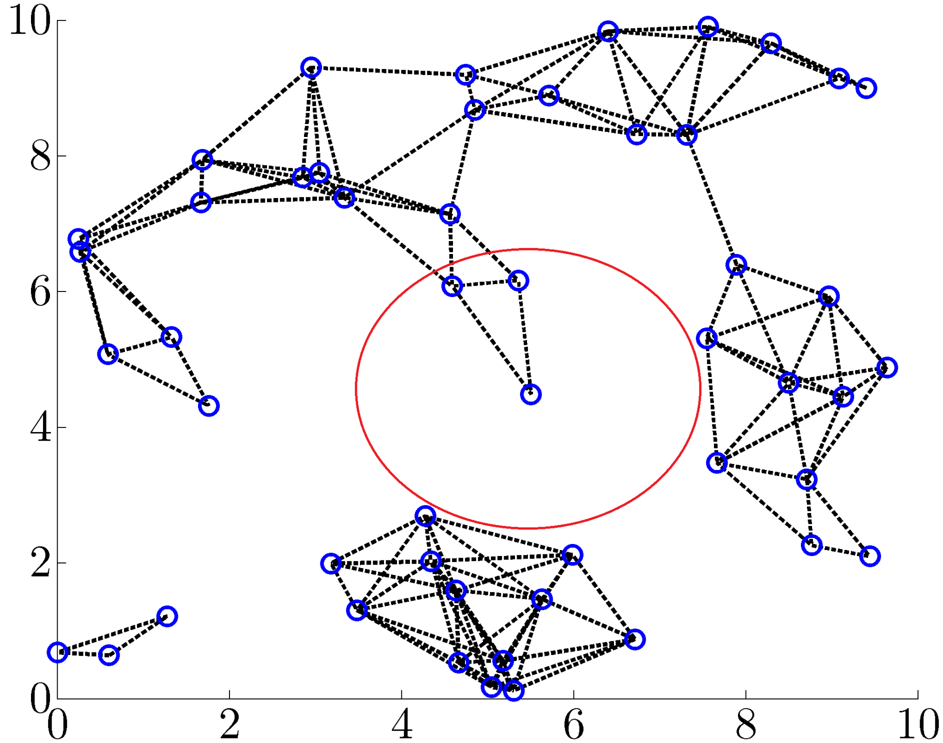


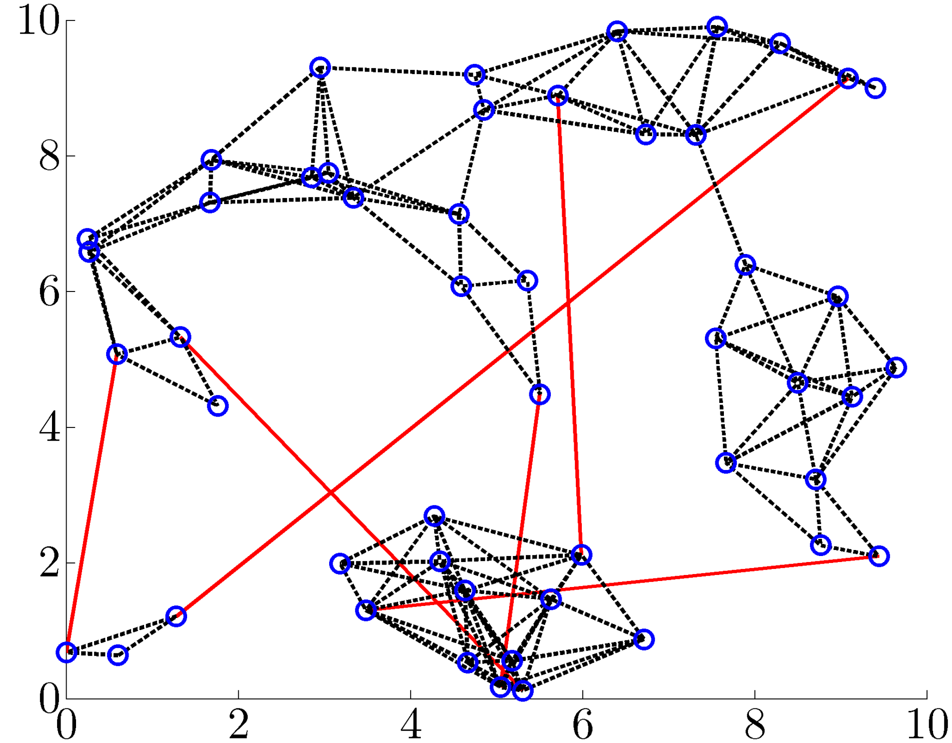
Dörfler, Jovanović, Chertkov, Bullo, IEEE TPS '13 (submitted)

Sparsity-promoting consensus algorithm

local performance graph:

identified communication graph:





$$Q \;=\; Q_{\mathrm{loc}} \,+\, \left(I \,-\, \frac{1}{N} \mathbb{1}\mathbb{1}^T \right)$$

$$\frac{J \,-\, J_{\mathrm{c}}}{J_{\mathrm{c}}} \;\approx\; 11\%$$

ALGORITHM

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**

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

ADMM:

$$\boxed{\begin{array}{rcl} F^{k+1} & := & \displaystyle\operatorname*{arg\,min}_F \; \mathcal{L}_\rho(F, G^k, \Lambda^k) \\[2ex] G^{k+1} & := & \displaystyle\operatorname*{arg\,min}_G \; \mathcal{L}_\rho(F^{k+1}, G, \Lambda^k) \\[2ex] \Lambda^{k+1} & := & \Lambda^k \,+\, \rho\left(F^{k+1} \,-\, G^{k+1}\right) \end{array}}$$

MANY MODERN APPLICATIONS

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

Boyd et al., Foundations and Trends in Machine Learning '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

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

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

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

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

Separability of G-minimization problem

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

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

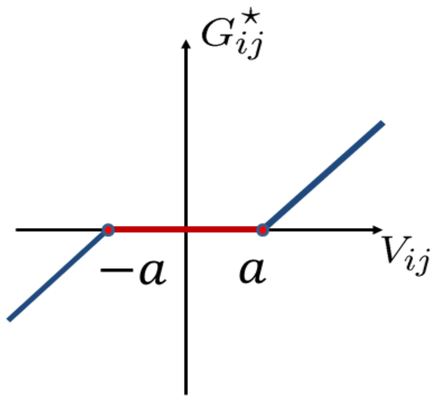
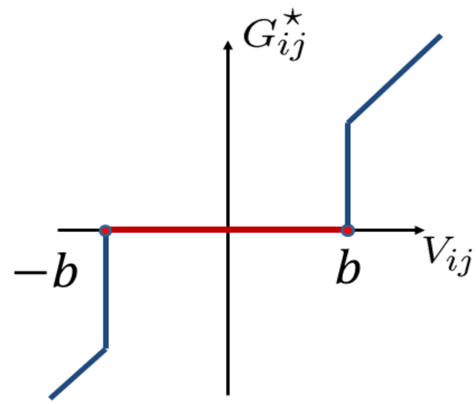
weighted ℓ_1: $\quad \underset{G_{ij}}{\text{minimize}} \quad \sum_{i,\,j} \left(\gamma\, W_{ij}\, |G_{ij}| \;+\; \frac{\rho}{2}\,(G_{ij} \;-\; V_{ij})^2 \right)$

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

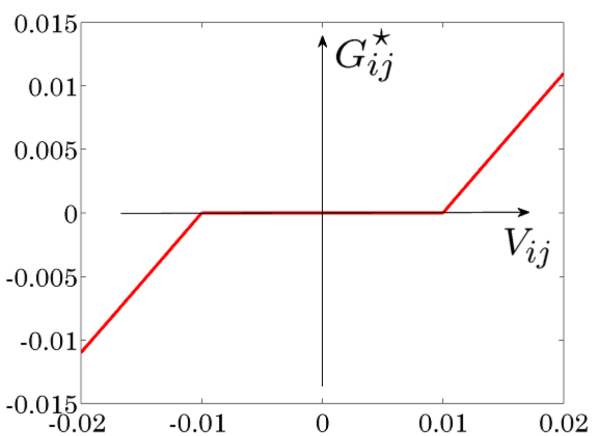
cardinality: $\quad \underset{G_{ij}}{\text{minimize}} \quad \sum_{i,\,j} \left(\gamma\, \mathbf{card}\,(G_{ij}) \;+\; \frac{\rho}{2}\,(G_{ij} \;-\; V_{ij})^2 \right)$

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

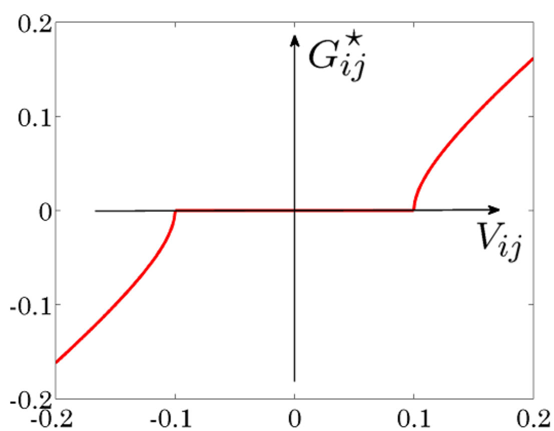
Solution to G-minimization problem

weighted ℓ_1: shrinkage	**cardinality: truncation**
	
$a = (\gamma/\rho) W_{ij}$	$b = \sqrt{2\gamma/\rho}$

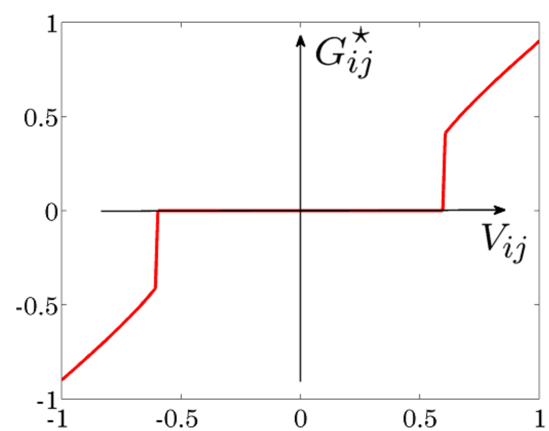
sum-of-logs (with $\rho = 100$, $\varepsilon = 0.1$):



$\gamma = 0.1$



$\gamma = 1$



$\gamma = 10$

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:

$$\boxed{\begin{aligned} (A - B_2 F) L + L (A - B_2 F)^T &= -B_1 B_1^T \\ (A - B_2 F)^T P + P (A - B_2 F) &= -(Q + F^T R F) \\ F L + \rho (2R)^{-1} F &= R^{-1} B_2^T P L + \rho (2R)^{-1} U \end{aligned}}$$

- ITERATIVE SCHEME

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

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

Summary

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

    Lin, Fardad, Jovanović, IEEE TAC ’13 (in press; arXiv:1111.6188)

  - ⋆ Software

    www.umn.edu/∼mihailo/software/lqrsp/

- **ONGOING EFFORT**
  - ⋆ Leader selection in large dynamic networks

    Lin, Fardad, Jovanović, IEEE TAC ’13 (conditionally accepted; arXiv:1302.0450)

  - ⋆ Optimal dissemination of information in social networks

    Fardad, Zhang, Lin, Jovanović, CDC ’12

  - ⋆ Wide-area control of power networks

    Dörfler, Jovanović, Chertkov, Bullo, IEEE TPS ’13 (submitted)

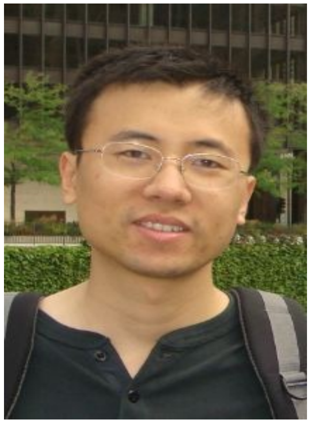
  - ⋆ Sparse or infrequently changing (in time) control signals

    Jovanović & Lin, ECC ’13 (WeC2.4; 17:20 – 17:40)

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

ADDITIONAL SLIDES

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

ℓ_1 **norm:** $$\sum_{i,\,j} |F_{ij}|$$

weighted ℓ_1 **norm:** $$\sum_{i,\,j} W_{ij}\,|F_{ij}|, \quad W_{ij} \;\geq\; 0$$

- CARDINALITY VS WEIGHTED ℓ_1 NORM

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

RE-WEIGHTED SCHEME

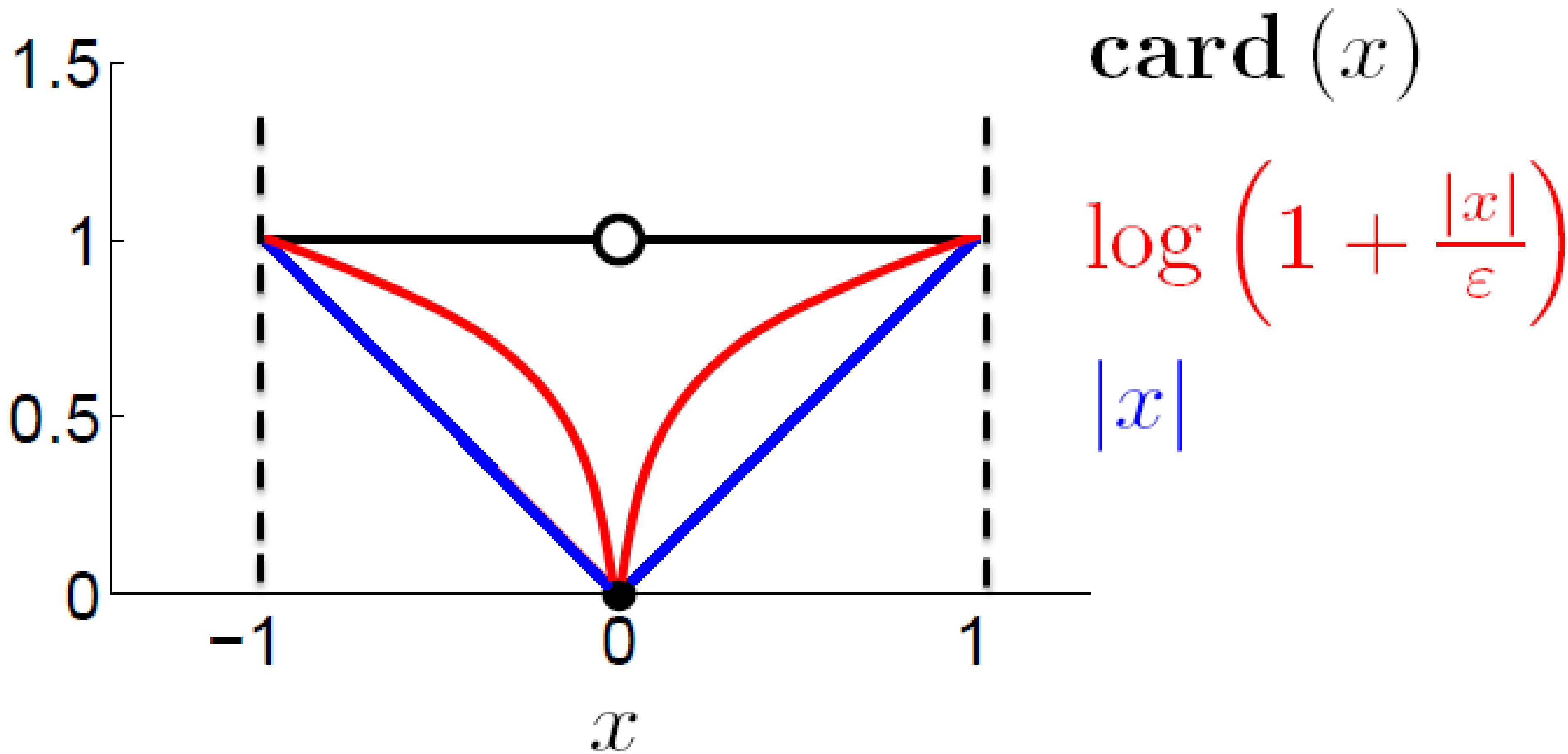
⋆ **Use feedback gains from previous iteration to form weights**

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

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:

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

$\Rightarrow$

relaxation:

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

ℓ_1	**weighted** ℓ_1	**sum-of-logs**