Compressive Sensing via Sparse Sensing

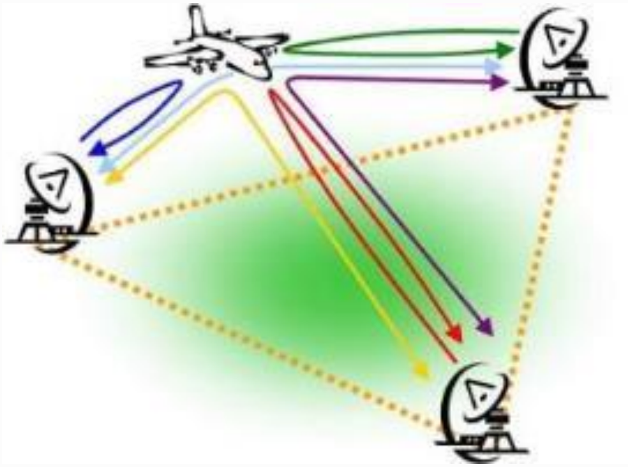
Geert Leus g.j.t.leus@tudelft.nl

Thanks: Sundeep Chepuri, Mario Coutiño, Guillermo Jimenez

29 Aug 2016

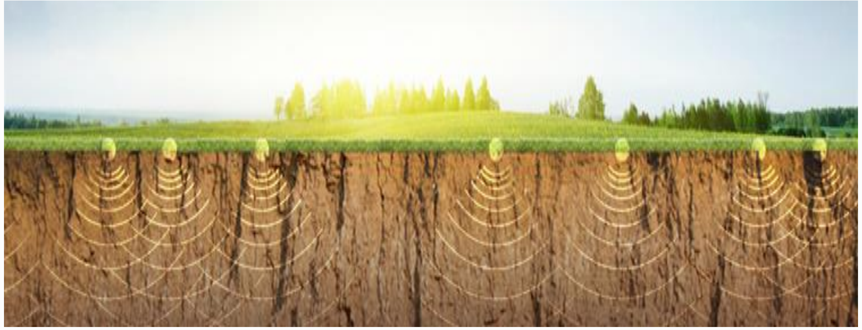
TU Delft Delft University of Technology

Motivation

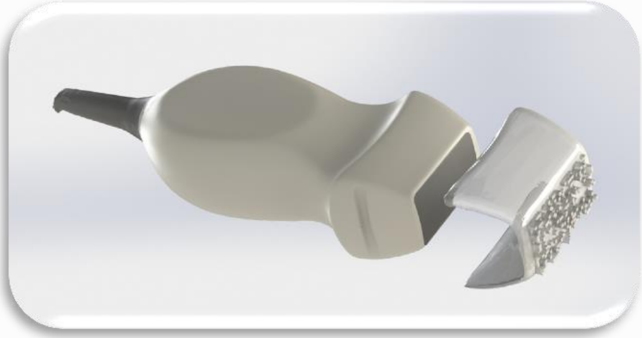


Distributed radar

Radio astronomy



Microseismic event detection



Ultrasound imaging

Problem statement

Design of structured (sparse) space-time samplers

or

sparse sensing design

The term “**sparse sensing = sampling**” has been used earlier:

- Sampling sparse signals *[Vetterli et al.-2008]*
- Covariance reconstruction and array processing *[Vaidyanathan et al.-2011]*

- T. Blu, P.L. Dragotti, M.Vetterli, P. Marziliano, and L. Coulot. “Sparse sampling of signal innovations,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31-40, Mar. 2008.
- P.P. Vaidyanathan and P. Pal. "Sparse sensing with co-prime samplers and arrays." *IEEE Transactions on Signal Processing,* vol. 59, no. 2, pp. 573-586, Feb. 2011.

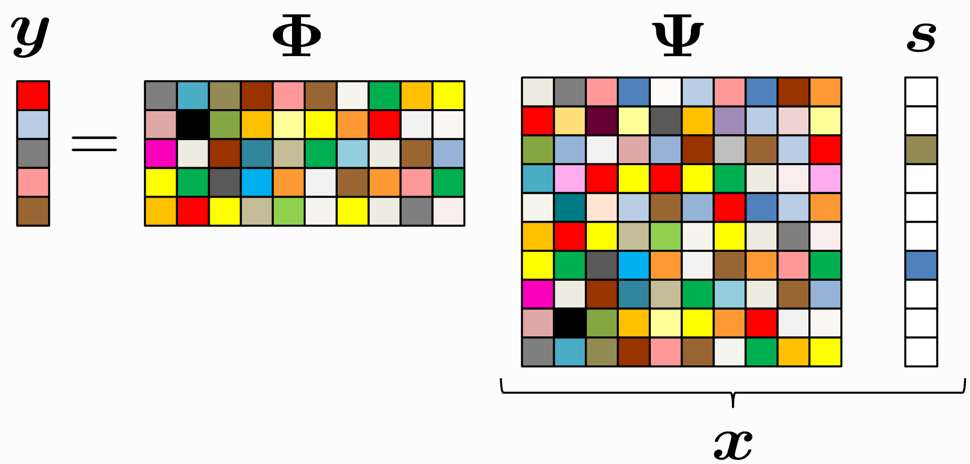
Why sparse sensing?

- **Economical** constraints (hardware cost)
- Limited **physical space**
- Limited data **storage space**
- Reduce **communications bandwidth**
- Reduce **processing overhead**

Compressive sensing

- **State-of-the-art tool** for sensing cost reduction

[Donoho 2006], [Candès 2006]



- Random linear projections of Nyquist rate samples
- Sparse signal reconstruction

$$\min_{\boldsymbol{s}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{s}\|_2^2 \quad \text{s.to} \quad \|\boldsymbol{s}\|_0 \leq \epsilon$$

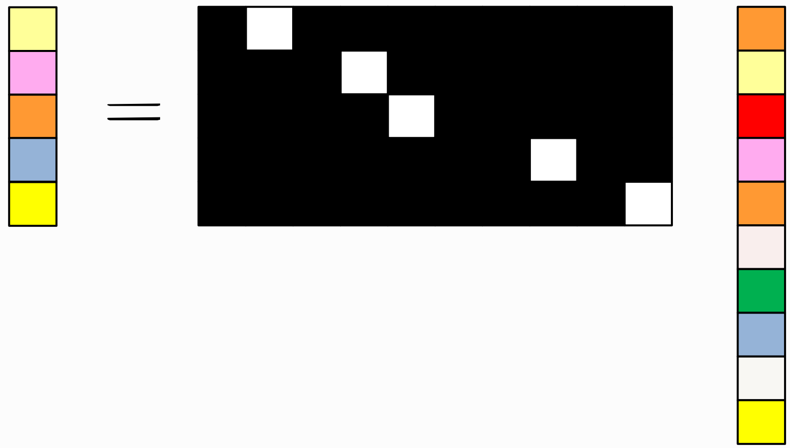
- D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, feb. 2006.

Sparse sensing vs. compressed sensing

	Compressed sensing	Sparse sensing
Sparse signal	**needed**	**not needed**
Samplers	**random**	**deterministic, sparse**
Compression	**robust**	**practical, controllable**
Signal processing task	**sparse signal reconstruction**	**any statistical inference**

Discrete sparse sensing

$$\boldsymbol{y} \qquad \boldsymbol{\Phi}(\boldsymbol{w}) \in \{0,1\}^{K\times M} \quad \boldsymbol{x}$$



- Candidate measurement set (samples, sensors, etc.):

$$\boldsymbol{x} = [x_1, x_2, \ldots, x_M]^T$$

- Sparse sensing vector (Boolean):

$$\boldsymbol{w} = [w_1, w_2, \ldots, w_M]^T \in \{0,1\}^M$$

Design problem

Select the “best” subset of sensors out of the candidate sensors that guarantee a certain desired inference performance.

Formulation 1

$$\begin{aligned} \arg\min_{\boldsymbol{w}} \quad & \|\boldsymbol{w}\|_0 \\ \text{s.to} \quad & f(\boldsymbol{w}) \leq \lambda \\ & \boldsymbol{w} \in \{0,1\}^M \end{aligned}$$

Formulation 2

$$\begin{aligned} \arg\min_{\boldsymbol{w}} \quad & f(\boldsymbol{w}) \\ \text{s.to} \quad & \|\boldsymbol{w}\|_0 = K \\ & \boldsymbol{w} \in \{0,1\}^M \end{aligned}$$

$f(\boldsymbol{w})$ inference performance metric

λ prescribed accuracy

K sample size

Nonconvex Boolean problem

Solutions to the combinatorial problem

Exact solutions:

$M = 30, 2^M = 10^9$

- Exhaustive search over
  - 2^M possible candidates for *formulation 1*
  - $\binom{M}{K}$ possible candidates for *formulation 2*

- Branch-and-bound methods

  [Lawler-Wood-1966], [Nguyen-Miller-1992]

  - long runtimes even for a modest sized problem

- E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Oper. Res.*, vol. 14, pp. 699–719, 1966.
- N. Nguyen and A. Miller, "A review of some exchange algorithms for constructing discrete D-optimal designs," *Comput. Statist. Data Anal.*, vol. 14, pp. 489–498, 1992

Solutions to the combinatorial problem

Suboptimal solutions:

- **Convex** optimization (polynomial time)

  [Joshi-Boyd-2009], [Chepuri-Leus-2015]

  - convex relaxation for $\|\boldsymbol{w}\|_0, \ \{0,1\}, f(\boldsymbol{w})$
  - thresholding, randomization to get back a Boolean solution
  - Semidefinite program (SDP) typically

- S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009
- S.P. Chepuri and G. Leus. “Sparsity-Promoting Sensor Selection for Non-linear Measurement Models,” *IEEE Trans. on Signal Processing*, vol. 63, no. 3, pp. 684-698, Feb. 2015.

Solutions to the combinatorial problem

Suboptimal solutions:

- **Submodular** optimization (linear search)

  [Krause-Singh-Guestrin-2008], [Ranieri-Chebira-Vetteri-2014]

  - greedy search
  - solution is near optimal

- A. Krause, A. Singh, and C. Guestrin, “Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies,” *J. Machine Learn. Res.*, vol. 9, pp. 235–284, Feb. 2008.
- J. Ranieri, A. Chebira, and M. Vetterli, “Near-optimal sensor placement for linear inverse problems,” *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1135–1146, Mar. 2014

Convex optimization

Requires $f(\cdot)$ to be convex function

- Boolean constraint is relaxed to the box constraint $[0,1]^M$
- ℓ_0(-quasi) norm is relaxed to either ℓ_1-norm: $\sum_{m=1}^{M} w_m$ or a form that is iteratively convex *[Candés-Wakin-Boyd-2008]*

Formulation 1

$$\arg\min_{\boldsymbol{w}} \ \mathbf{1}^T \boldsymbol{w}$$

$$\text{s.to} \quad f(\boldsymbol{w}) \leq \lambda$$

$$\boldsymbol{w} \in [0,1]^M$$

Formulation 2

$$\arg\min_{\boldsymbol{w}} \ f(\boldsymbol{w})$$

$$\text{s.to} \quad \mathbf{1}^T \boldsymbol{w} = K$$

$$\boldsymbol{w} \in [0,1]^M$$

- E. Candés, M. Wakin, and S. Boyd, “Enhancing sparsity by reweighted minimization l1-minimization,” *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.

Submodular optimization

Requires $f(\cdot)$ to be submodular monotonically increasing function

- Define the sampling set:

$$\mathcal{X} := \mathcal{S} = \{m | w_m = 1, m = 1, 2, \ldots, M\}$$

or $$\mathcal{X} := \mathcal{M} \backslash \mathcal{S} = \{m | w_m = 0, m = 1, 2, \ldots, M\}$$

- *Set* function $f(\mathcal{X})$ is submodular, if $\forall \mathcal{X} \subseteq \mathcal{Y} \subset \mathcal{M},\ s \in \mathcal{M} \backslash \mathcal{Y}$

$$f(\mathcal{X} \cup \{s\}) - f(\mathcal{X}) \geq f(\mathcal{Y} \cup \{s\}) - f(\mathcal{Y})$$

- If $f(\mathcal{X})$ is monotonically increasing, i.e., $f(\mathcal{X} \cup \{s\}) \geq f(\mathcal{X})$

Submodular optimization

Algorithm 1 Greedy algorithm

1. **Require** $\mathcal{X} = \emptyset, L$.
2. **for** $k = 1$ to L
3. $\quad s^* = \arg\max_{s \notin \mathcal{X}} f(\mathcal{X} \cup \{s\})$
4. $\quad \mathcal{X} \leftarrow \mathcal{X} \cup \{s^*\}$
5. **end**
6. **Return** $\mathcal{X}$

Linear search

$$L = K \text{ or } L = M - K$$

Then, greedy algorithm is near-optimal

$$f(\mathcal{X}) \geq \underbrace{(1 - 1/e)}_{63\%} \max_{|\mathcal{Y}|=L} f(\mathcal{Y})$$

[Nemhauser-Wolsey-Fisher-1978]

- G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions— I," Mathematical Programming, vol. 14, no. 1, pp. 265–294, 1978.

Sparse sensing for estimation

- Unknown parameter vector $\boldsymbol{\theta} \in \mathbb{C}^N$ follows

$$y_m = w_m \overbrace{h_m(\boldsymbol{\theta}, n_m)}^{x_m \sim p_m(x;\boldsymbol{\theta})}, \; m = 1, 2, \ldots, M$$

- Linear observations with indep. additive Gaussian noise

[Joshi-Boyd-09]

- What about more general cases?
- Exact MSE is hard to optimize and depends on algorithm
- Use the **Cramér-Rao bound** as the performance metric

$$\mathbb{E}\{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\} \geq \boldsymbol{C} = \boldsymbol{F}^{-1}$$

Fisher Information matrix (FIM)

- S. Joshi and S. Boyd, "Sensor Selection via Convex Optimization," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 451–462, Feb. 2009

Statistically independent observations

- Independent observations: FIM is additive

$$\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta}) = \sum_{m=1}^{M} w_m \boldsymbol{F}_m(\boldsymbol{\theta})$$

- Dependent Gaussian observations: $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{h}(\boldsymbol{\theta}), \boldsymbol{\Sigma}\right)$

$$\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta}) = \left[\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{J}(\boldsymbol{\theta})\right]^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{w}) \left[\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{J}(\boldsymbol{\theta})\right]$$

$$\boldsymbol{J}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \qquad \boldsymbol{\Sigma}^{-1}(\boldsymbol{w}) = \Big(\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{\Sigma}\boldsymbol{\Phi}^T(\boldsymbol{w})\Big)^{-1}$$

Use $\boldsymbol{\Sigma} = a\boldsymbol{I} + \boldsymbol{S}$ with $a > 0$ and $\boldsymbol{S} \succ \boldsymbol{0}$ to obtain proxy

$$\begin{bmatrix} \boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) \\ \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1} & \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) \end{bmatrix}$$

This is again additive

$f(\boldsymbol{w})$ for estimation – scalar measures

- Prominent scalar measures
  - **E-optimality** measure (worst case error)

$$f(\boldsymbol{w}) := \lambda_{\mathsf{max}}\{\boldsymbol{F}^{-1}(\boldsymbol{w}, \boldsymbol{\theta})\} \longrightarrow \text{Convex (SDP)}$$

  - **A-optimality** measure (average error)

$$f(\boldsymbol{w}) := \mathsf{tr}\{\boldsymbol{F}^{-1}(\boldsymbol{w}, \boldsymbol{\theta})\} \longrightarrow \text{Convex (SDP)}$$

  - **D-optimality** measure (error volume)

$$f(\boldsymbol{w}) := \ln\det\{\boldsymbol{F}^{-1}(\boldsymbol{w}, \boldsymbol{\theta})\}$$

Convex

Submodular

  - Frame potential (approx. error) $\longrightarrow$ Submodular

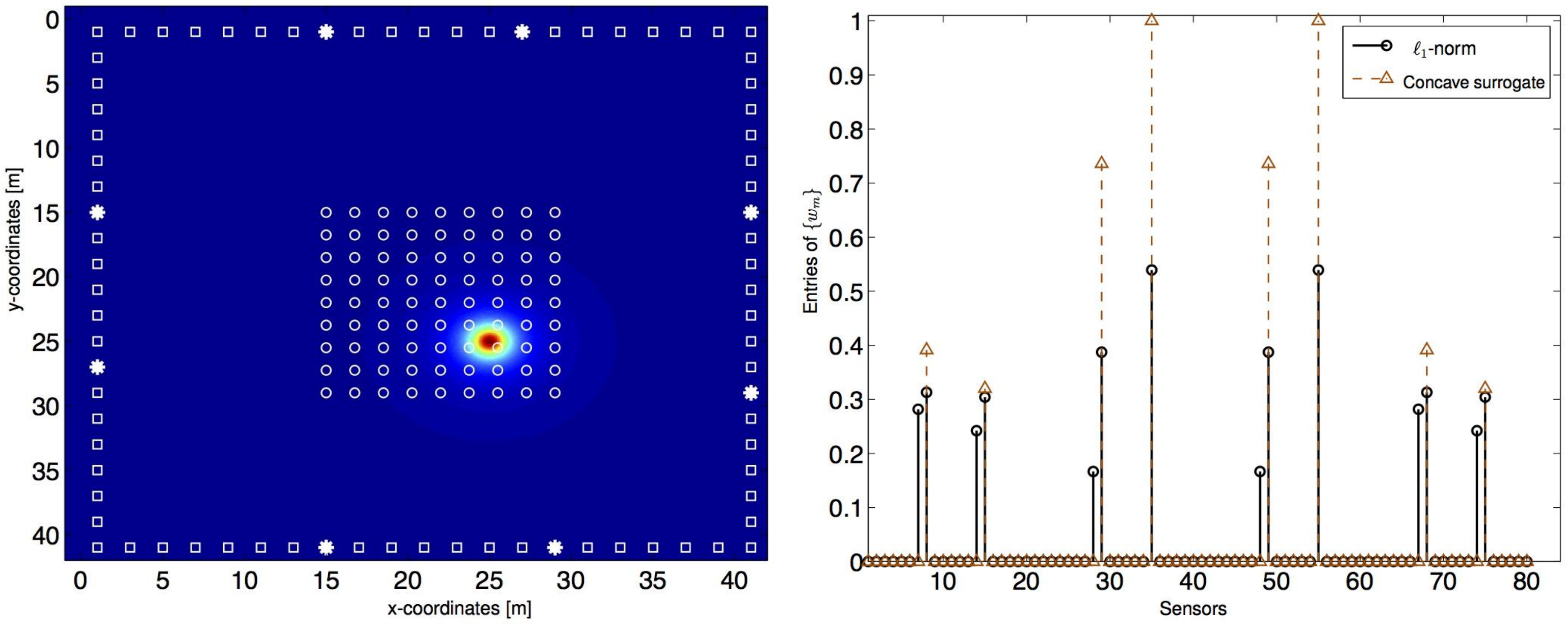
Sparse sensing for estimation

Setting	Convex	Submodular
Optimization criterion	Trace of inverse FIM = MSE Mimimum eigenvalue FIM	Logdet of FIM Frame potential
Independent Gaussian observations, linear	**SDP** using LMI	**Greedy** method
Independent observations, nonlinear	**SDP** - One LMI per possible solution - Single LMI for Bayesian cost	**Greedy** on Bayesian cost
Dependent Gaussian observations, linear	**SDP** using extended LMI	**Greedy** on extended Logdet of FIM
Dependent Gaussian observations, nonlinear	**SDP** - One ext. LMI per possible solution - Single ext. LMI for Bayesian cost	**Greedy** on Bayesian extended Logdet of FIM

- S.P. Chepuri and G. Leus, “Sparsity-Promoting Sensor Selection for Non-linear Measurement Models,” *IEEE Trans. on Signal Processing*, vol. 63, no. 3, pp. 684-698, Feb. 2015.
- S. Liu, S.P. Chepuri, M. Fardad, E. Masazade, G. Leus, and P.K. Varshney, “Sensor Selection for Estimation with Correlated Measurement Noise,” *IEEE Transactions on Signal Processing*, Mar. 2016.
- S. Rao, S.P. Chepuri, and G. Leus, “Greedy Sensor Selection for Non-Linear Models,” In *Proc. to the IEEE Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc.* (CAMSAP 2015), Cancun, Mexico, December 2015.

Example: RSS target localization

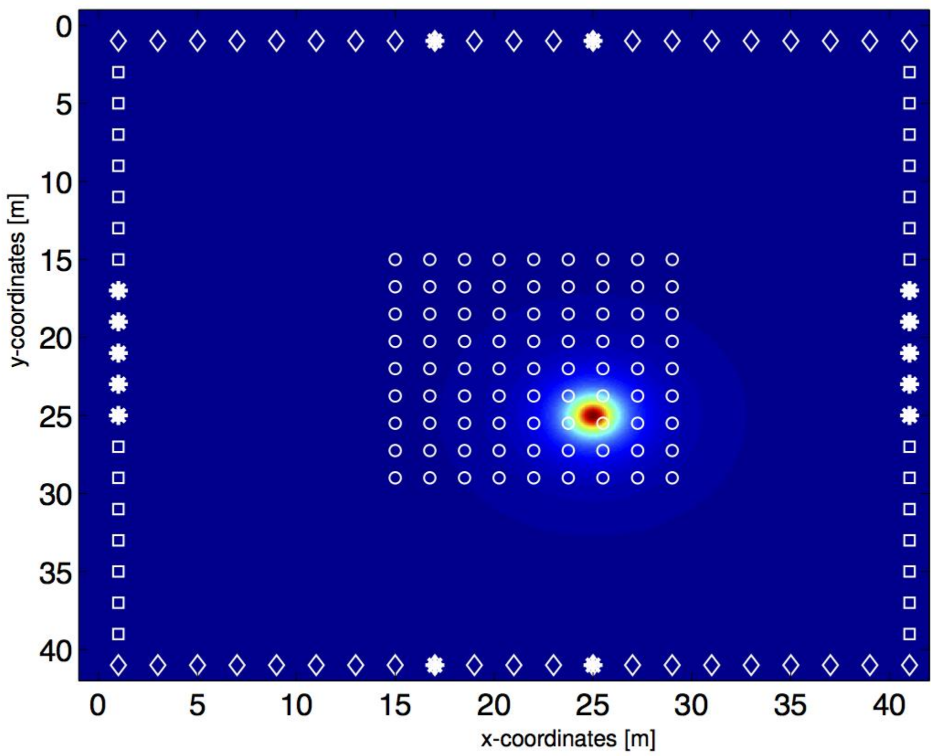
- Sensors along all edges are not correlated

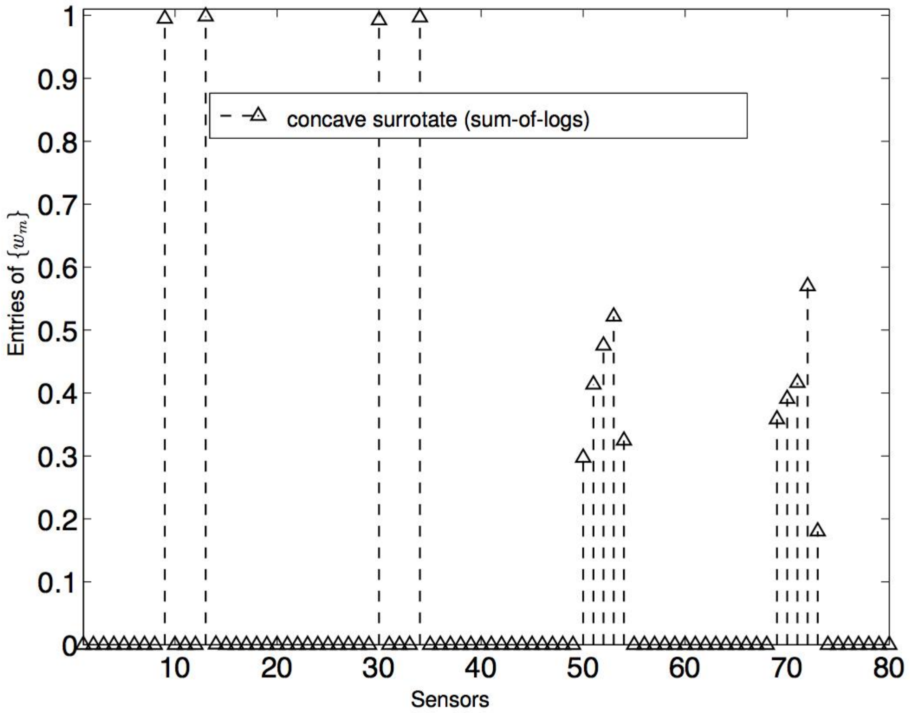


Out of 80 available access point locations, 8 access points are selected.

Example: RSS target localization

- Sensors along horizontal edges are equicorrelated (correlation coefficient 0.5)
- Sensors along vertical edges are not correlated





Out of 80 available uncorrelated and correlated access point locations, 14 access points are selected.

Sparse sensing for detection

- Observations follow (binary hypothesis testing)

$$\mathcal{H}_0 : \quad y_m = w_m x_m; x_m \sim p_m(x|\mathcal{H}_0),\ m = 1, 2, \ldots, M$$
$$\mathcal{H}_1 : \quad y_m = w_m x_m; x_m \sim p_m(x|\mathcal{H}_1),\ m = 1, 2, \ldots, M$$

- Independent Gaussian observations

[Cambanis-Masry-83], [Yu-Varshney-97], [Bajovic-Sinopoli-Xavier-11]

- What about more general cases?

- S. Cambanis and E. Masry, “Sampling designs for the detection of signals in noise,” *IEEE Trans. Inf. Theory*, vol. 29, no. 1, pp. 83–104, Jan. 1983.
- C.-T. Yu and P. K. Varshney, “Sampling design for Gaussian detection problems,” *IEEE Trans. Signal Process.*, vol. 45, no. 9, pp. 2328–2337, 1997.
- D. Bajovic, B. Sinopoli, and J. Xavier, “Sensor selection for event detection in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4938–4953, Oct. 2011.

Sparse sensing for detection

Neyman Pearson setting

$$\arg \min_{\boldsymbol{w} \in \{0,1\}^M} \|\boldsymbol{w}\|_0$$
$$\text{s.to} \quad P_f(\boldsymbol{w}) \leq \alpha,$$
$$P_m(\boldsymbol{w}) \leq \beta$$

$$P_m = 1 - P(\widehat{\mathcal{H}} = \mathcal{H}_1 | \mathcal{H}_1)$$
$$P_f = P(\widehat{\mathcal{H}} = \mathcal{H}_1 | \mathcal{H}_0)$$

Bayesian setting

$$\arg \min_{\boldsymbol{w} \in \{0,1\}^M} \|\boldsymbol{w}\|_0$$
$$\text{s.to} \quad P_e(\boldsymbol{w}) \leq e$$

π_0, π_1 prior probabilities

$$P_e = \pi_0 P_f + \pi_1 P_m$$

- Exact error probabilities hard to optimize
- Seek weaker performance measures
  - Kullback-Leibler distance
  - J-divergence
  - Bhattacharyya distance

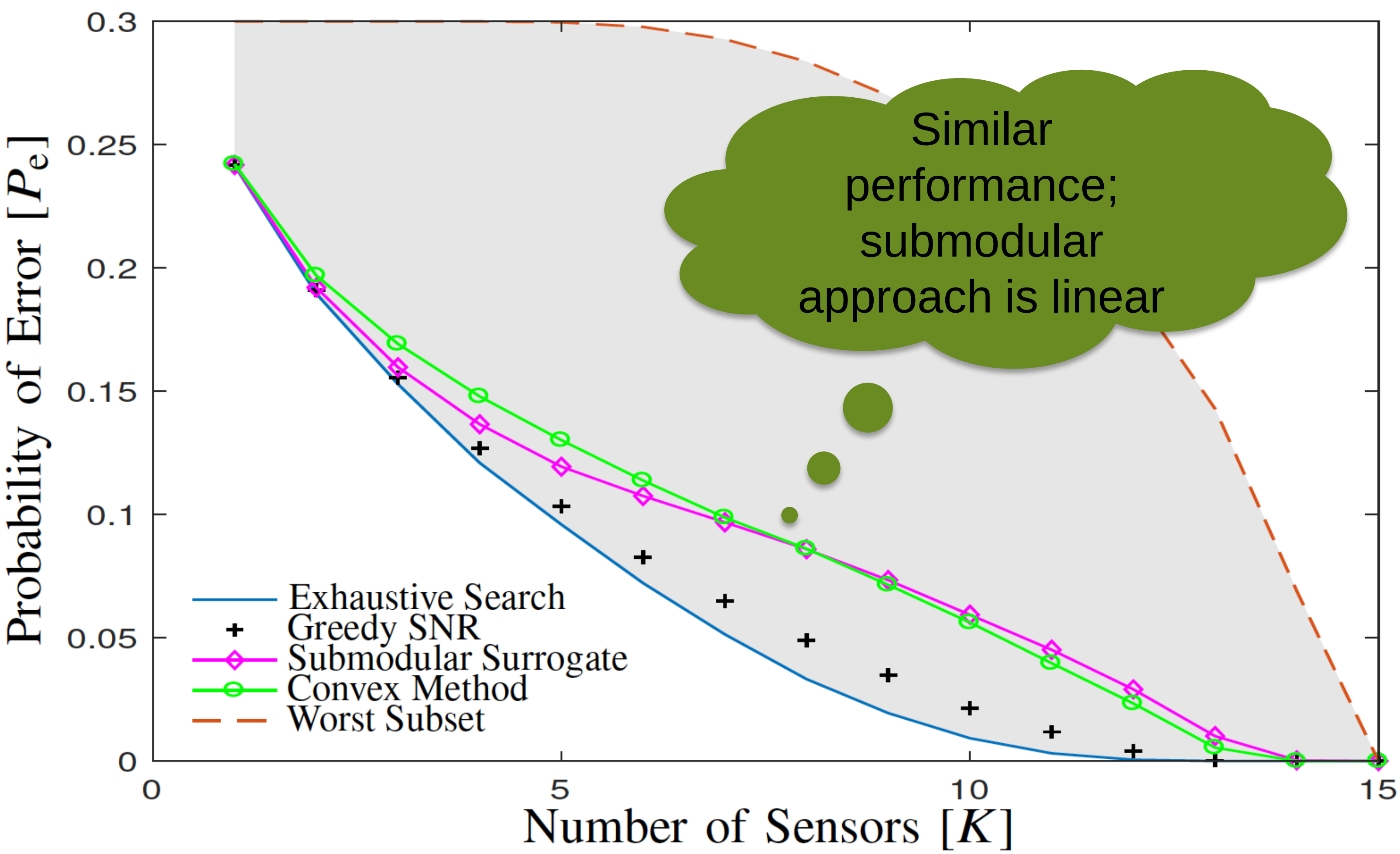
Sparse sensing for detection

Setting	Neyman-Pearson	Bayesian
Optimization criterion	Kullback-Leibler distance or J-divergence	Bhattacharyya distance or J-divergence
Independent observations	**Ordering** distances	**Ordering** distances
Dependent Gaussian observations (uncommon means)	**Convex:** SNR matrix **Greedy:** Logdet SNR matrix	**Convex:** SNR matrix **Greedy:** Logdet SNR matrix
Dependent Gaussian observations (uncommon variances)	**Sup-sub**: Kullback-Leibler dist. **Convex:** J-divergence	**Sup-sub**: Bhattacharyya dist. **Convex**: J-divergence

- M. Coutino, S.P. Chepuri, and G. Leus, "Near-Optimal Sparse Sensing for Gaussian Detection with Correlated Observations," *IEEE Transactions on Signal Processing*, vol 66, no. 15, pp. 4025-4039, Aug. 2018.
- S.P. Chepuri and G. Leus, "Sparse Sensing for Distributed Detection," *IEEE Trans. on Signal Processing*, vol. 16, no. 6, pp. 1446-1460, Mar. 2016.
- S.P. Chepuri and G. Leus, "Sparse Sensing for Distributed Gaussian Detection," In *Proc. of the International Conference on Acoustics, Speech, and Signal Processing* (ICASSP 2015), Brisbane, Australia, April 2015. (ICASSP best student paper award)

Example: Convex or submodular?

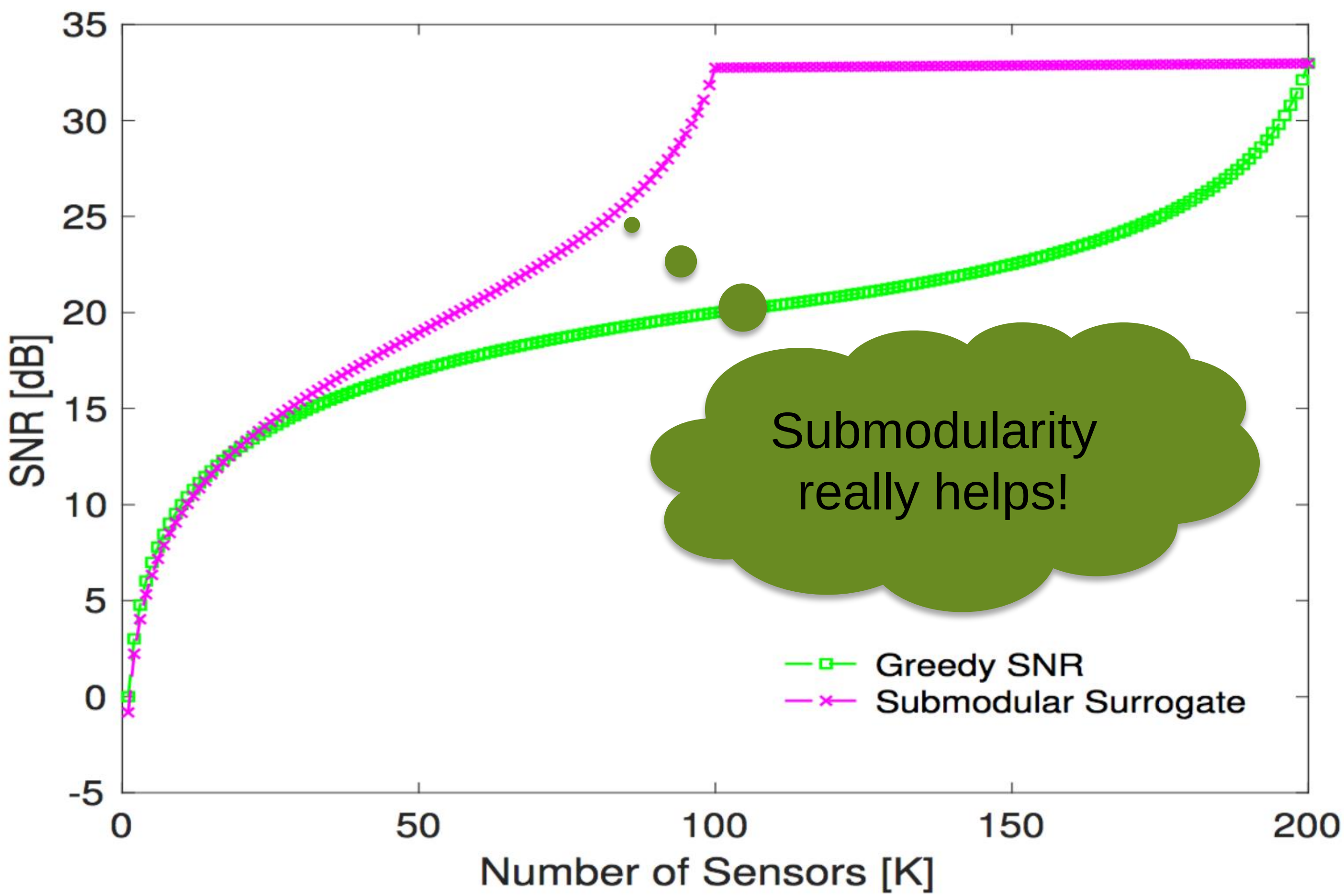
Random Toeplitz correlation matrices based on array processing



Example: Can “simple” greedy fail?

Subset of calibrated sensors

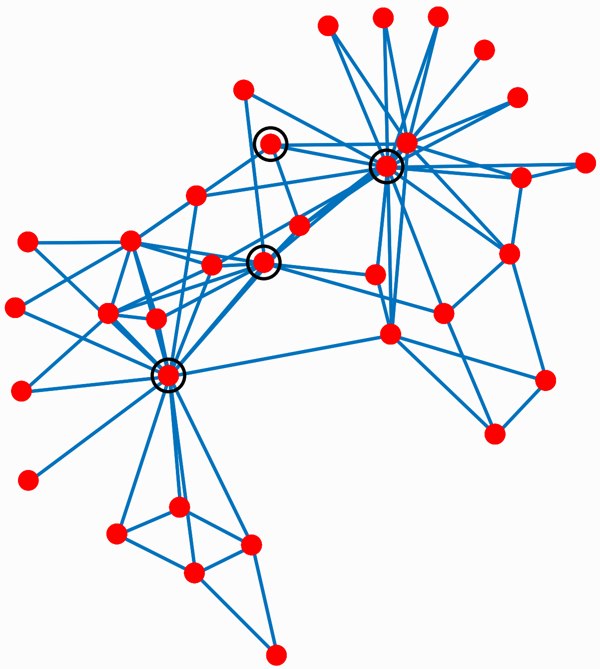
$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{T} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \qquad \boldsymbol{T} = \mathrm{Toeplitz}([1, \rho^1, \rho^2, \ldots, \rho^{\mathrm{M}/2-1}])$$



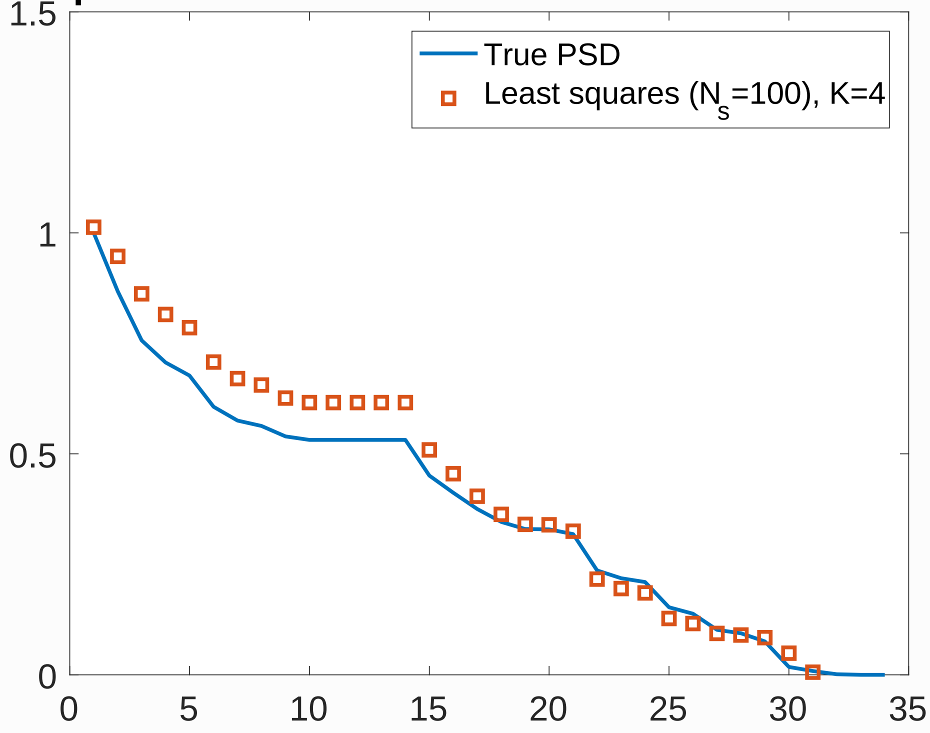
Recent developments

Subsampling signals on graphs

Sample 4 out of 34 nodes



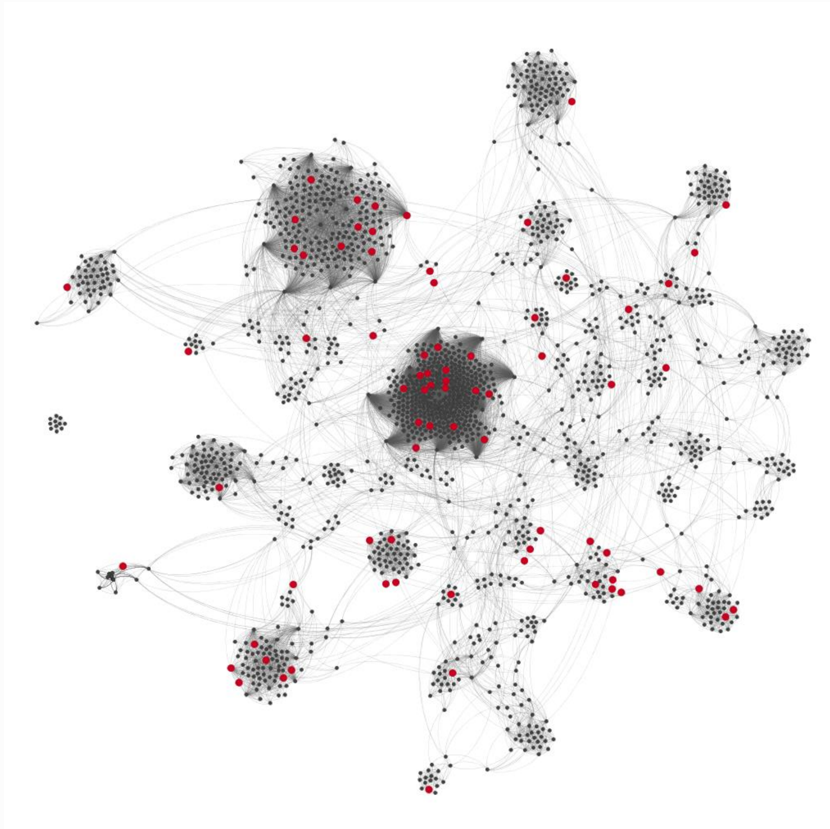
MA parametric PSD estimation from subset of nodes

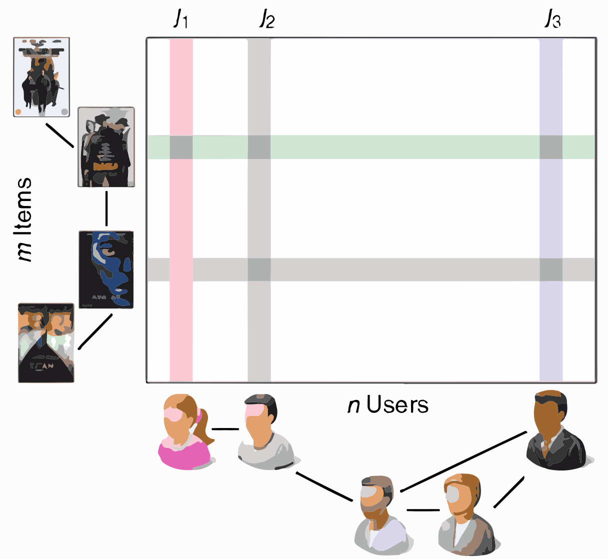


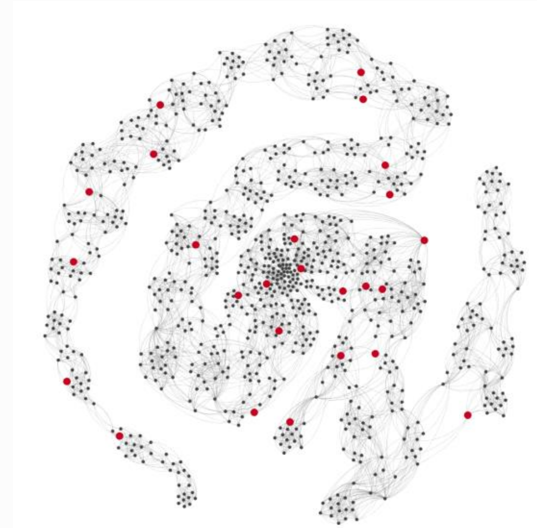
- S.P. Chepuri and G. Leus, "Graph Sampling for Covariance Estimation," *IEEE Jour. on Sel. Topics in Sig. Proc. and IEEE Trans. on Sig. and Info. Proc. over Networks, joint special issue on Graph Signal Processing*, vol. 3, no. 3, pp. 451-466, Sep. 2017.

Recent developments

Structured selection for tensors







Similar performance as state-of-the-art methods
but 1.875 measurements vs. 80.000 measurements

- G.Ortiz-Jimenez, M. Coutino, S.P. Chepuri, and G. Leus, “Sparse Sampling for Inverse Problems with Tensors,” *IEEE Transactions on Signal Processing*, submitted, Jun. 2018.

Conclusions

	Sparse sensing
Sparse signal	**not needed**
Samplers	**deterministic and sparse**
Compression	**practical, controllable**
Signal processing task	**any statistical inference**

- Design space-time sparse samplers

  Extend Nyquist-based classical sensing techniques

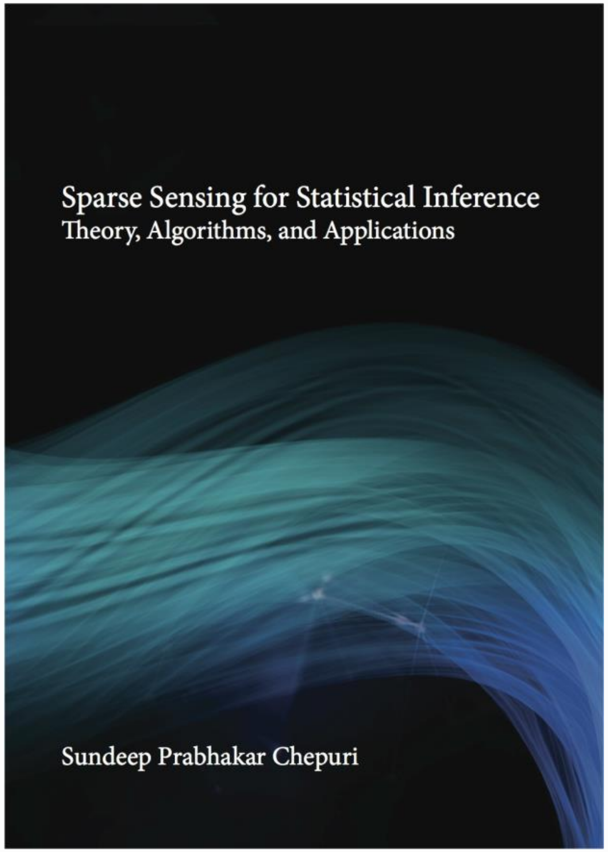
- Basic statistical inference problems

  Estimation, filtering, and detection

Reference material

PhD thesis

http://theses.eurasip.org/theses/648/sparse-sensing-for-statistical-inference-theory/



Monograph in Foundations and Trends in Signal Processing

Thank You!
Questions?