EUSIPCO 2016

Sparse Sensing for Statistical Inference

Geert Leus **Sundeep Chepuri**

TU Delft
Delft University of Technology

29 Aug 2016

Thanks!

Collaborators at TU Delft

- Alle-Jan van der Veen
- Andrea Simonetto
- Georg Kail
- Hadi Jamali Rad
- Inna Ivashko
- Rocio Arroyo-Valles
- Shilpa Rao
- Venkat Roy

Collaborators outside TU Delft

- Georgios Giannakis (Uni. of Minnesota)
- Pramod Varshney (Uni. of Syracuse)
- Sijia Liu (Uni. of Michigan)
- Xiaoli Ma (Georgia Tech.)
- Yu Zhang (Uni. of California, Berkley)

STW connecting innovators

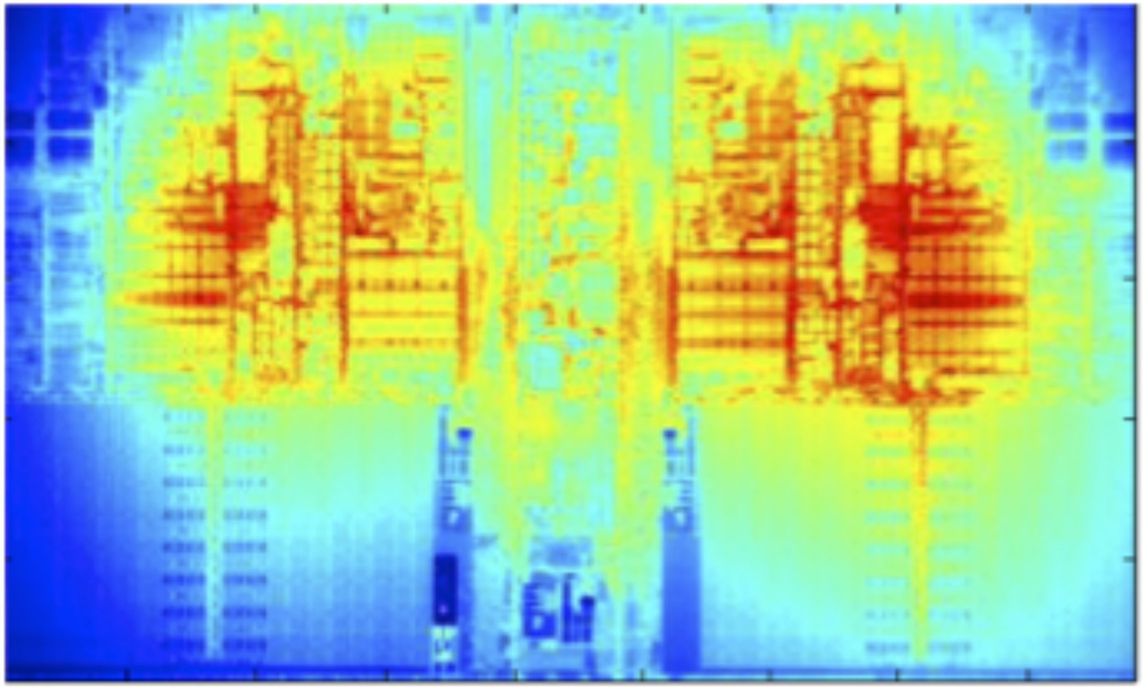
NWO Netherlands Organisation for Scientific Research

Roadmap

Introduction

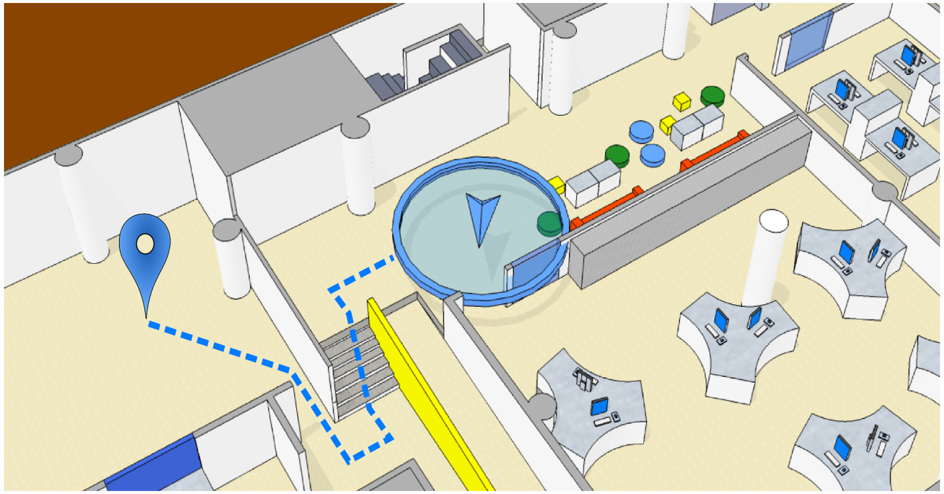
- ❑ Overview
- ❑ Sparse sensing
- ❑ Inverse problems
- ❑ Filtering
- ❑ Detection
- ❑ Off-the grid sparse sensing
- ❑ Related problems
- ❑ Conclusions, future directions, Q&A

How to optimally deploy sensors?

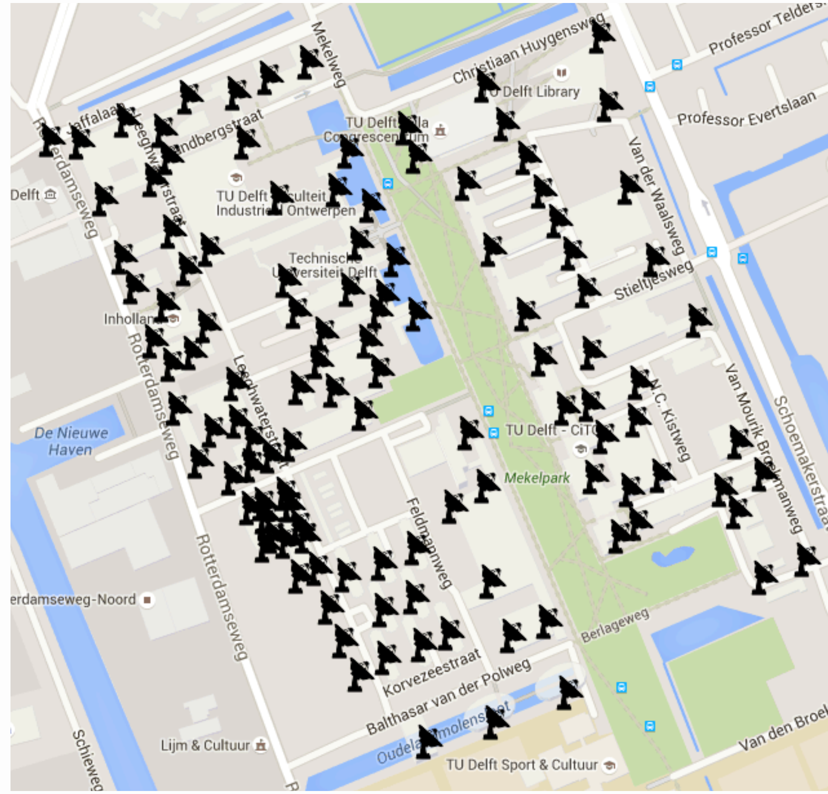


Thermal map of a processor

- Field estimation/filtering: localize (varying) heat source(s)
- Field detection: detect hot spot(s)

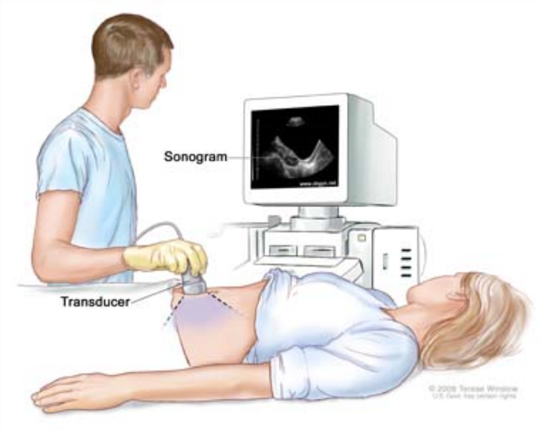


Indoor localization (e.g., museum)



Distributed radar (TU Delft campus)

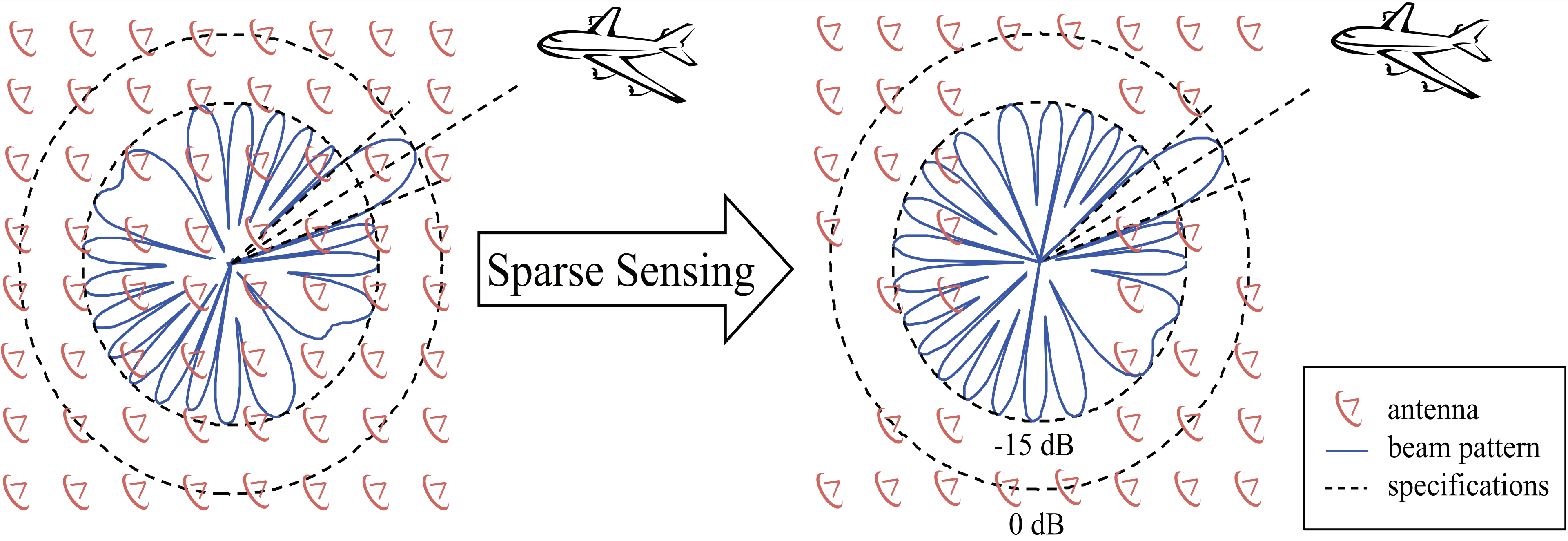
Microseismic event detection



Ultrasound imaging

Radio astronomy (e.g., LOFAR, SKA)

Power networks (PMU placement)



Design **structured** (sparse) space-time samplers

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]*

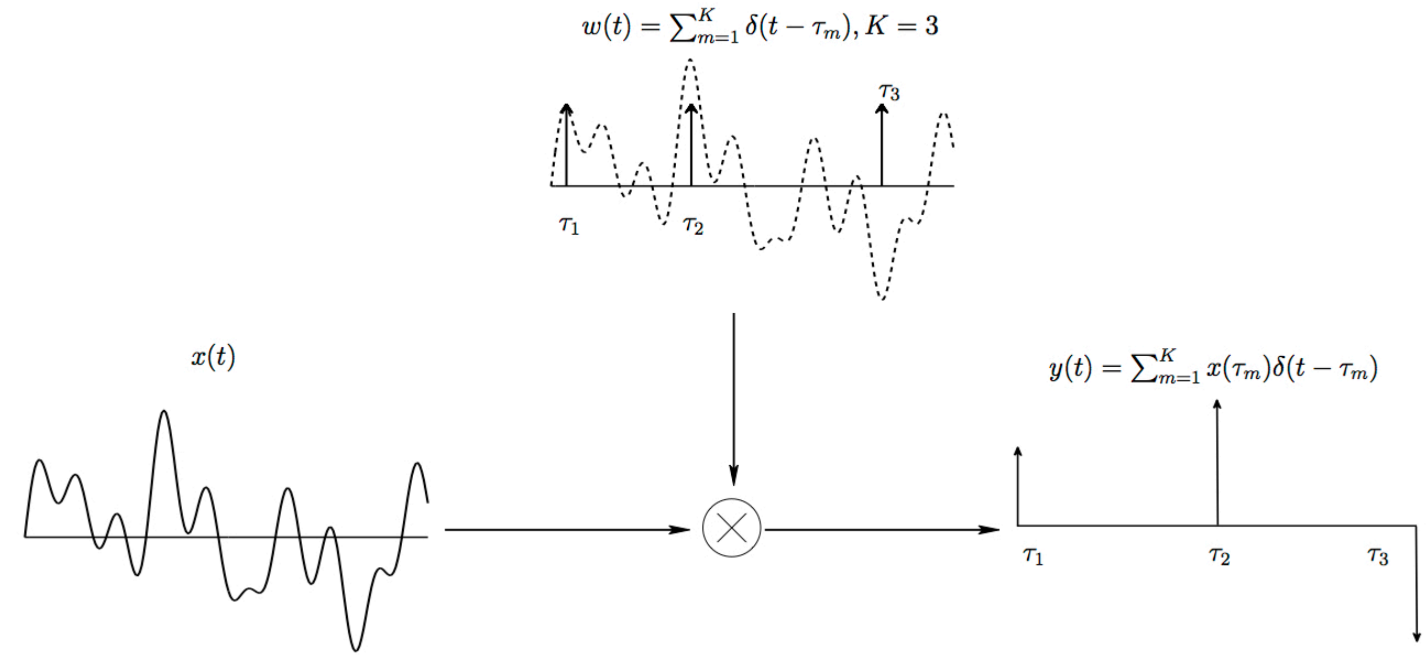
- T. Blu, P.L. Dragotti, M.Vetterli, P. Marziliano, and L. Coulot. "Sparse sampling of signal innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31-40, Mar. 2008.
- P.P. Vaidyanathan and P. Pal. "Sparse sensing with co-prime samplers and arrays." *IEEE Transactions on Signal Processing*,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

What is sparse sensing?

Find the best indices $\{\tau_m\}$ to sample $x(t)$ such that a desired inference performance is achieved.



Design a **sparse sampler** $w(t) = \sum_m \delta(\tau - \tau_m)$ to acquire

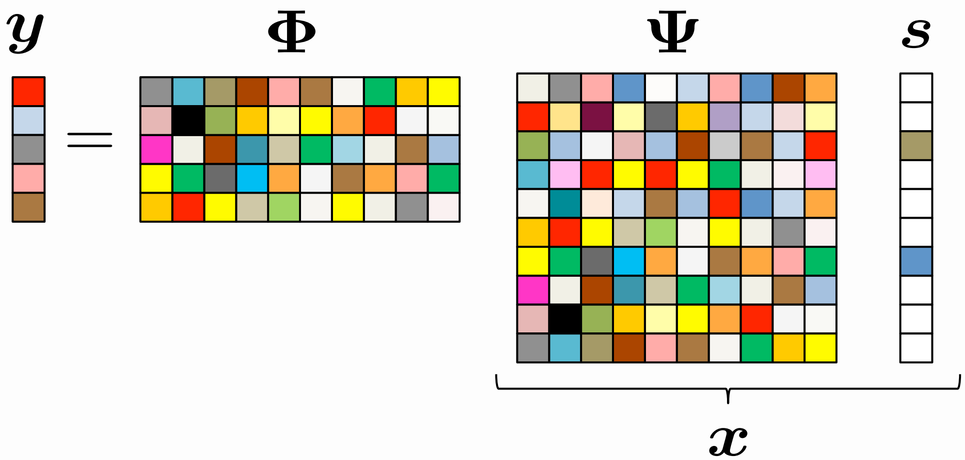
$$y(t) = w(t)x(t) = \sum_m x(\tau_m)\delta(\tau - \tau_m)$$

Inference tasks can estimation, filtering, and detection

Compressive sensing

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

[Donoho 2006], [Candès 2006]

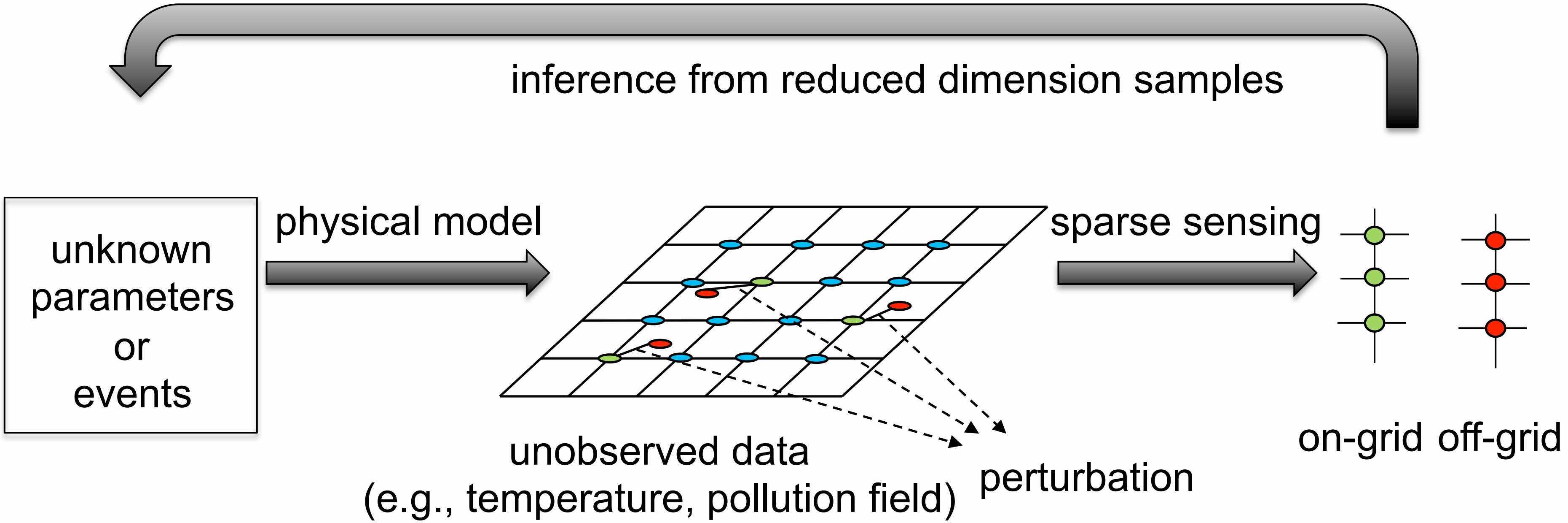


- Random linear projections of Nyquist rate samples
- Sparse signal reconstruction

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

- 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.

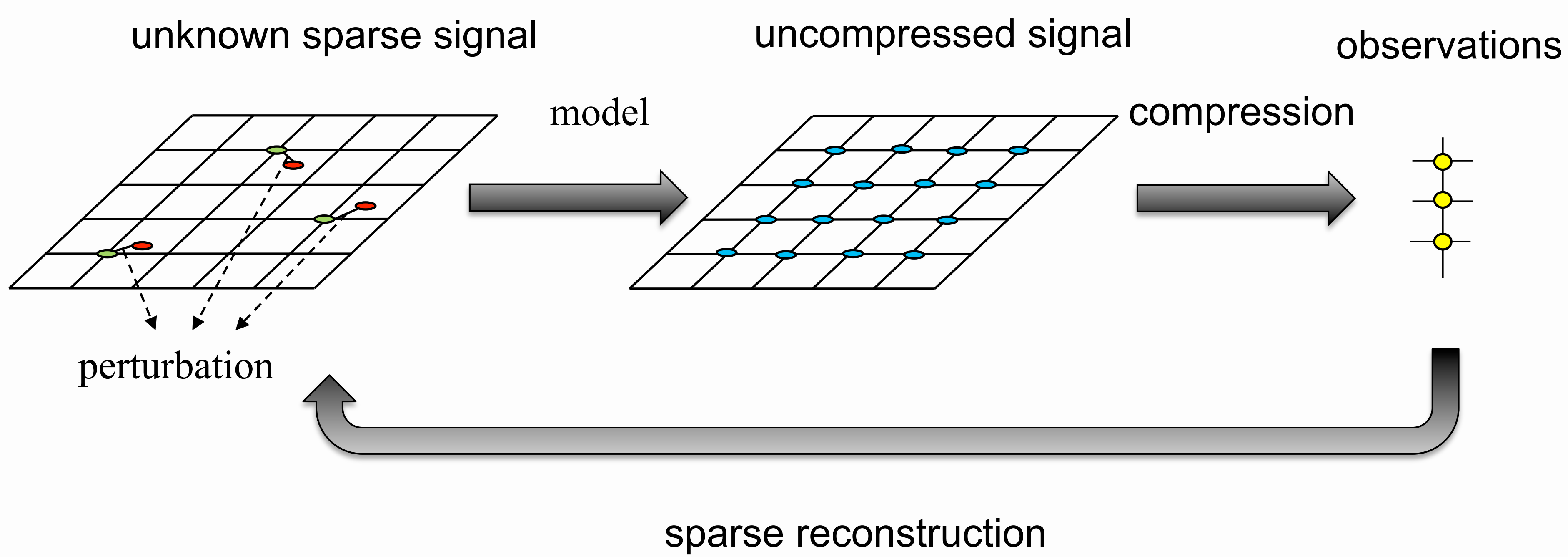
Sparse sensing vs. compressed sensing



Sparse sensing

Sparse sensing vs. compressed sensing

Compressed sensing

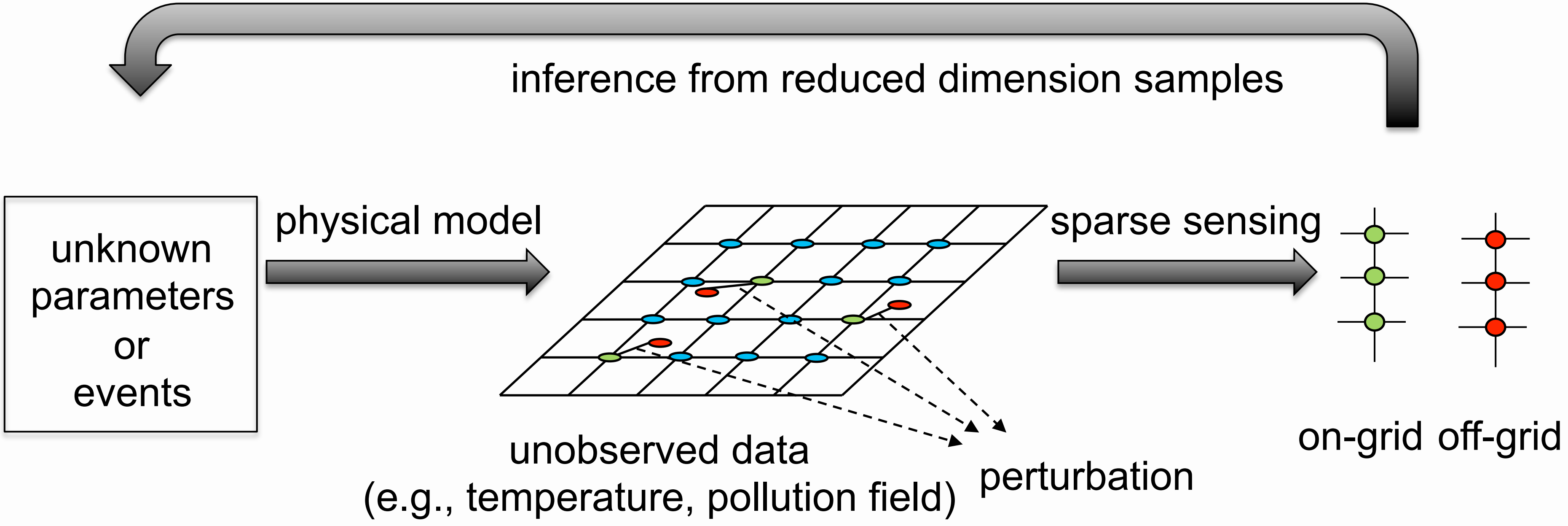


Sparse sensing vs. compressed sensing

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

Sparse sensing paradigms

1. Discrete (or on-grid) sparse sensing
2. Continuous (or off-grid) sparse sensing



Discrete Sparse Sensing

Discrete sparse sensing - model

- To design continuous-domain $w(t)$

$M \gg K$

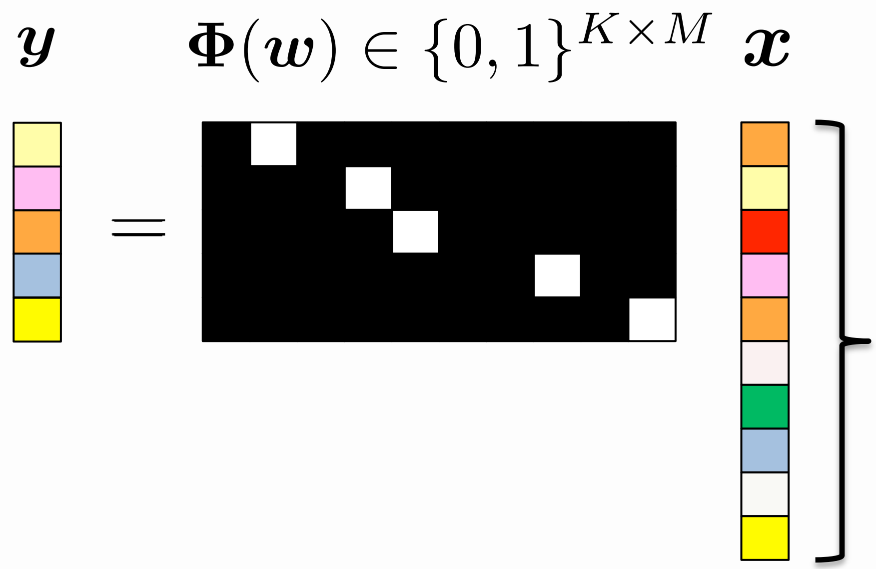
- **Assume** $\{\tau_1, \tau_2, \cdots, \tau_K\}$ lie on a discrete grid $\{t_1, t_2, \cdots, t_M\}$

- Design the discrete sensing vector

$$\boldsymbol{w} = [w(t_1), w(t_2), \ldots, w(t_M)]^T$$
$$= [w_1, w_2, \ldots, w_M]^T \in \{0, 1\}^M$$

M number of candidate sensors
$w_m = (0)1$ sensor is (not) selected

Discrete sparse sensing - model



- Sensor selection
- Sensor placement
- Sample selection
- Antenna selection

➢ Input is also discretized $\boldsymbol{x} = [x(t_1), x(t_2), \ldots, x(t_M)]^T$

➢ Sparse sensing structure

- only one nonzero entry per row
- many zero columns

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

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

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

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 $\{0,1\}, f(\boldsymbol{w})$
  - thresholding, randomization to get back a Boolean solution
  - Semidefinite program (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 time)

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

  - ❑ greedy search
  - ❑ solution is $1 - 1/e$ 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 of its argument

- Boolean constraint is relaxed to the box constraint $[0,1]^M$

- ℓ_0(-quasi) norm is relaxed to either:
  a) ℓ_1-norm: $\sum_{m=1}^{M} w_m$

Convex optimization

Requires $f(\cdot)$ to be convex function of its argument

- Boolean constraint is relaxed to the box constraint $[0,1]^M$

- ℓ_0(-quasi) norm is relaxed to either:
  a) ℓ_1-norm: $\sum_{m=1}^{M} w_m$
  b) sum-of-logarithms: $\sum_{m=1}^{M} \ln(w_m + \delta)$

$$\ln(w_m + \delta) \leq \ln(w_m[i-1] + \delta) + \frac{w_m - w_m[i-1]}{w_m[i-1] + \delta}$$

[Candés-Wakin-Boyd-2008]

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

Convex optimization

Requires $f(\cdot)$ to be convex function of its argument

- Boolean constraint is relaxed to the box constraint $[0, 1]^M$

- ℓ_0(-quasi) norm is relaxed to either:
  a) ℓ_1-norm: $\sum_{m=1}^{M} w_m$
  b) sum-of-logarithms: $\sum_{m=1}^{M} \ln(w_m + \delta)$
  c) Your favorite approximation

Convex optimization

Requires $f(\cdot)$ to be convex function of its argument

- Boolean constraint is relaxed to the box constraint $[0,1]^M$

- ℓ_0(-quasi) norm is relaxed to either:
  a) ℓ_1-norm: $\sum_{m=1}^{M} w_m$

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

$$\arg\min_{\boldsymbol{w}} \ f(\boldsymbol{w})$$
$$\text{s.to} \quad \mathbf{1}^T \boldsymbol{w} = K$$
$$\boldsymbol{w} \in [0,1]^M$$

The question!

What is the convex function $f(w)$ that we can **minimize** for

- ✓ Estimation
- ✓ Filtering
- ✓ Detection

Submodular optimization

Requires $f(\cdot)$ to be **submodular** function of its argument

- 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

If $f(\cdot)$ is **submodular** and **monotonic**

Linear time

Algorithm 1 Greedy algorithm

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

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

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.

The question!

What is the submodular function $f(\mathcal{X})$ that we can **maximize** for

- ✓ Estimation
- ✓ Filtering
- ✓ Detection

Model-driven vs. data-driven

Model-driven

- Performance metric (**ensemble**)
  - ✓ Based on **model information** only
  - ✓ Design doesn't depend on the actual data
  - ✓ **Offline** sensing design **to acquire** data

Data-driven

- Performance metric (**Instantaneous**)
  - ✓ Based on **model** + **data** information
  - ✓ Design depends on actual data
  - ✓ **Sketch** or **censor** already acquired data (**big data**)

[Rago-Willett-Shalom-96], [Msechu-Giannakis-12]

- Rago, C., Willett, P. and Bar-Shalom, Y., 1996. Censoring sensors: A low-communication-rate scheme for distributed detection. IEEE Transactions on Aerospace and Electronic Systems, Vol. 32, no. 2, pp.554-568.
- Msechu, E.J. and Giannakis, G.B., 2012. Sensor-centric data reduction for estimation with WSNs via censoring and quantization. *IEEE Transactions on Signal Processing*, Vol. 60, no. 1, pp.400-414.

Estimation

1. 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.

2. S.P. Chepuri and G. Leus. **Sparse Sensing for Estimation with Correlated Observations**. To appear in Asilomar Conf. Signals, systems, and Computers (Asilomar 2015), Pacific Grove, California, USA, November 2015.

3. 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.

4. 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.

Inverse problem

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

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

Candidate sensing locations

x_m m-th spatial or temporal sensor measurement;

h_m (in general) non-linear function;

n_m (additive/multiplicative) noise process.

$f(w)$ for estimation

Best **sampling** locations yield the lowest estimation error

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

Example: *Linear additive Gaussian model*

$$h_m(\boldsymbol{\theta}, n_m) := \mathbf{h}_m^T \boldsymbol{\theta} + n_m \quad \text{and} \quad x_m \sim \mathcal{N}(\mathbf{h}_m^T \boldsymbol{\theta}, \sigma_m^2)$$

Error of the least-squares estimate

$$\boldsymbol{C} = \left(\sum_{m=1}^{M} w_m \sigma_m^{-2} \mathbf{h}_m \mathbf{h}_m^T \right)^{-1}$$

$f(w)$ for estimation – Cramér-Rao bound

- Closed-form expression for C is not always available
  – e.g., non-linear, non-Gaussian measurement models

- 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

✓ well-suited for offline design problems

✓ reveals (local) identifiability

✓ improves performance of any practical algorithm

✓ equal to the error covariance for the linear additive Gaussian case

Statistically independent observations

- Suppose the observations are independent

$$\ln p(\boldsymbol{y};\boldsymbol{\theta}) = \ln \prod_{m=1}^{M} p(y_m;\boldsymbol{\theta})^{w_m} = \sum_{m=1}^{M} w_m \ln p(y_m;\boldsymbol{\theta})$$

- Consequence, Fisher information (FIM) is additive

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

$$\boldsymbol{F}_m(\boldsymbol{\theta}) = \mathbb{E}\left\{\left(\frac{\partial \ln p_m(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln p_m(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right\} \in \mathbb{R}^{N\times N}$$

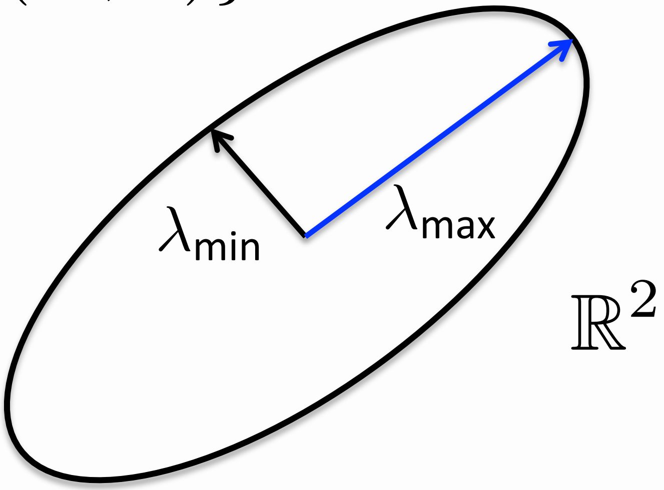
- For non-linear models and/or specific distributions, FIM **depends** on the true **unknown** parameter

$f(w)$ for estimation – scalar measures

- Prominent scalar measures

  - **E-optimality** measure (worst case error)

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



Example: *Linear additive Gaussian model*

$$f(\boldsymbol{w}) = \lambda_{\text{max}}\left\{\left(\sum_{m=1}^{M} w_m \sigma_m^{-2} \mathbf{h}_m \mathbf{h}_m^T\right)^{-1}\right\}$$

$f(w)$ for estimation – scalar measures

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

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

- SDP problem based on ℓ_1-norm heuristics

$$\begin{aligned} &\arg\min_{\boldsymbol{w}} \quad \mathbf{1}^T \boldsymbol{w} \\ &\text{s.to} \quad \sum_{m=1}^{M} w_m \boldsymbol{F}_m(\boldsymbol{\theta}) - \lambda \boldsymbol{I}_N \succeq 0, \quad \forall \boldsymbol{\theta} \in \mathcal{T}, \\ &\qquad 0 \leq w_m \leq 1, \quad m = 1, \ldots, M. \end{aligned}$$

Linear matrix inequality

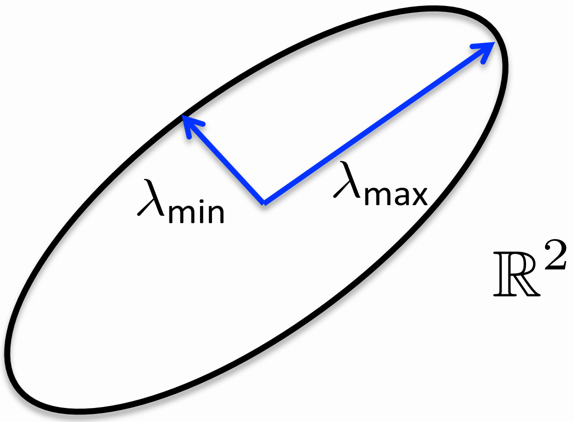
Set of possible solutions

$f(w)$ for estimation – scalar measures

- Prominent scalar measures

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

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



Example: *Linear additive Gaussian model*

$$f(\boldsymbol{w}) = \text{tr}\left(\sum_{m=1}^{M} w_m \sigma_m^{-2} \mathbf{h}_m \mathbf{h}_m^T\right)^{-1}$$

$f(w)$ for estimation – scalar measures

- Prominent scalar measures

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

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

- SDP problem based on ℓ_1-norm heuristics

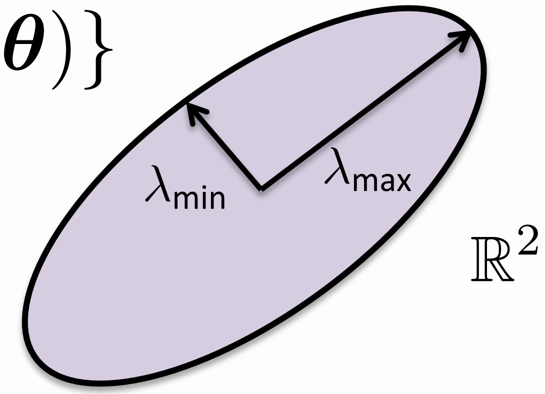
$$\begin{aligned} \arg \min_{\boldsymbol{w} \in \mathbb{R}^M, \, \boldsymbol{x} \in \mathbb{R}^N} \quad & \|\boldsymbol{w}\|_1 \\ \text{s.to} \quad & \begin{bmatrix} \sum_{m=1}^M w_m \boldsymbol{F}_m(\boldsymbol{\theta}) & \boldsymbol{e}_n \\ \boldsymbol{e}_n^T & x_n \end{bmatrix} \succeq \boldsymbol{0}_{N+1}, n = 1, 2, \ldots, N, \, \forall \boldsymbol{\theta} \in \mathcal{T}, \\ & \boldsymbol{1}_N^T \boldsymbol{x} \leq \lambda, \\ & 0 \leq w_m \leq 1, \; m = 1, 2, \ldots, M. \end{aligned}$$

$f(w)$ for estimation – scalar measures

- Prominent scalar measures

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

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



Example: *Linear additive Gaussian model*

$$f(\boldsymbol{w}) = \ln\det\left(\sum_{m=1}^{M} w_m \sigma_m^{-2}\mathbf{h}_m\mathbf{h}_m^T\right)^{-1}$$

✓ **Convex function to be minimized**

Example: *Linear additive Gaussian model*

$$f(\mathcal{X}) = \ln\det\sum_{m\in\mathcal{X}} \sigma_m^{-2}\mathbf{h}_m\mathbf{h}_m^T$$

✓ **Submodular function to be maximized**

[Shamaiah-Banerjee-Vikalo-2010]

- M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," *in Proc. Of 49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 2572–2577.

$f(w)$ for estimation – scalar measures

- Prominent scalar measures
  - **D-optimality** measure (error volume)

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

- Convex problem based on ℓ_1-norm heuristics

$$\arg \min_{\boldsymbol{w} \in \mathbb{R}^M} \quad \mathbf{1}^T \boldsymbol{w}$$

$$\text{s.to } \ln\det\{\sum_{m=1}^{M} w_m \boldsymbol{F}_m(\boldsymbol{\theta})\} \geq \lambda, \forall \boldsymbol{\theta} \in \mathcal{T},$$

$$0 \leq w_m \leq 1, \; m = 1, \ldots, M,$$

Concave function

$f(w)$ for estimation – scalar measures

- Recall, near-optimal greedy algorithm for **submodular** log-det

  - **D-optimality** measure (error volume) for non-linear additive Gaussian model

$$x_m = h_m(\boldsymbol{\theta}) + n_m, m = 1, \cdots, M$$

✓ Linearize $h_m(\boldsymbol{\theta})$ around $\boldsymbol{\theta} \in \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_D\} \subseteq \mathcal{T}$

✓ Weighted log-det $\sum_{d=1}^{D} \pi_d \left(\ln \det \sum_{i \in \mathcal{X}} \boldsymbol{h}_{i,d} \boldsymbol{h}_{i,d}^T\right)$

$\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_d) = \pi_d$ with $0 \leq \pi_d \leq 1$

[Rao-Chepuri-Leus-2015]

✓ Modified log-det

$$f(\mathcal{X}) = \sum_{d=1}^{D} \pi_d \left[\log \det\left(\sum_{i \in \mathcal{X}} \boldsymbol{h}_{i,d} \boldsymbol{h}_{i,d}^T + \epsilon \boldsymbol{I}_N\right) - N \log \epsilon\right]$$

- S. Rao, S.P. Chepuri, and G. Leus. “Greedy Sensor Selection for Non-Linear Models,” In Proc. of the *IEEE Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc. (CAMSAP 2015)*, Cancun, Mexico, December 2015.

$f(w)$ for estimation – scalar measures

- Prominent scalar measures
  - **E-optimality** measure (worst case error)
  - **A-optimality** measure (average error)
  - **D-optimality** measure (error volume)

Think about this figure!

Can't say much about how one performs w.r.t. the other

$f(w)$ for estimation – scalar measures

- Some more scalar measures

  - **T-optimality** measure (to be maximized)

$$f(\boldsymbol{w}) = \mathrm{tr}\{\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta})\}$$

Example: *Linear additive Gaussian model*

$$f(\boldsymbol{w}) = \mathrm{tr}\left\{\sum_{m=1}^{M} w_m \sigma_m^{-2} \mathbf{h}_m \mathbf{h}_m^T\right\}$$

✓ **Linear function to be maximized**

$f(w)$ for estimation – frame potential

- Some more scalar measures (linear additive Gaussian)

$$x_m = \boldsymbol{h}_m^T \boldsymbol{\theta} + n_m, m = 1, \cdots, M$$

Frame potential

$$\sum_{m,n \in \mathcal{X}} |\boldsymbol{h}_m^T \boldsymbol{h}_n|^2$$

✓ orthogonality measure (minimizes MSE)
✓ to be minimized
✓ submodular w.r.t. to the ***complement*** set $\mathcal{X} = \mathcal{M} \backslash \mathcal{S}$

$$f(\mathcal{X}) := \sum_{m,n \in \mathcal{M}} |\boldsymbol{h}_m^T \boldsymbol{h}_n|^2 - \sum_{m,n \in \mathcal{S}} |\boldsymbol{h}_m^T \boldsymbol{h}_n|^2$$

Ensures $f(\mathcal{X})$ is zero for empty input set

[Ranieri-Chebira-Vetterli-2014]

- 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

$f(w)$ for estimation – frame potential

- Frame potential for non-linear additive Gaussian model

$$x_m = h_m(\boldsymbol{\theta}) + n_m, m = 1, \cdots, M$$

- Linearize $h_m(\boldsymbol{\theta})$ around $\boldsymbol{\theta} \in \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_D\} \subseteq \mathcal{T}$

- Weighted frame potential

$$\sum_{d=1}^{D} \pi_d \left(\sum_{i,j \in \mathcal{X}} |\boldsymbol{h}_{i,d}^T \boldsymbol{h}_{j,d}|^2 \right)$$

$\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_d) = \pi_d$ with $0 \leq \pi_d \leq 1$

$f(w)$ for estimation – frame potential

- Modified frame potential for non-linear additive Gaussian case

$$f(\mathcal{X}) := \sum_{d=1}^{D} \pi_d \left(\sum_{i,j \in \mathcal{M}} |\boldsymbol{h}_{i,d}^T \boldsymbol{h}_{j,d}|^2 \right) - \sum_{d=1}^{D} \pi_d \left(\sum_{i,j \in \mathcal{S}} |\boldsymbol{h}_{i,d}^T \boldsymbol{h}_{j,d}|^2 \right)$$

Submodular w.r.t. to the ***complement*** set $\mathcal{X} = \mathcal{M} \backslash \mathcal{S}$

[Rao-Chepuri-Leus-2015]

Hint: weighted sum of submodular functions is submodular

- Frame potential tends to discard rows with a larger norm

  - ✓ Rows with larger norm are more relevant for reducing the MSE
  - ✓ Worse performance when rows have different norms

- S. Rao, S.P. Chepuri, and G. Leus. "Greedy Sensor Selection for Non-Linear Models," In Proc. of the *IEEE Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc. (CAMSAP 2015)*, Cancun, Mexico, December 2015.

Bayesian setting

Suppose prior probability $p(\boldsymbol{\theta})$ is known (e.g., MMSE, MAP)

- Use the Bayesian Cramér-Rao bound

$$\text{Bayesian FIM:} \quad \boldsymbol{F}_{\mathrm{p}} + \sum_{m=1}^{M} w_m \mathbb{E}_{\boldsymbol{\theta}}\{\boldsymbol{F}_m(\boldsymbol{\theta})\}$$

[Chepuri-Leus-2015]

$$\boldsymbol{F}_{\mathrm{p}} = -\mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{\ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\right\}$$

✓ Doesn't depend on the true parameter (averaged over the prior)

Example: *Linear additive Gaussian model*

$$h_m(\boldsymbol{\theta}, n_m) := \mathbf{h}_m^T\boldsymbol{\theta} + n_m \quad \text{and} \quad \boldsymbol{\theta} \sim \mathcal{N}(0, \boldsymbol{F}_p^{-1})$$

Inverse error (FIM) of the MAP/MMSE estimate

$$\boldsymbol{F} = \boldsymbol{F}_p + \sum_{m=1}^{M} w_m \sigma_m^{-2} \boldsymbol{h}_m \boldsymbol{h}_m^T$$

• 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.

Bayesian setting

Suppose prior probability $p(\boldsymbol{\theta})$ is known (e.g., MMSE, MAP)

Example: *Linear additive Gaussian model*

$$h_m(\boldsymbol{\theta}, n_m) := \mathbf{h}_m^T \boldsymbol{\theta} + n_m \quad \text{and} \quad \boldsymbol{\theta} \sim \mathcal{N}(0, \boldsymbol{F}_p^{-1})$$

SDP problem:

$$\arg\min_{\boldsymbol{w}} \quad \mathrm{tr}\left\{\left[\boldsymbol{F}_{\mathrm{p}} + \sum_{m=1}^{M} \sigma_m^{-2} w_m \boldsymbol{h}_m \boldsymbol{h}_m^T\right]^{-1}\right\}$$

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

$$0 \leq w_m \leq 1, \quad m = 1, \ldots, M.$$

Extensions:

- ✓ Underdetermined system (Wiener interpolation or Krigging)
- ✓ Case where $\boldsymbol{F}_{\mathrm{p}}$ does not exist

[Roy-Simonetto-Leus-2016]

- V. Roy, A. Simonetto, and G. Leus. "Spatio-temporal sensor management for environmental field estimation," Signal Processing, vol. 128, pp.369-381, Nov. 2016.

MATLAB script - CVX

```
function selected_locations = cvx_Eoptimality(A,lambda)

M = length(A(:,1)); %nr. of candidate sensors
N = length(A(1,:)); %nr. of parameters
threshold = 0.01;
cvx_begin sdp
variable w(M);
    minimize sum(w)
    subject to
        F = zeros(N,N); %or F_p (prior info. matrix)
        for m=1:M
          F = F + w(m).*(A(:,m)'*A(:,m));
        end
        F >= lambda*eye(N);
        w>=0;
        w<=1;
cvx_end
%deterministic thresholding
what=(w>threshold);
selected_locations = find(what==1);
```

Randomized rounding

Solution from the convex solver

1. Generate L candidates $w_{m,l} = 1$ with a probability $\widehat{w}_m$

2. Define index set

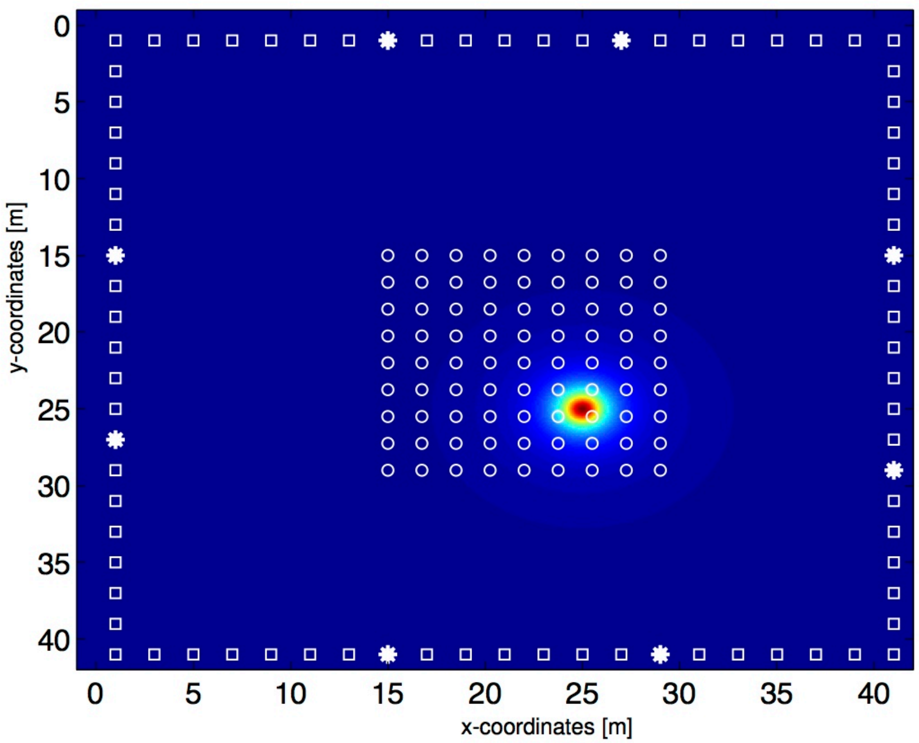
$$\Omega \triangleq \{l \mid f(\boldsymbol{w}_l) \leq \lambda, \forall \boldsymbol{\theta} \in \mathcal{T}, l = 1, 2, \ldots, L\}$$

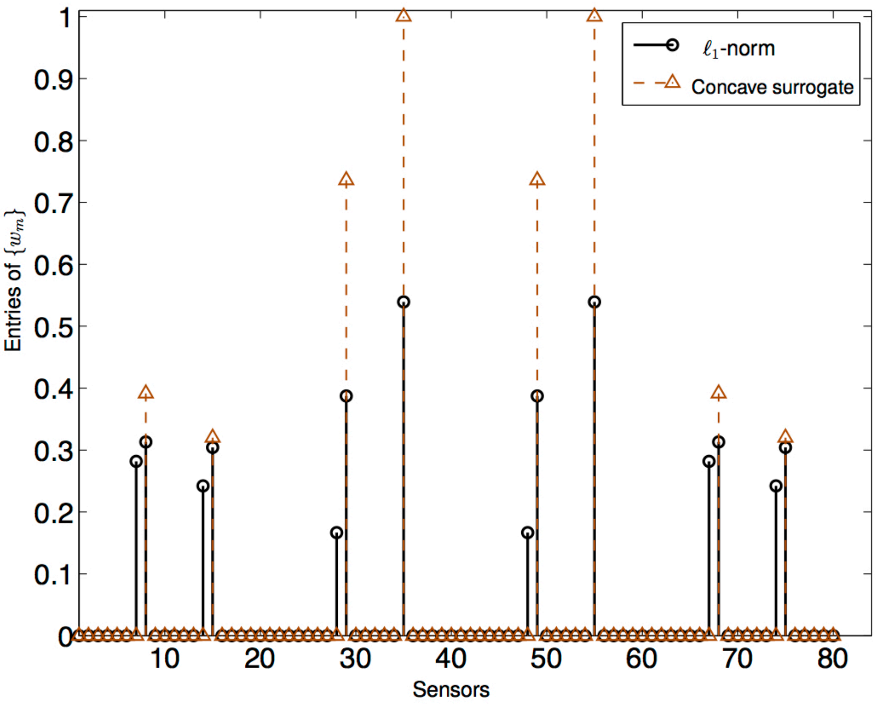
3. If the above set is empty, go back to step 1.

4. Suboptimal Boolean estimate: $\arg\min_{l \in \Omega} \|\boldsymbol{w}_l\|_0$

Examples

Target localization (based on RSS measurements)

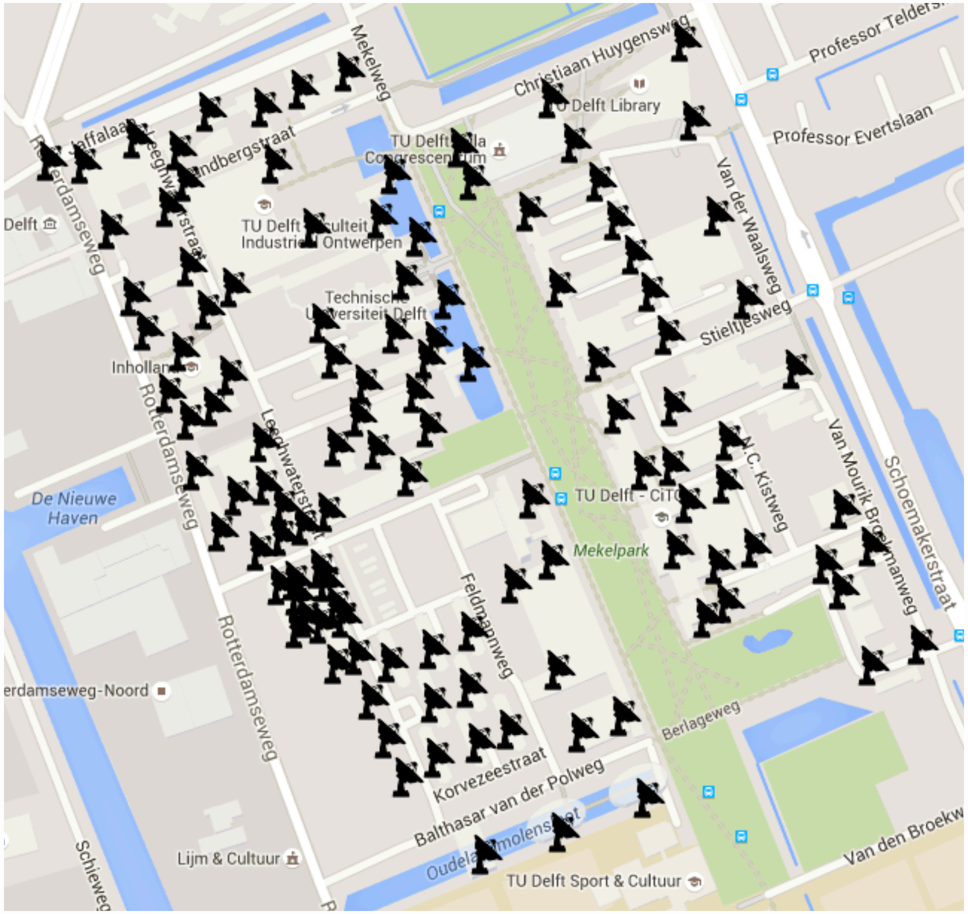


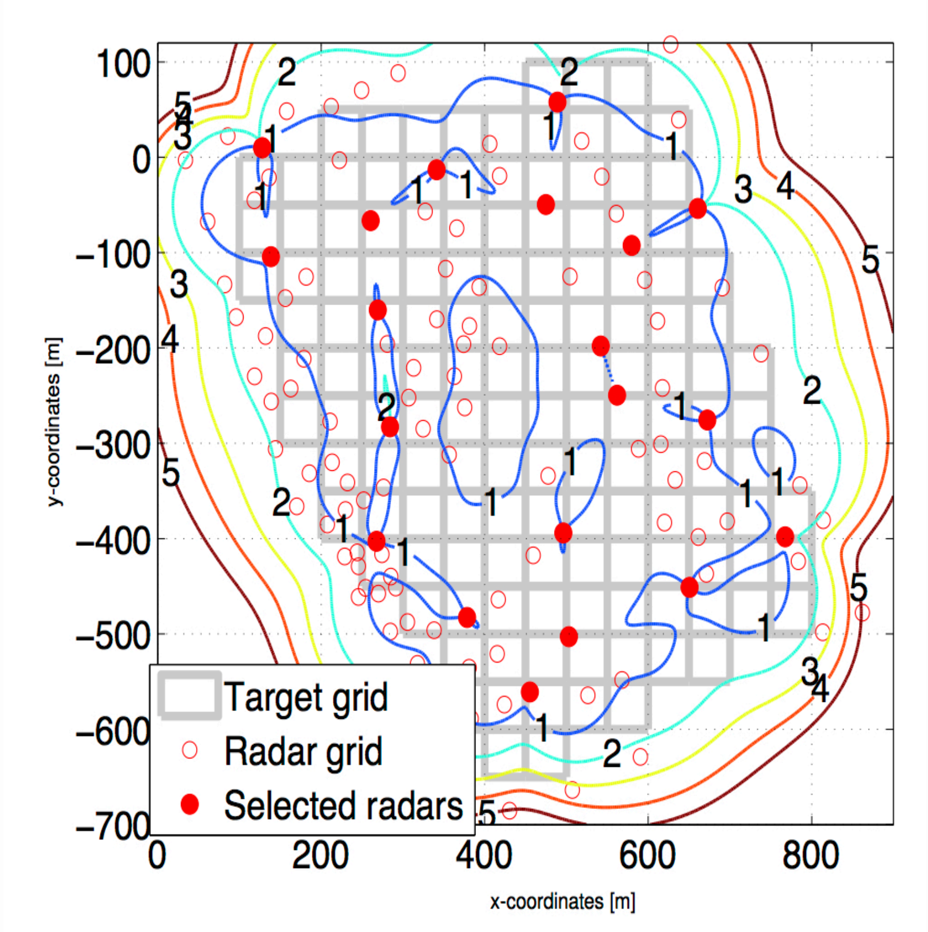


Out of 80 available access point locations, 8 access points are selected.

Examples

Multi-static FMCW radar





Out of 117 available radar positions, 20 radar positions are selected.

[Inna-Leus-Yarovoy-2016]

- I. Ivashko, G. Leus, and A. Yarovoy. "Radar network topology optimization for joint target position and velocity estimation," *Elsevier signal processing*, To appear May 2016.

Statistically dependent observations

Suppose that the errors are not independent

- Observations follow:

$$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{h}(\boldsymbol{\theta}), \boldsymbol{\Sigma}\right)$$

- Fisher information matrix

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

is **no more additive/linear.**

$$\boldsymbol{J}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{w}) = \left(\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{\Sigma}\boldsymbol{\Phi}^T(\boldsymbol{w})\right)^{-1}$$

$F(w, \theta)$ in its current form is non convex in w

Statistically dependent observations

- Express the noise covariance matrix

$$\boldsymbol{\Sigma} = a\boldsymbol{I} + \boldsymbol{S} \quad \text{for any} \quad a \neq 0 \in \mathbb{R} \quad \text{such that} \quad \boldsymbol{S} \quad \text{is invertible}$$

[Chepuri-Leus-2016]

- FIM will be

$$\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta}) = \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{\Phi}^T(\boldsymbol{w})\left(a\boldsymbol{I} + \boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{S}\boldsymbol{\Phi}^T(\boldsymbol{w})\right)^{-1}\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{J}(\boldsymbol{\theta})$$

use matrix inversion lemma and $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \mathrm{diag}(\boldsymbol{w})$

$$\boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) - \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\left[\boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}^{-1}\boldsymbol{J}^T(\boldsymbol{\theta})$$

- S.P. Chepuri and G. Leus. "Sparse Sensing for Estimation with Correlated Observations," *In proc. of Asilomar Conf. Signals, systems, and Computers (Asilomar 2015)*, Pacific Grove, California, USA, November 2015.

Statistically dependent observations

- (E-optimal design) constraint $\lambda_{\min}\{\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta})\} \geq \lambda$

$$\boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) - \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\left[\boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}^{-1}\boldsymbol{J}^T(\boldsymbol{\theta}) \succeq \lambda \boldsymbol{I}_N$$

is equivalent to

$$\begin{bmatrix} \boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) \\ \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1} & \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) - \lambda\boldsymbol{I}_N \end{bmatrix} \succeq \boldsymbol{0}$$

- ✓ an LMI —linear/convex in $\boldsymbol{w}$
- ✓ choose $a > 0$ and $\boldsymbol{S} \succ \boldsymbol{0}$

[Chepuri-Leus-2016]

- S.P. Chepuri and G. Leus. “Sparse Sensing for Estimation with Correlated Observations,” *In proc. of Asilomar Conf. Signals, systems, and Computers (Asilomar 2015)*, Pacific Grove, California, USA, November 2015.

Sampler design: SDP problem

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

$$\text{s.to} \begin{bmatrix} \boldsymbol{S}^{-1} + a^{-1}\text{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) \\ \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1} & \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) - \lambda \boldsymbol{I}_N \end{bmatrix} \succeq \mathbf{0}, \ \forall \boldsymbol{\theta} \in \mathcal{T},$$

$$0 \leq w_m \leq 1, \quad m = 1, \ldots, M.$$

[Chepuri-Leus-2016]

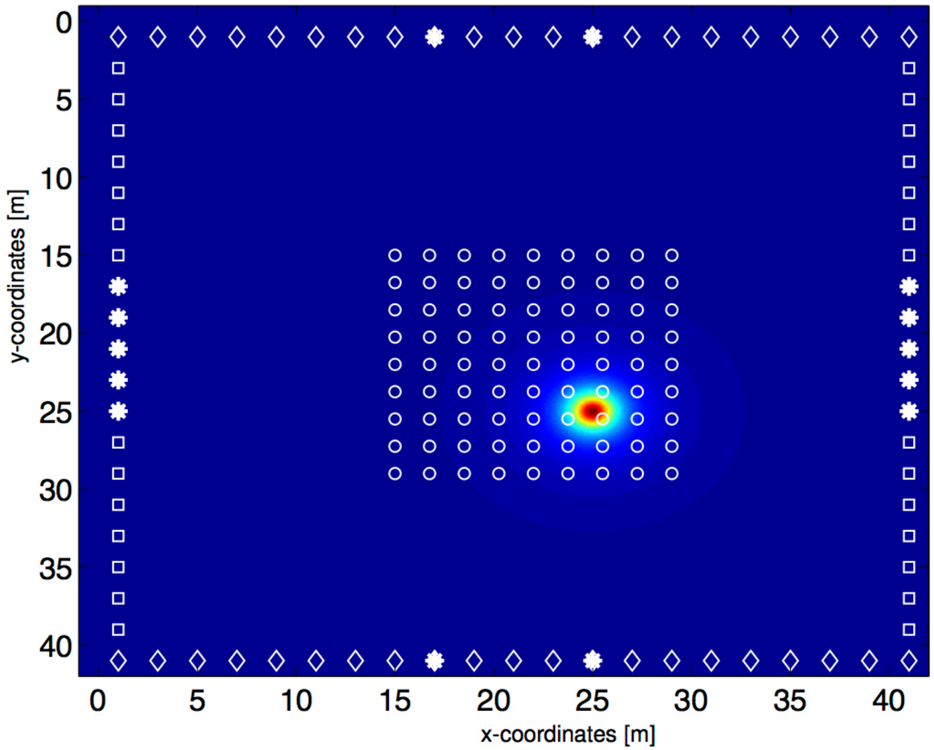
- ✓ Linear model doesn't depend on the true parameter
- ✓ Bayesian setting: prior information can be used as before.

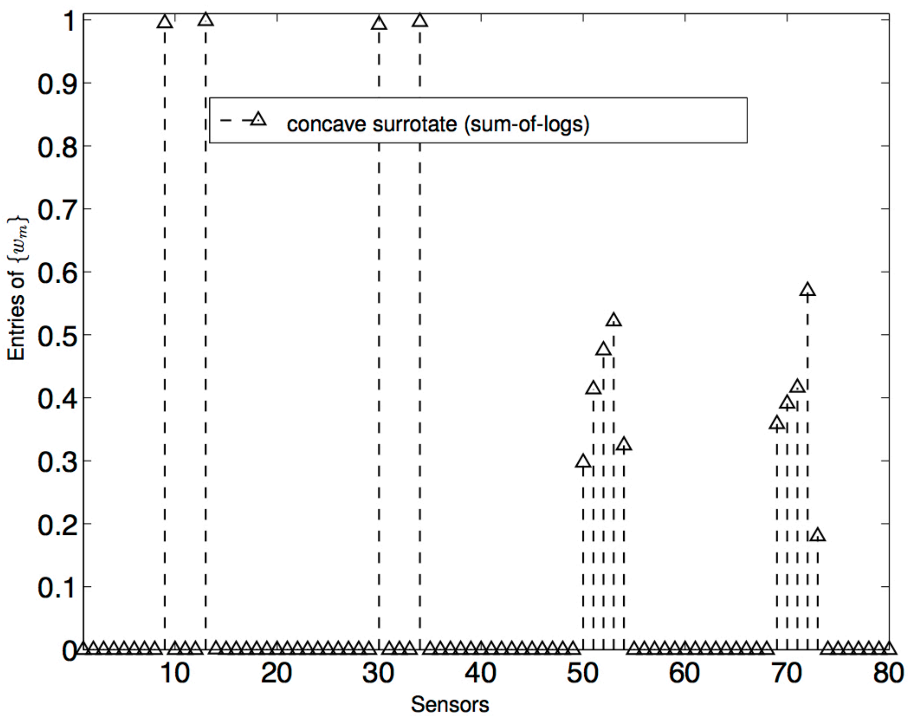
[Liu-Chepuri-Leus-Varshney-2016]

- S.P. Chepuri and G. Leus. "Sparse Sensing for Estimation with Correlated Observations," In proc. of *Asilomar Conf. Signals, systems, and Computers (Asilomar 2015)*, Pacific Grove, California, USA, November 2015.
- S. Liu, S.P. Chepuri, M. Fardad, E. Maşazade, G. Leus G, P.K. Varshney. "Sensor Selection for Estimation with Correlated Measurement Noise," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3509-22, Aug. 2015.

Example: Target localization

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

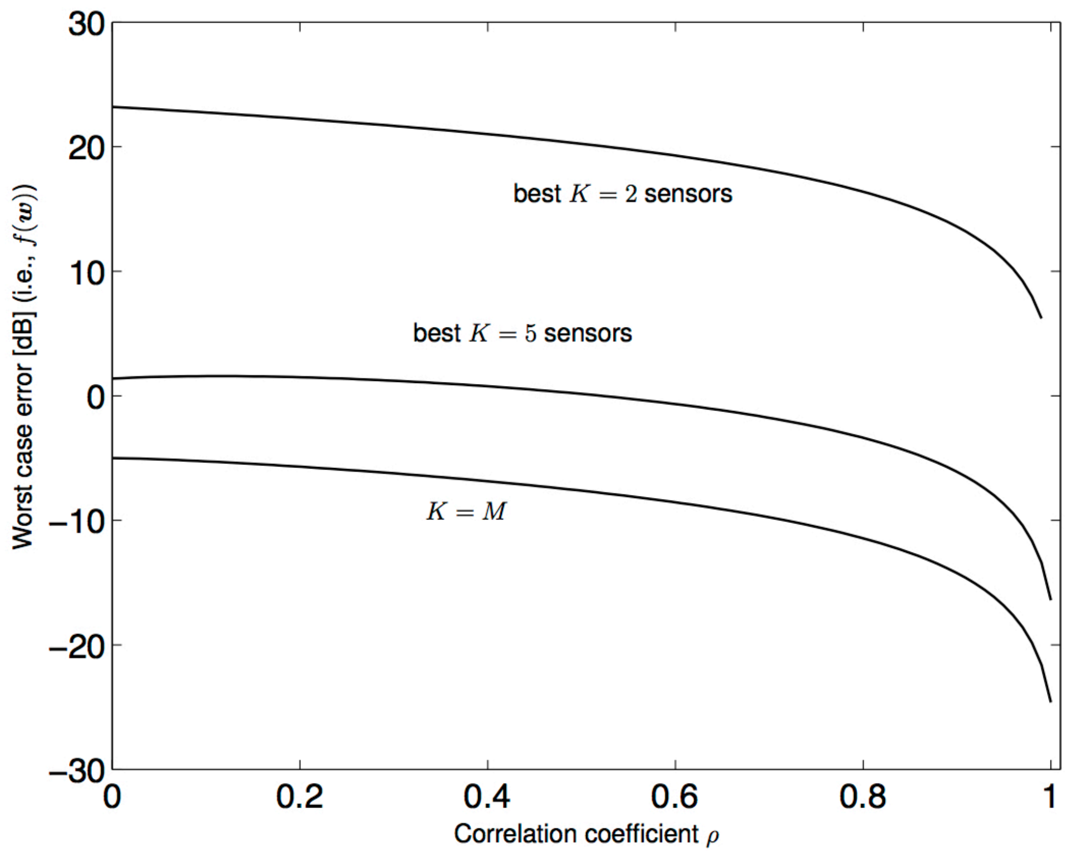




Out of 80 available uncorrelated and correlated access point locations, 14 access points are selected.

Is correlation good?

- Linear model, Gaussian regression matrix
- Equicorrelated correlation matrix: $\boldsymbol{\Sigma} = \left[(1-\rho)\boldsymbol{I} + \rho\mathbf{1}\mathbf{1}^T\right]$



Nr. of required sensors (worst case error) reduces as sensors become more coherent

Filtering

1. S.P. Chepuri and G. Leus. **Sensor Selection for Estimation, Filtering, and Detection**. In Proc. of the International Conference on Signal Processing and Communications (SPCOM 2014), Bangalore, India, July 2014.

2. S.P. Chepuri and G. Leus. **Sparsity-Promoting Adaptive Sensor Selection for Non-linear Filtering**. In Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), Florence, Italy, May 2014.

Time-varying states

- Unknown parameter $\boldsymbol{\theta}_k$ obeys the state-space equations

$$\text{measurements: } y_{k,m} = w_{k,m} \overbrace{h_{k,m}(\boldsymbol{\theta}_k, n_{k,m})}^{x_{k,m} \sim p_{k,m}(x;\boldsymbol{\theta}_k)}, \quad m = 1, 2, \ldots, M,$$

$$\text{dynamics: } \boldsymbol{\theta}_{k+1} = \boldsymbol{A}_k \boldsymbol{\theta}_k + \boldsymbol{u}_k.$$

$$\mathbb{E}\{\boldsymbol{u}_k \boldsymbol{u}_k^H\} = \boldsymbol{\Sigma}_{\boldsymbol{u}}$$

- **Design** the **sequence** of time-varying sensing vectors:

$$\boldsymbol{w}_k = [w_{k,1}, w_{k,2}, \ldots, w_{k,M}]^T \in [0, 1]^M$$

based on the entire **history** of measurements up to that point

Performance metric – Posterior CRB

- Use the **Posterior Cramér-Rao bound** as the metric

$$\mathbb{E}\left\{(\widehat{\boldsymbol{\theta}}_{k|k}-\boldsymbol{\theta}_k)(\widehat{\boldsymbol{\theta}}_{k|k}-\boldsymbol{\theta}_k)^T\right\} \geq \boldsymbol{C}_k = \boldsymbol{F}_k^{-1}$$

- Posterior-FIM can be expressed as

$$\boldsymbol{F}_k(\boldsymbol{w}_k, \{\boldsymbol{\theta}_{\kappa-1}\}_{\kappa=1}^{k}, \boldsymbol{\theta}_k) = \overbrace{(\boldsymbol{\Sigma}_u + \boldsymbol{A}_k \boldsymbol{F}_{k-1}^{-1}(\{\boldsymbol{\theta}_{\kappa-1}\}_{\kappa=1}^{k})\boldsymbol{A}_k^T)^{-1}}^{\boldsymbol{F}_{\mathrm{p},k-1}(\{\boldsymbol{\theta}_{\kappa-1}\}_{\kappa=1}^{k})} + \underbrace{\sum_{m=1}^{M} w_{k,m}\boldsymbol{F}_{k,m}(\boldsymbol{\theta}_k)}$$

Depends on all the previous states

related to measurements
(measurements are statistically independent)

Performance metric – Posterior CRB

- To reduce the computational complexity, the prior Fisher can be simply evaluated at the **past estimate** $\tilde{\theta}_{k-1} := \widehat{\theta}_{k-1|k-1}$

$$\boldsymbol{F}_k(\boldsymbol{w}_k, \boldsymbol{\theta}_k) \approx \overbrace{(\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{A}_k \boldsymbol{F}_{k-1}^{-1}(\tilde{\boldsymbol{\theta}}_{k-1}) \boldsymbol{A}_k^T)^{-1}}^{\boldsymbol{F}_{\mathrm{prior},k-1}(\tilde{\boldsymbol{\theta}}_{k-1})} + \boldsymbol{F}_{\mathrm{obs},k}(\boldsymbol{w}_k, \boldsymbol{\theta}_k)$$

[Chepuri-Leus-2014]

✓ Equal to prediction error of the extended Kalman filter for $\boldsymbol{\theta}_k := \widehat{\boldsymbol{\theta}}_{k|k-1}$ (prediction)

[Masazade-Farad-Varshney-2012]

- S.P. Chepuri and G. Leus. "Sparsity-Promoting Adaptive Sensor Selection for Non-linear Filtering," *In Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italy, May 2014.
- E. Masazade, M. Fardad, and P.K. Varshney. "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 845-848, Dec. 2012.

Performance metric – Posterior CRB

- To reduce the computational complexity, the prior Fisher can be simply evaluated at the **past estimate** $\tilde{\boldsymbol{\theta}}_{k-1} := \widehat{\boldsymbol{\theta}}_{k-1|k-1}$

$$\boldsymbol{F}_k(\boldsymbol{w}_k, \boldsymbol{\theta}_k) \approx \overbrace{(\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{A}_k \boldsymbol{F}_{k-1}^{-1}(\tilde{\boldsymbol{\theta}}_{k-1}) \boldsymbol{A}_k^T)^{-1}}^{\boldsymbol{F}_{\text{prior},k-1}(\tilde{\boldsymbol{\theta}}_{k-1})} + \boldsymbol{F}_{\text{obs},k}(\boldsymbol{w}_k, \boldsymbol{\theta}_k)$$

Example: *Linear additive Gaussian filtering*

$$h_{k,m}(\boldsymbol{\theta}_k, n_{k,m}) := \boldsymbol{h}_{k,m}^T \boldsymbol{\theta}_k + n_{k,m} \quad \text{and} \quad \boldsymbol{u}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{u}})$$

Posterior FIM (equal to posterior error covariance of a Kalman filter)

$$\boldsymbol{F}_k(\boldsymbol{w}_k) = (\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{A}_k \boldsymbol{F}_{k-1}^{-1} \boldsymbol{A}_k^T)^{-1} + \frac{1}{\sigma^2} \sum_{m=1}^{M} w_{k,m} \boldsymbol{h}_{k,m} \boldsymbol{h}_{k,m}^T$$

Doesn’t depend on the past or present states

Sampler design

- SDP problem (E-optimality)

around the prediction

$$\arg \min_{\boldsymbol{w}_k \in [0,1]^M} \mathbf{1}^T \boldsymbol{w}_k$$

$$\text{s.to} \quad \boldsymbol{F}_{\mathrm{p},k-1} + \sum_{m=1}^{M} w_{k,m} \boldsymbol{F}_{k,m}(\boldsymbol{\theta}_k) \succeq \lambda \boldsymbol{I}_N, \forall \boldsymbol{\theta}_k \in \mathcal{T}_k$$

$$0 \leq w_m \leq 1, \quad m = 1, \ldots, M.$$

- Greedy algorithm (D-optimality)

How about
“**frame potential**”?

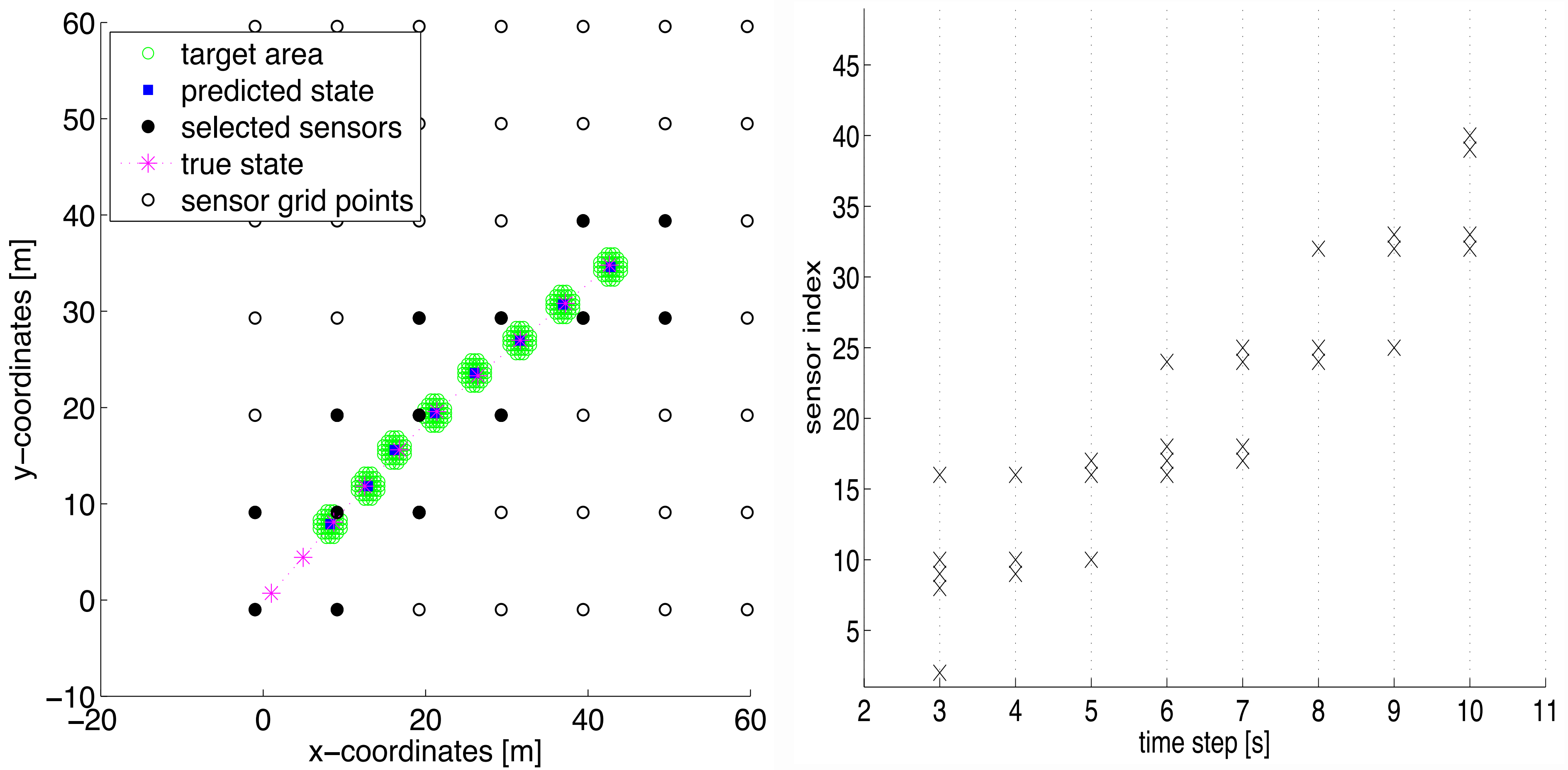
Example: *Linear additive Gaussian filtering*

$$f(\mathcal{X}) = \ln\det\left(\boldsymbol{F}_{\mathrm{p},k-1} + \frac{1}{\sigma^2}\sum_{i\in\mathcal{X}} \boldsymbol{h}_{k,i}\boldsymbol{h}_{k,i}^T\right)$$

[Shamaiah-Banerjee-Vikalo-2010]

- M. Shamaiah, S. Banerjee, and H. Vikalo, “Greedy sensor selection: Leveraging submodularity,” *in Proc. Of 49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 2572–2577.

Scheduling example – Target tracking



Schedule 49 available access points

Dependent measurements case

Measurements are not statistically independent

$$\boldsymbol{x}_k \sim \mathcal{N}\left(\boldsymbol{h}_k(\boldsymbol{\theta}_k), \boldsymbol{\Sigma}_k\right)$$

- Recall the decomposition

$\boldsymbol{\Sigma}_k = a_k \boldsymbol{I} + \boldsymbol{S}_k$ for any $a_k \neq 0 \in \mathbb{R}$ such that $\boldsymbol{S}_k$ is invertible

- E-optimality constraint can be equivalently expressed as

$$\begin{bmatrix} \boldsymbol{S}_k^{-1} + a_k^{-1}\mathrm{diag}(\boldsymbol{w}_k) & \boldsymbol{S}_k^{-1}\boldsymbol{J}_k(\boldsymbol{\theta}_k) \\ \boldsymbol{J}_k^T(\boldsymbol{\theta}_k)\boldsymbol{S}_k^{-1} & \boldsymbol{F}_{\mathrm{prior},k-1} + \boldsymbol{J}_k^T(\boldsymbol{\theta})\boldsymbol{S}_k^{-1}\boldsymbol{J}_k(\boldsymbol{\theta}_k) - \lambda \boldsymbol{I}_N \end{bmatrix} \succeq \boldsymbol{0}_{M+N}$$

Structured time-varying states

Time-varying compressive sensing

- Unknown parameter $\boldsymbol{\theta}_k$ obeys the state-space equations

measurements: $\boldsymbol{y}_k = \mathrm{diag}_{\mathrm{r}}(\boldsymbol{w}_k)\boldsymbol{H}_k\boldsymbol{\theta}_k + \boldsymbol{n}_k$

dynamics: $\boldsymbol{\theta}_k = \boldsymbol{A}_k\boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k$

pseudo-measurement: $0 = r(\boldsymbol{\theta}_k) + e_k$

- $r(\boldsymbol{\theta}_k)$ enforces **structure** (e.g., sparsity, smoothness,...)

[Carmi-Gurfil-Kanevsky-2010]

- Can we use sparse sensing framework to design *sparse compressive sensing matrices*?

- A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2405–2409, April 2010.

Performance metric

- Inverse error covariance

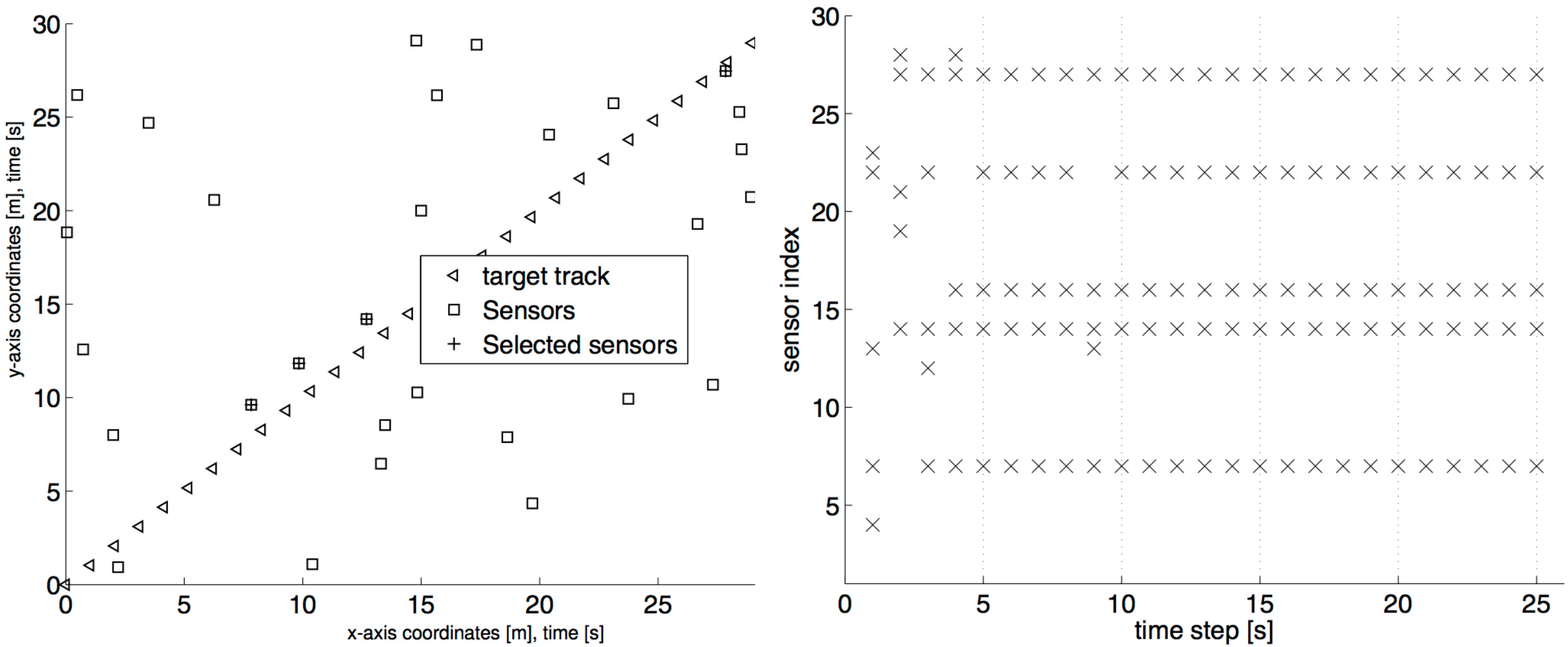
$$\boldsymbol{P}_{k|k}^{-1} = \underbrace{\boldsymbol{P}_{k|k-1}^{-1}}_{\text{dynamics}} + \underbrace{\partial r(\widehat{\boldsymbol{\theta}}_{k|k-1})\partial r(\widehat{\boldsymbol{\theta}}_{k|k-1})^T}_{\text{sparsity prior/ pseudo-measurement}} + \underbrace{\sum_{m=1}^{M} w_{k,m}\boldsymbol{h}_{k,m}\boldsymbol{h}_{k,m}^T}_{\text{measurements}}$$

$\boldsymbol{h}_{k,m}$: mth row of the dictionary $\boldsymbol{H}_k$

$\partial r(\widehat{\boldsymbol{\theta}}_{k|k-1})$: (sub)gradient of $r(\boldsymbol{\theta}_k)$ towards $\boldsymbol{\theta}_k$ at $\widehat{\boldsymbol{\theta}}_{k|k-1}$

- Performance metric $f(\boldsymbol{w}_k) = \mathrm{tr}\{\boldsymbol{P}_{k|k}\}$ depends on $\boldsymbol{\theta}_k$

Example – grid based target localization



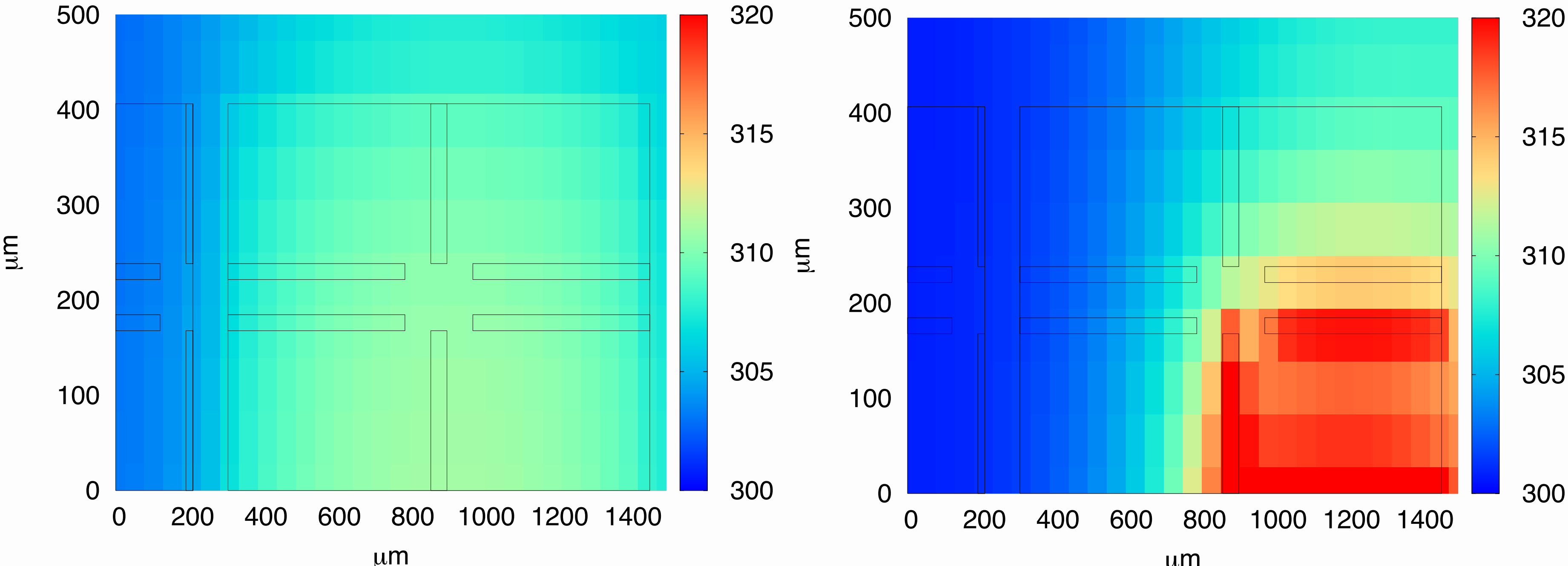
Out of 30 available sensors, 5 sensors are selected (5 measurement equations in 30 unknowns)

Detection

1. S.P. Chepuri and G. Leus. **Sparse Sensing for Distributed Detection**. IEEE Trans. on Signal Processing, 16(6): 1446-1460, Mar. 2016.

2. 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**)

Binary hypothesis testing



$\mathcal{H}_0$: No Hot spot

$\mathcal{H}_1$: Hot spot

Other applications:

- ✓ spectrum sensing, anomaly detection
- ✓ radar and sonar systems

Detection problem

- Observations follow (binary hypothesis testing)

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

- Gaussian observations

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

- Generalization to other distributions

  [Chepuri-Leus-16]

- 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.
- 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.

Sensing design for detection

Neyman Pearson setting

$$\arg \min_{\boldsymbol{w} \in \{0,1\}^M} \|\boldsymbol{w}\|_0$$

$$\text{s.to} \quad P_f(\boldsymbol{w}) \le \alpha, \quad P_m(\boldsymbol{w}) \le \beta$$

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

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

Bayesian setting

$$\arg \min_{\boldsymbol{w} \in \{0,1\}^M} \|\boldsymbol{w}\|_0$$

$$\text{s.to} \quad P_e(\boldsymbol{w}) \le e$$

π_0, π_1 prior probabilities

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

- Error probabilities (in general) do not admit expressions suitable for numerical optimization
  - ✓ Seek weaker performance measures

Bayesian setting

- Decision is based upon the log-likelihood ratio test

$$\log l(\boldsymbol{y}) = \log \frac{p(\boldsymbol{y}|\mathcal{H}_1)}{p(\boldsymbol{y}|\mathcal{H}_0)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \log \frac{\pi_0}{\pi_1}$$

- Best error exponent: **Chernoff distance**

$$\begin{aligned} \mathcal{C}(\mathcal{H}_1 \| \mathcal{H}_0) &= -\log \min_{0 \le n \le 1} \int [p(\boldsymbol{y}|\mathcal{H}_1)]^n [p(\boldsymbol{y}|\mathcal{H}_0)]^{1-n} d\boldsymbol{y} \\ &= -\log \underbrace{\min_{0 \le n \le 1}} \mathbb{E}_{|\mathcal{H}_0}\{[l(\boldsymbol{y})]^n\} \end{aligned}$$

Minimization complicates the sensing design

$$n = 0.5?$$

Bhattacharyya distance

- Special case of Chernoff distance: Bhattacharyya distance

$$\mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = -\log \rho$$

- Hellinger or Bhattacharyya coefficient: **average root likelihood ratio**

$$\rho = \int \sqrt{p(\boldsymbol{y}|\mathcal{H}_1)p(\boldsymbol{y}|\mathcal{H}_0)}d\boldsymbol{y} = \int p(\boldsymbol{y}|\mathcal{H}_0)\sqrt{\frac{p(\boldsymbol{y}|\mathcal{H}_1)}{p(\boldsymbol{y}|\mathcal{H}_0)}}d\boldsymbol{y}$$

$$= \mathbb{E}_{|\mathcal{H}_0}\left\{\sqrt{l(\boldsymbol{y})}\right\}$$

✓ **Symmetric:**

$$\mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = \mathcal{B}(\mathcal{H}_0 \| \mathcal{H}_1)$$

Bhattacharyya distance

- Upper bounds the error probabilities

$$\frac{1}{2}\min(\pi_0, \pi_1)\rho^2 \leq P_e \leq \sqrt{\pi_0 \pi_1}\rho$$

[Kadota-Shepp-1967], [Kailath-1967]

- T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two Gaussian signals," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 278–284, 1967.
- T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 1, pp. 52–60, Feb. 1967.

Neyman Pearson setting

- Decision is based upon the log-likelihood ratio test

$$\log l(\boldsymbol{y}) = \log \frac{p(\boldsymbol{y}|\mathcal{H}_1)}{p(\boldsymbol{y}|\mathcal{H}_0)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma$$

- Best error exponent – **Kullback-Leibler distance**

[Cover-Thomas-2012]

$$\log P_m \overset{as.}{=} -\mathcal{K}(\mathcal{H}_1 \| \mathcal{H}_0) \quad \text{for} \quad P_m \to 0$$

- **Average log-likelihood ratio**

$$\begin{aligned}\mathcal{K}(\mathcal{H}_1 \| \mathcal{H}_0) &= \mathbb{E}_{|\mathcal{H}_1}\{\log l(\boldsymbol{y})\} \\ &= \int \log l(\boldsymbol{y}) p(\boldsymbol{y}|\mathcal{H}_1) d\boldsymbol{y}\end{aligned}$$

- T. M. Cover and J. A. Thomas. “Elements of Information Theory,” New York, NY, USA: Wiley, 2012.

Kullback-Leibler distance

- Upper bound on *probability of miss detection*

[Chepuri-Leus-2016]

$$P_m \leq \frac{1}{1 + \frac{(\mathcal{K}(\mathcal{H}_1 \| \mathcal{H}_0) - \log \gamma)^2}{v^2}}$$

v^2: variance of the log-likehood ratio

- 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.
- T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two Gaussian signals," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 278–284, 1967.

Kullback-Leibler distance

- **J-divergence**: symmetric form of the Kullback-Leibler distance

$$\mathcal{D}(\mathcal{H}_1 \| \mathcal{H}_0) = \mathcal{K}(\mathcal{H}_1 \| \mathcal{H}_0) + \mathcal{K}(\mathcal{H}_0 \| \mathcal{H}_1)$$

  - ✓ reasonable for $\pi_0 = \pi_1 = 0.5$ and for Gaussian observations (upper bounds error prob.)

[Kadota-Shepp-1967]

- 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.
- T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two Gaussian signals," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 278–284, 1967.

Performance metric

- Kullback-Leibler dist., J-divergence, Bhattacharyya dist.
  - ✓Don't depend on the actual data
  - ✓Depend on the model
  - ✓Can be computed offline

Example: *Gaussian - uncommon means and common covariances*

$$\mathcal{H}_0: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}_0, \sigma^2 \mathbf{I}) \quad \text{and} \quad \mathcal{H}_1: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}_1, \sigma^2 \mathbf{I})$$

Bhattacharyya distance, Kullback-Leibler distance, J-divergence are all the **same** up to a constant:

$$\mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = s(\boldsymbol{w})/8 \quad \mathcal{K}(\mathcal{H}_1 \| \mathcal{H}_0) = s(\boldsymbol{w})/2 \quad \mathcal{D}(\mathcal{H}_1 \| \mathcal{H}_0) = s(\boldsymbol{w})$$

$$\text{SNR:} \quad s(\boldsymbol{w}) = \sigma^{-2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathrm{diag}(\boldsymbol{w})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$$

Statistically independent observations

- Assuming conditionally independent observations

$$l(\boldsymbol{y}) = \prod_{m=1}^{M} \left[\frac{p(y_m|\mathcal{H}_1)}{p(y_m|\mathcal{H}_0)} \right]^{w_m}$$

- Distance measures are **additive**:

(KL distance) $\mathcal{K}(\mathcal{H}_1 \| \mathcal{H}_0) = \mathbb{E}_{|\mathcal{H}_1}\{\log l(\boldsymbol{y})\}$

$$= \sum_{m=1}^{M} w_m \underbrace{\mathbb{E}_{|\mathcal{H}_1}\{\log l_m(x)\}}_{\mathcal{K}_m}$$

Local distance

(Bhattacharyya distance) $\mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = -\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l(\boldsymbol{y})}\}$

$$= -\sum_{m=1}^{M} w_m \underbrace{\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l_m(x)}\}}_{\mathcal{B}_m}$$

Recall: additive property of FIM

Sensing design

- Linear program

$$\begin{aligned} &\arg\min_{\boldsymbol{w}} \quad \|\boldsymbol{w}\|_0 \\ &\text{s.to} \quad \sum_{m=1}^{M} w_m d_m \geq \lambda, \\ &\qquad\quad w_m \in \{0,1\}, m = 1, 2, \ldots, M \end{aligned}$$

Classical setting $d_m := \{\mathcal{K}_m\}_{m=1}^{M}$
Bayesian setting $d_m := \{\mathcal{B}_m\}_{m=1}^{M}$

Do we need convex relaxation(s) or submodular greedy algorithms?

Sensing design

- Linear program with **explicit** solution

$$\begin{aligned} \arg\min_{\boldsymbol{w}} \quad & \|\boldsymbol{w}\|_0 \\ \text{s.to} \quad & \sum_{m=1}^{M} w_m d_m \geq \lambda, \\ & w_m \in \{0,1\}, m = 1, 2, \ldots, M \end{aligned}$$

Classical setting $d_m := \{\mathcal{K}_m\}_{m=1}^{M}$
Bayesian setting $d_m := \{\mathcal{B}_m\}_{m=1}^{M}$

- *Solution: Ordering!*

- *The best subset of sensors: sensors with largest average log/root local likelihood ratio*

Example

Example: *Gaussian - uncommon means and common covariances*

$$\mathcal{H}_0: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}_0, \sigma^2 \mathbf{I}) \quad \text{and} \quad \mathcal{H}_1: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}_1, \sigma^2 \mathbf{I})$$

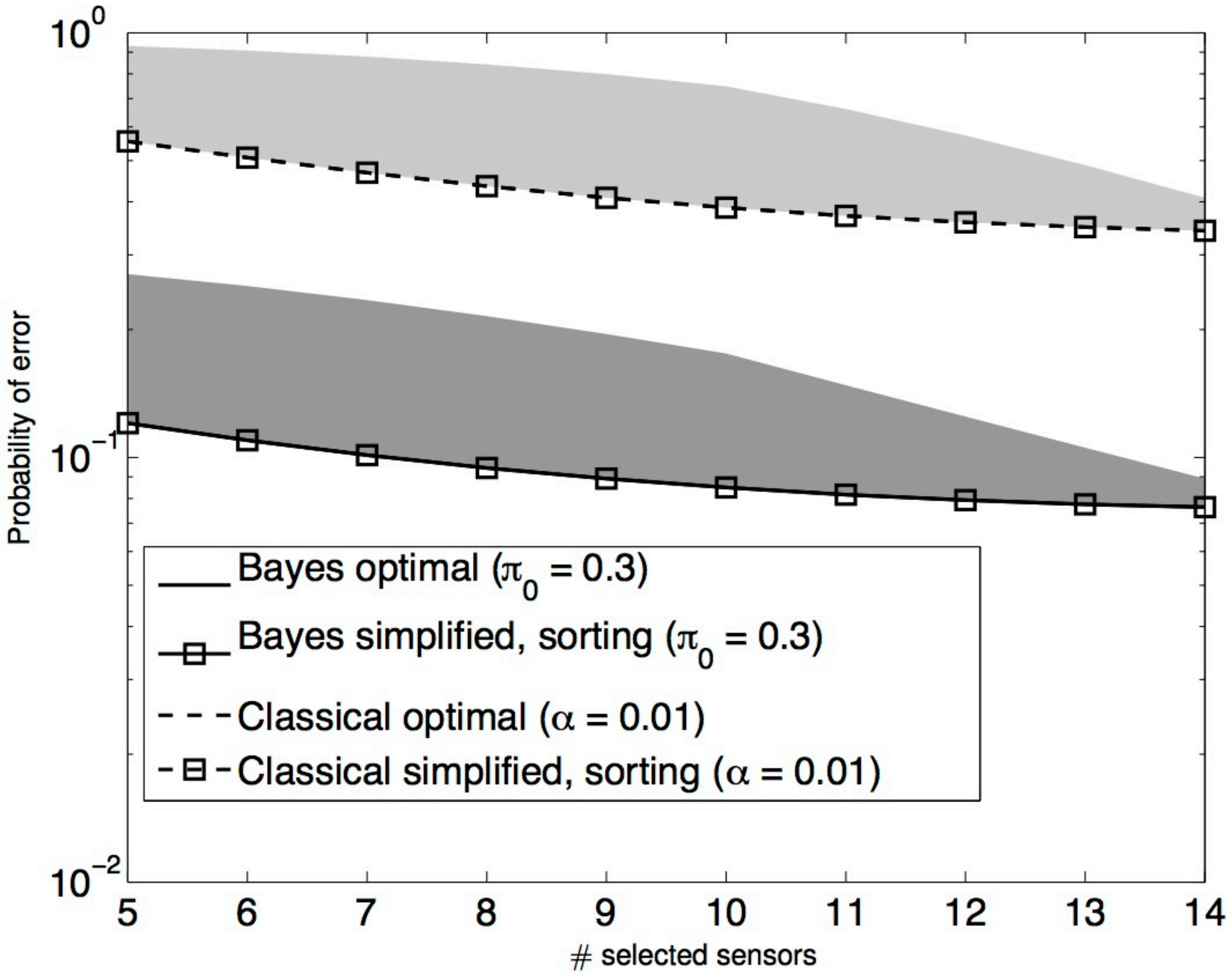
Bhattacharyya distance, Kullback-Leibler distance, J-divergence are all the **same** up to a constant:

$$\mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = s(\boldsymbol{w})/8 \quad \mathcal{K}(\mathcal{H}_1 \| \mathcal{H}_0) = s(\boldsymbol{w})/2 \quad \mathcal{D}(\mathcal{H}_1 \| \mathcal{H}_0) = s(\boldsymbol{w})$$

SNR: $s(\boldsymbol{w}) = \sigma^{-2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathrm{diag}(\boldsymbol{w})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$

Sensors with the largest local SNR values are optimal

Example



Shaded regions denote the performance of the possible suboptimal samplers

Dependent (Gaussian) observations

- Suppose

$$\mathcal{H}_0: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}) \quad \text{vs.} \quad \mathcal{H}_1: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}_1, \boldsymbol{\Sigma})$$

Not diagonal or scaled identity

- Distance measure (or SNR) is **not additive**

$$s(w) = [\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{m}]^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{w}) [\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{m}]$$

$$\boldsymbol{m} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$$
$$\boldsymbol{\Sigma}(\boldsymbol{w}) = \boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{\Sigma}\boldsymbol{\Phi}^T(\boldsymbol{w})$$

Dependent (Gaussian) observations

- Express (as before)

$$\boldsymbol{\Sigma} = a\boldsymbol{I} + \boldsymbol{S} \quad \text{for any} \quad a \neq 0 \in \mathbb{R} \quad \text{such that} \quad \boldsymbol{S} \quad \text{is invertible}$$

- Constraint $d(\boldsymbol{w}) \geq \lambda$

$$\boldsymbol{m}^T \boldsymbol{S}^{-1} \boldsymbol{m} - \boldsymbol{m}^T \boldsymbol{S}^{-1} \left[\boldsymbol{S}^{-1} + a^{-1} \mathrm{diag}(\boldsymbol{w})\right]^{-1} \boldsymbol{S}^{-1} \boldsymbol{m} \;\geq\; \lambda$$

is equivalent to an LMI:

$$\begin{bmatrix} \boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{m} \\ \boldsymbol{m}^T \boldsymbol{S}^{-1} & \boldsymbol{m}^T \boldsymbol{S}^{-1} \boldsymbol{m} - \lambda \end{bmatrix} \succeq \boldsymbol{0}$$

Dependent case: solver

Relaxed SDP

How about submodular algorithms?

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

$$\text{s.to} \begin{bmatrix} \boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{m} \\ \boldsymbol{m}^T\boldsymbol{S}^{-1} & \boldsymbol{m}^T\boldsymbol{S}^{-1}\boldsymbol{m} - \lambda \end{bmatrix} \succeq \mathbf{0},$$

$$0 \le w_m \le 1, \quad m = 1, \ldots, M.$$

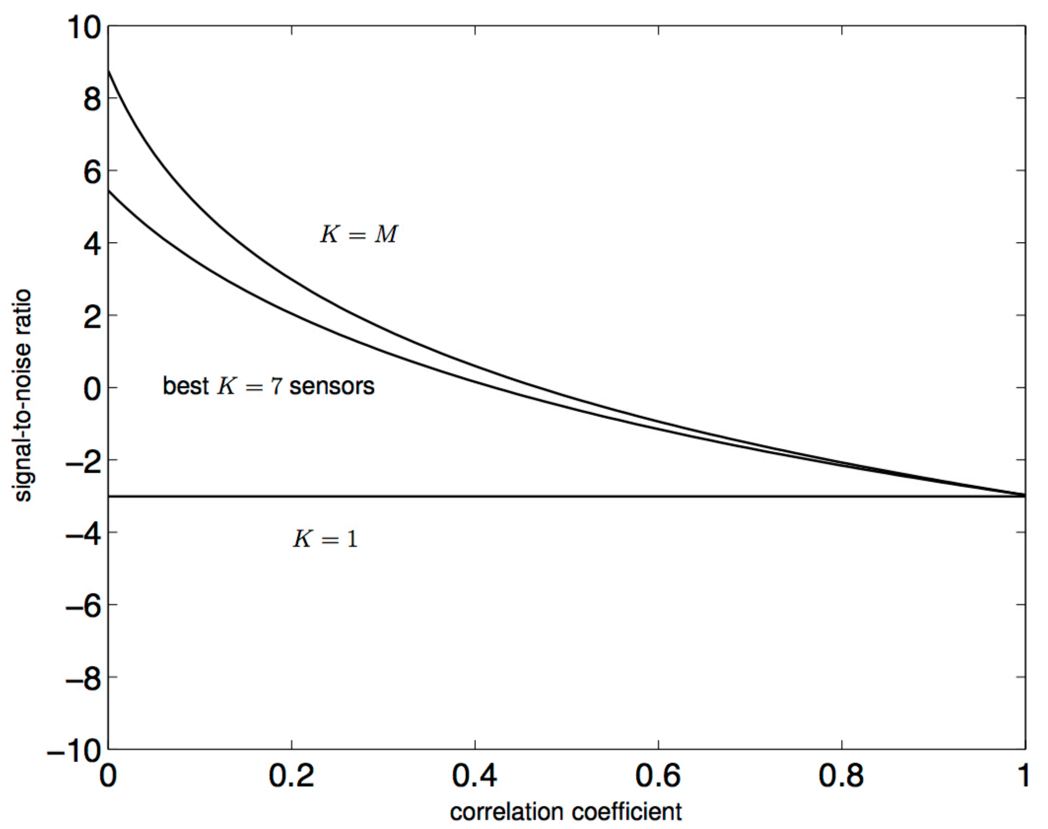
Randomized rounding (similar to estimation)

MATLAB – CVX (Formulation 2)

```
function sel_loc = cvx_detection(theta1,theta0,Sigma,K)

M = length(theta1);
a=0.11; S = Sigma - a*eye(M);

cvx_begin sdp
    variable w(M)
    variable t
    minimize t;
    subject to
    sum(w) ==  K;
    [inv(S)+ a^-1*diag(w), inv(S)*(theta1-theta0);
    (theta1-theta0)'*inv(S) , t] >= semidefinite(M+1);
    w<=1;
    w>=0;
cvx_end

%deterministic thresholding
wsort=sort(w); threshold =wsort(M-K); what=(w>threshold);
sel_loc = find(what==1);
```

Is correlation good or bad?

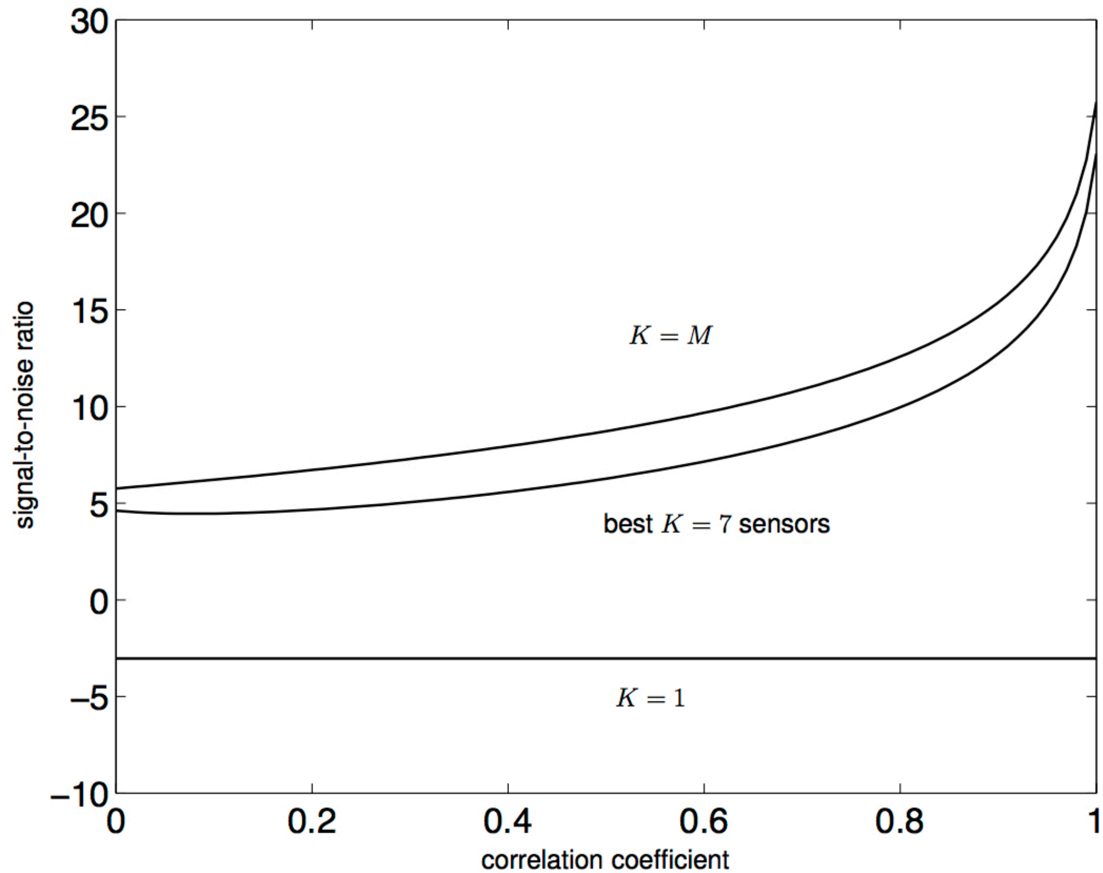
Equicorrelated correlation matrix: $\boldsymbol{\Sigma} = \left[(1-\rho)\boldsymbol{I} + \rho\mathbf{1}\mathbf{1}^T\right]$



Identical sensors: $\boldsymbol{\theta}_0 = c_0\mathbf{1}$ and $\boldsymbol{\theta}_1 = c_1\mathbf{1}$

Is correlation good or bad?

Equicorrelated correlation matrix: $\boldsymbol{\Sigma} = \left[(1-\rho)\boldsymbol{I} + \rho\mathbf{1}\mathbf{1}^T\right]$



Non-Identical sensors: $\boldsymbol{\theta}_1 \neq c_1\mathbf{1}$ and $\boldsymbol{\theta}_0 \neq c_2\mathbf{1}$

Nr. of required sensors reduces as sensors become more coherent

Dependent (Gaussian) observations

- Suppose the conditional observations have uncommon variances

$$\mathcal{H}_0: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}_0)$$
$$\mathcal{H}_1: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}_1)$$

Not diagonal

- Distance measures are different, and complicated

- For example, J-divergence will be

$$\mathcal{D}(\mathcal{H}_1 || \mathcal{H}_0) = \frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{w})\boldsymbol{\Sigma}_1(\boldsymbol{w})\} + \frac{1}{2}\text{tr}\{\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{w})\boldsymbol{\Sigma}_0(\boldsymbol{w})\} - \|\boldsymbol{w}\|_0$$

Signal-to-noise ratio matrix

Dependent (Gaussian) observations

- Express $\boldsymbol{\Sigma}_0 = a_0\boldsymbol{I} + \boldsymbol{S}_0$ and $\boldsymbol{\Sigma}_1 = a_1\boldsymbol{I} + \boldsymbol{S}_1$

- For a fixed K, **maximizing J-divergence** is same as minimizing

$$\frac{1}{2}\mathrm{tr}\{\boldsymbol{S}_0^{-1}\left[\boldsymbol{S}_0^{-1} + a_0^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}_0^{-1}\boldsymbol{\Sigma}_1\}$$
$$+\frac{1}{2}\mathrm{tr}\{\boldsymbol{S}_1^{-1}\left[\boldsymbol{S}_1^{-1} + a_1^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}_1^{-1}\boldsymbol{\Sigma}_0\}$$

- Introduce two variables

$$\boldsymbol{Z}_0 = \boldsymbol{\Sigma}_1^{T/2}\boldsymbol{S}_0^{-1}\left[\boldsymbol{S}_0^{-1} + a_0^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}_0^{-1}\boldsymbol{\Sigma}_1^{1/2};$$
$$\boldsymbol{Z}_1 = \boldsymbol{\Sigma}_0^{T/2}\boldsymbol{S}_1^{-1}\left[\boldsymbol{S}_1^{-1} + a_1^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}_1^{-1}\boldsymbol{\Sigma}_0^{1/2},$$

Convex solver

- SDP based on ℓ_1-norm heuristics

$$\arg \min_{\boldsymbol{w}, \boldsymbol{Z}_0, \boldsymbol{Z}_1} \frac{1}{2}\mathrm{tr}\{\boldsymbol{Z}_0\} + \frac{1}{2}\mathrm{tr}\{\boldsymbol{Z}_1\}$$

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

LMIs

$$\boldsymbol{\Sigma}_1^{T/2} \boldsymbol{S}_0^{-1} \left[\boldsymbol{S}_0^{-1} + a_0^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1} \boldsymbol{S}_0^{-1} \boldsymbol{\Sigma}_1^{1/2} \preceq \boldsymbol{Z}_0$$

$$\boldsymbol{\Sigma}_0^{T/2} \boldsymbol{S}_1^{-1} \left[\boldsymbol{S}_1^{-1} + a_1^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1} \boldsymbol{S}_1^{-1} \boldsymbol{\Sigma}_0^{1/2} \preceq \boldsymbol{Z}_1$$

$$0 \leq w_m \leq 1, m = 1, 2, \ldots, M.$$

$$\begin{bmatrix} \boldsymbol{Z}_0 & \boldsymbol{S}_0^{-1}\boldsymbol{\Sigma}_1^{1/2} \\ \boldsymbol{\Sigma}_1^{T/2}\boldsymbol{S}_0^{-1} & \boldsymbol{S}_0^{-1} + a_0^{-1}\mathrm{diag}(\boldsymbol{w}) \end{bmatrix} \succeq \mathbf{0}, \qquad \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{S}_1^{-1}\boldsymbol{\Sigma}_0^{1/2} \\ \boldsymbol{\Sigma}_0^{T/2}\boldsymbol{S}_1^{-1} & \boldsymbol{S}_1^{-1} + a_1^{-1}\mathrm{diag}(\boldsymbol{w}) \end{bmatrix} \succeq \mathbf{0}.$$

Sparse sensing for detection

Setting	Neyman-Pearson	Bayesian
Optimization criterion	Kullback-Leibler distance or J-divergence	Bhattacharyya distance
Independent observations	**Ordering** distances	**Ordering** distances
Dependent Gaussian observations (uncommon means)	Signal-to-noise (**convex**) optimization	Signal-to-noise (**convex**) optimization
Dependent Gaussian observations (uncommon covariances)	Nonconvex (Kullback-Leibler distance) or **convex** (**J-divergence**) optimization	Nonconvex (Bhattacharyya distance) or **convex** (**J-divergence**) optimization

Off-the-grid Sparse Sensing

S.P. Chepuri and G. Leus. **Continuous Sensor Placement**. IEEE Signal Processing Letters, 22(5): 544-548, May 2015.

Rough gridding

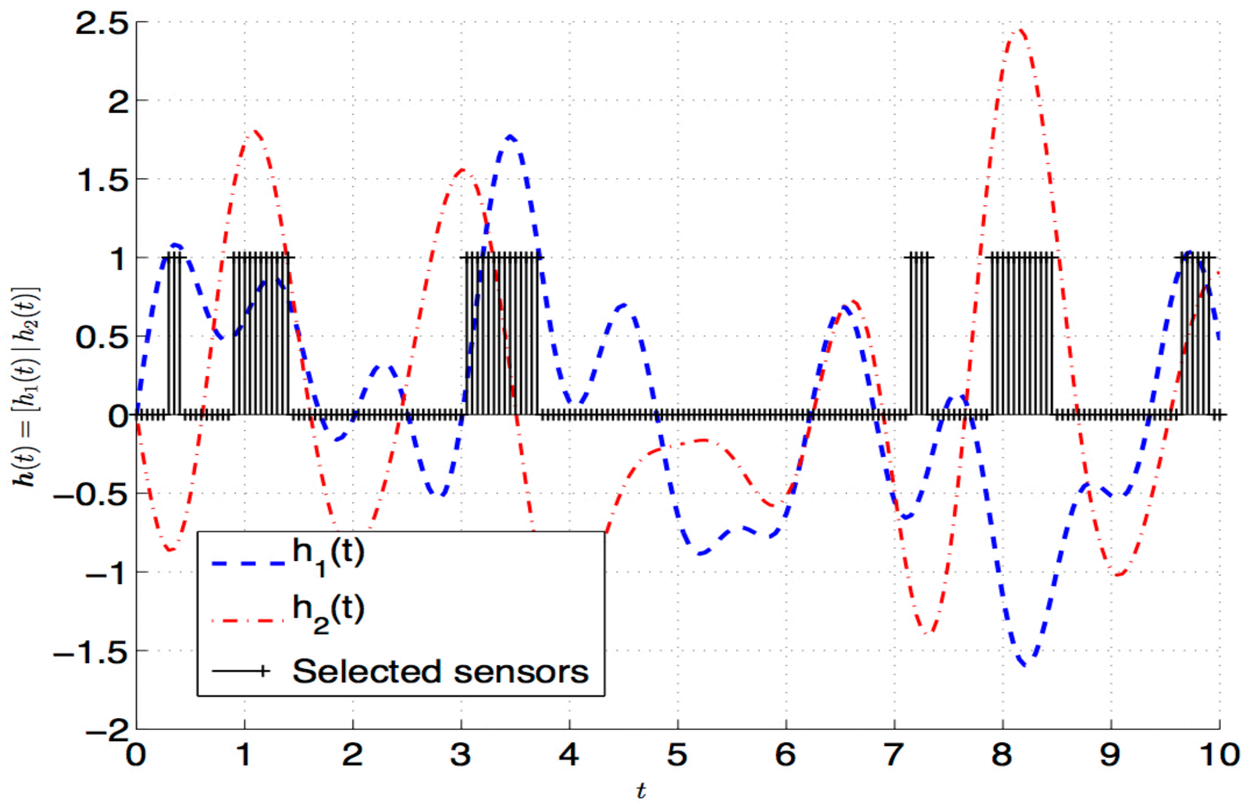
- So far, the focus was on discrete sparse sensing
  - start with a discrete set of candidates to pick the best ones
- Rough grid for complexity savings
  - candidate set is too small and/or resolution is too coarse
  - desired performance might not be achieved

Fine gridding

- Suppose

$$y(t) = w(t)[\boldsymbol{h}^H(t)\boldsymbol{\theta} + n(t)]$$

- How about fine (or dense) gridding?



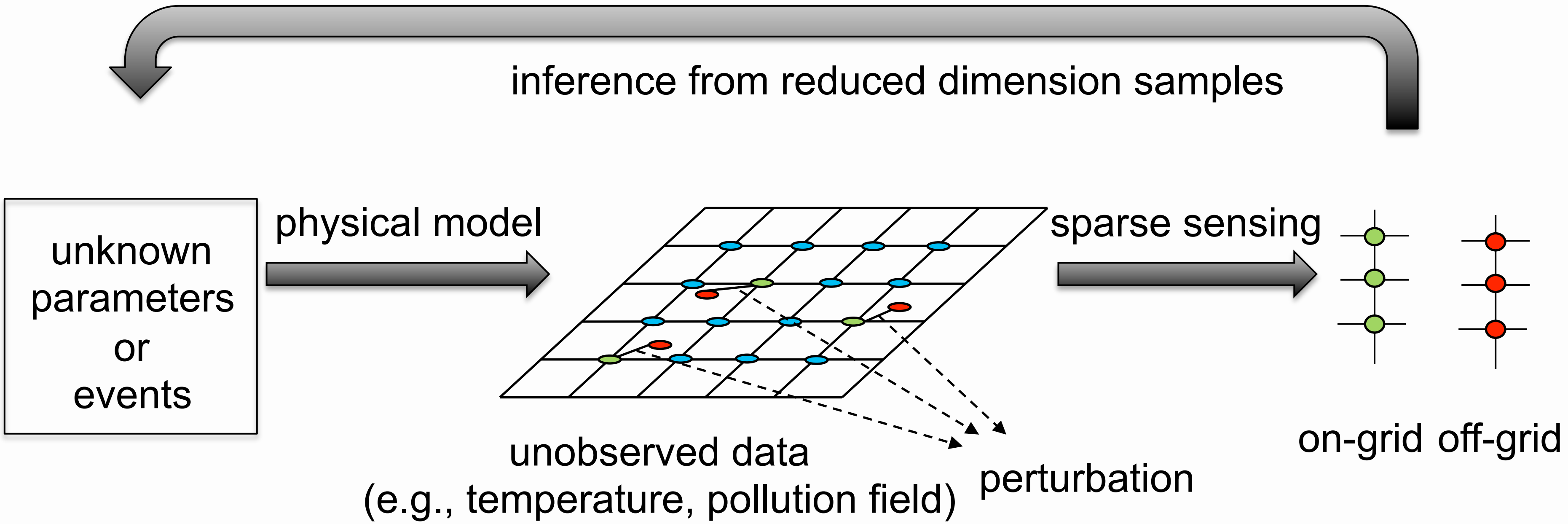
Off-the-grid sensor placement

➢ Off-the-grid sampling point = on-grid point + perturbation

$$\boldsymbol{y} = \mathrm{diag_r}(\boldsymbol{w})(\boldsymbol{x} + \mathrm{diag}(\boldsymbol{x}')\boldsymbol{p})$$

$\boldsymbol{x}'$ derivative of $x(t)$ towards t

$\boldsymbol{p}$ perturbation of the grid points



Off-the-grid sensor placement

- Off-the-grid sampling point = on-grid point + perturbation

$$\boldsymbol{y} = \mathrm{diag}_{\mathrm{r}}(\boldsymbol{w})(\boldsymbol{x} + \mathrm{diag}(\boldsymbol{x}')\boldsymbol{p})$$

$\boldsymbol{x}'$ derivative of $x(t)$ towards t
$\boldsymbol{p}$ perturbation of the grid points

- Similar to sparse total-least-squares, continuous basis pursuit
  - ✓ off-the-grid on the input space

[Zhu-Leus-Giannakis-2011], [Ekanadham-Tranchina-Simoncelli-2011]

- **Sensor placement**: off-the-grid on the **output** space

[Chepuri-Leus-2015]

- H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.
- C. Ekanadham, D. Tranchina, and E. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.
- S.P. Chepuri and G. Leus, "Continuous Sensor Placement," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 544-548, May 2015.

Off-the-grid sensor placement

- Off-the-grid sampling point = on-grid point + perturbation

$$\boldsymbol{y} = \mathrm{diag}_{\mathrm{r}}(\boldsymbol{w})(\boldsymbol{x} + \mathrm{diag}(\boldsymbol{x}')\boldsymbol{p})$$

$\boldsymbol{x}'$ derivative of $x(t)$ towards t
$\boldsymbol{p}$ perturbation of the grid points

- For $y(t) = w(t)[\boldsymbol{h}^H(t)\boldsymbol{\theta} + n(t)]$ off-the-grid sample would be

$$\begin{aligned} y_m &= w_m(\boldsymbol{h}_m^H + p_m \boldsymbol{h}_m'^H)\boldsymbol{\theta} + w_m n_m \\ &= (w_m \boldsymbol{h}_m + v_m \boldsymbol{h}_m')^H \boldsymbol{\theta} + w_m n_m \end{aligned}$$

$$v_m := w_m p_m$$

Least squares estimate

- Mean-squared error of the least-squares estimate

$$f(\boldsymbol{w}, \boldsymbol{v}) = \sigma^2 \mathrm{tr}\left\{ \left(\sum_{m=1}^{M} w_m \boldsymbol{h}_m \boldsymbol{h}_m^H + v_m^2 \boldsymbol{h}'_m \boldsymbol{h}'^H_m \right.\right.$$
$$\left.\left. + v_m (\boldsymbol{h}'_m \boldsymbol{h}_m^H + \boldsymbol{h}_m \boldsymbol{h}'^H_m) \right)^{-1} \right\}.$$

- Optimize jointly w.r.t. $\boldsymbol{w}, \boldsymbol{v}$
  - ✓ We are interested in v_m only when w_m is not zero
  - ✓ Non-linear models: FIM

Sensing design problem

$$\arg \min_{\boldsymbol{Z}=[\boldsymbol{w},\boldsymbol{v}]} \quad \|\boldsymbol{Z}\|_{0,2}$$

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

$$w_m \in \{0,1\}, m = 1,2,\ldots,M,$$

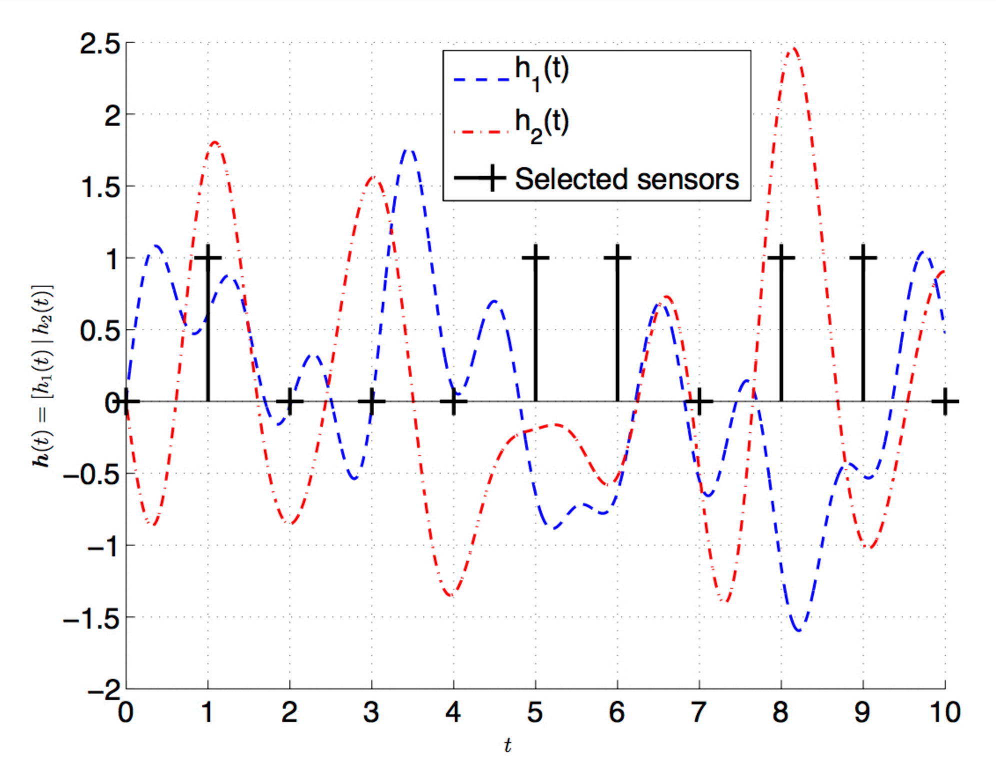
$$v_m \in [-r,r], m = 1,2,\ldots,M.$$

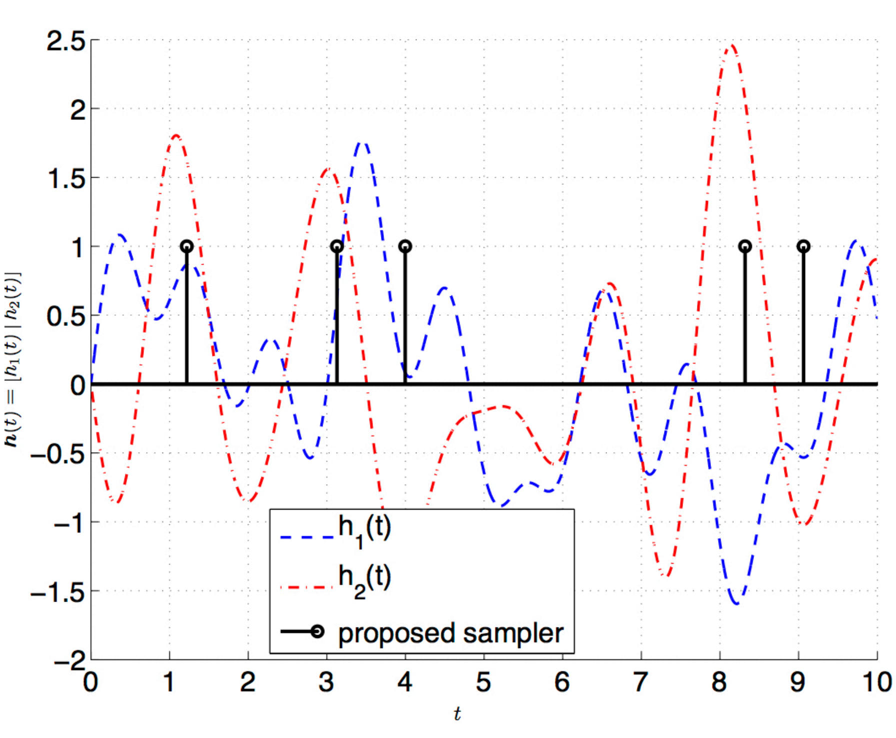
Bin size (replace grid by a bin)

$\|\boldsymbol{Z}\|_{0,2}$: # non-zero rows of $\boldsymbol{Z}$

Example

On-grid points $\{t_m = 1, 2, 3, \ldots, 11\}$





Discrete sparse sensing

$$\mathtt{mse}(\boldsymbol{\theta}) \approx 0.47$$

Off-the-grid sparse sensing

$$\mathtt{mse}(\boldsymbol{\theta}) \approx 0.36$$

Model Mismatch

1. G. Kail, S.P. Chepuri, and G. Leus. **Robust Censoring Using Metropolis-Hastings Sampling**. IEEE Journ. of Selec. Topics in Signal Processing, 10(2): 270-283, Mar. 2016.

2. S.P. Chepuri, Y. Zhang, G. Leus, and G.B. Giannakis. **Big Data Sketching with Model Mismatch**. In proc. of Asilomar Conf. Signals, systems, and Computers (Asilomar 2015), Pacific Grove, California, USA, November 2015.

Robust sparse sensing

- Till now, we assume model information is perfectly known
- What if the model parameters are partially known?
  - ✓ Regressors, noise variance, correlation matrices
- Or simply, if the data doesn't follow the postulated model
  - ✓ Outliers

Robust sparse sensing (model driven)

Example: *Linear additive Gaussian model*

$$x_m = \bar{\boldsymbol{h}}_m^T \boldsymbol{\theta} + n_m, \; m = 1, 2, \ldots, M \quad \text{and} \quad n_m = \mathcal{N}(0, \sigma^2)$$

Regressors are known up to a bounded uncertainty

$$\bar{\boldsymbol{h}}_m = \underbrace{\boldsymbol{h}_m}_{\text{known}} + \underbrace{\boldsymbol{\alpha}_m}_{\text{unknown}, \|\boldsymbol{\alpha}_m\|_2 \leq \eta}$$

- Minimize worst case **ensemble** error volume (D-optimality)

$$\max_{\boldsymbol{w} \in \{0,1\}^M} \; \min_{\|\boldsymbol{\alpha}_m\|_2 \leq \eta} \ln\det\left\{\sum_{m=1}^{M} w_m (\boldsymbol{h}_m + \boldsymbol{\alpha}_m)(\boldsymbol{h}_m + \boldsymbol{\alpha}_m)^T\right\}$$

$$\text{s.to} \quad \|\boldsymbol{w}\|_0 = K$$

Solver: convex relaxation using S-procedure

[Joshi-Boyd-2009]

- S. Joshi and S. Boyd, “Sensor selection via convex optimization,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009

Robust sparse sensing (data driven): Censoring or Sketching

Example: *Linear additive Gaussian model*

$$x_m = \bar{\boldsymbol{h}}_m^T \boldsymbol{\theta} + n_m, \; m = 1, 2, \ldots, M \quad \text{and} \quad n_m = \mathcal{N}(0, \sigma^2)$$

Regressors are known up to a bounded uncertainty

$$\bar{\boldsymbol{h}}_m = \underbrace{\boldsymbol{h}_m}_{\text{known}} + \underbrace{\boldsymbol{\alpha}_m}_{\text{unknown}, \|\boldsymbol{\alpha}_m\|_2 \le \eta}$$

➢ Minimize worst case **residual**

$$\min_{\boldsymbol{w} \in \{0,1\}^M, \boldsymbol{\theta}} \; \max_{\|\boldsymbol{\alpha}_m\|_2 \le \eta} \sum_{m=1}^{M} w_m \left(x_m - (\boldsymbol{h}_m + \boldsymbol{\alpha}_m)^T \boldsymbol{\theta}\right)^2$$

$$\text{s.to} \quad \|\boldsymbol{w}\|_0 = K$$

Two-step solver: ordering + convex/SOCP

[Chepuri-Zhang-Leus-Giannakis-2015]

- S.P. Chepuri, Y. Zhang, G. Leus, and G.B. Giannakis. Big Data Sketching with Model Mismatch. To appear in Asilomar Conf. Signals, systems, and Computers (Asilomar 2015), Pacific Grove, California, USA, November 2015

Robust sparse sensing (data driven) – outlier rejection

Example: *Linear additive Gaussian model*

We know the (**uncontaminated**) data model

$$\bar{x}_m = \boldsymbol{h}_m^T \boldsymbol{\theta} + n_m, \; m = 1, 2, \ldots, M \quad \text{and} \quad n_m = \mathcal{N}(0, \sigma^2)$$

Output data $\{x_m\}$ is possibly contaminated with up to o outliers

Given $\{x_m\}, \{\boldsymbol{h}_m\}$, and noise pdf:

a) Design $\boldsymbol{w}$ to **censor** less-informative samples and reject outliers
b) Estimate $\boldsymbol{\theta}$ using the uncensored data

Robust sparse sensing (data driven) – outlier rejection

Example: *Linear additive Gaussian model*

We know the (**uncontaminated**) data model

$$\bar{x}_m = \boldsymbol{h}_m^T \boldsymbol{\theta} + n_m,\ m = 1, 2, \ldots, M \quad \text{and} \quad n_m = \mathcal{N}(0, \sigma^2)$$

Output data $\{x_m\}$ is possibly contaminated with up to o outliers

Data samples with smaller residuals are informative

$$\min_{\mathbf{w} \in \{0,1\}^M, \boldsymbol{\theta}} \sum_{m=1}^{M} w_m \left(x_m - \mathbf{h}_m^T \boldsymbol{\theta}\right)^2 \quad \text{s.to} \quad \|\boldsymbol{w}\|_0 = K, \qquad \Leftrightarrow \qquad \min_{\mathbf{w} \in \{0,1\}^M} r(\boldsymbol{w}) \quad \text{s.to} \quad \|\boldsymbol{w}\|_0 = K,$$

$$r(\mathbf{w}) = \boldsymbol{x}_w^{\mathrm{T}} \left(\mathbf{I} - \boldsymbol{H}_w \left(\boldsymbol{H}_w^T \boldsymbol{H}_w\right)^{-1} \boldsymbol{H}_w^T\right) \boldsymbol{x}_w,\ \boldsymbol{H}_w = \boldsymbol{\Phi} \boldsymbol{H}; \boldsymbol{x}_w = \boldsymbol{\Phi} \boldsymbol{x}.$$

[Kail-Chepuri-Leus-2016]

- G. Kail, S.P. Chepuri, and G. Leus. "Robust Censoring Using Metropolis-Hastings Sampling," *IEEE Journ. of Selec. Topics in Signal Processing*, vol. 10, no. 2, pp. 270-283, Mar. 2016.

Robust sparse sensing (data driven) – outlier rejection

$$\min_{\mathbf{w}\in\{0,1\}^M,\boldsymbol{\theta}} \sum_{m=1}^{M} w_m \left(x_m - \mathbf{h}_m^T \boldsymbol{\theta}\right)^2 \quad \text{s.to} \quad \|\boldsymbol{w}\|_0 = K, \qquad \Leftrightarrow \qquad \min_{\mathbf{w}\in\{0,1\}^M} \; r(\boldsymbol{w}) \quad \text{s.to} \quad \|\boldsymbol{w}\|_0 = K,$$

$$r(\mathbf{w}) = \boldsymbol{x}_w^{\mathrm{T}}\left(\mathbf{I} - \boldsymbol{H}_w\left(\boldsymbol{H}_w^T\boldsymbol{H}_w\right)^{-1}\boldsymbol{H}_w^T\right)\boldsymbol{x}_w, \; \boldsymbol{H}_w = \boldsymbol{\Phi}\boldsymbol{H}; \boldsymbol{x}_w = \boldsymbol{\Phi}\boldsymbol{x}.$$

- Also known as least trimmed squares
- This problem is non convex in general
  - ✓ Can be convexified using to sparsity based outlier rejection *[Fuchs-1999]*
  - ✓ Markov chain Monte Carlo methods (e.g., Metropolis-Hastings sampling) *[Kail-Chepuri-Leus-2016]*

- J.-J. Fuchs, “An inverse problem approach to robust regression,” *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Phoenix, AZ, USA, Mar. 1999, pp. 1809–1812.
- G. Kail, S.P. Chepuri, and G. Leus. “Robust Censoring Using Metropolis-Hastings Sampling,” *IEEE Journ. of Selec. Topics in Signal Processing*, vol. 10, no. 2, pp. 270-283, Mar. 2016.

Hybrid model-data-driven

- Data-driven samplers are robust to outliers, but are not MSE optimal
- Model-driven samplers are MSE optimal, but are not robust to outliers
- Hybrid model-data-driven designs

  + Designs are robust to outliers and are MSE optimal

  - Samplers need to be designed for each data realization

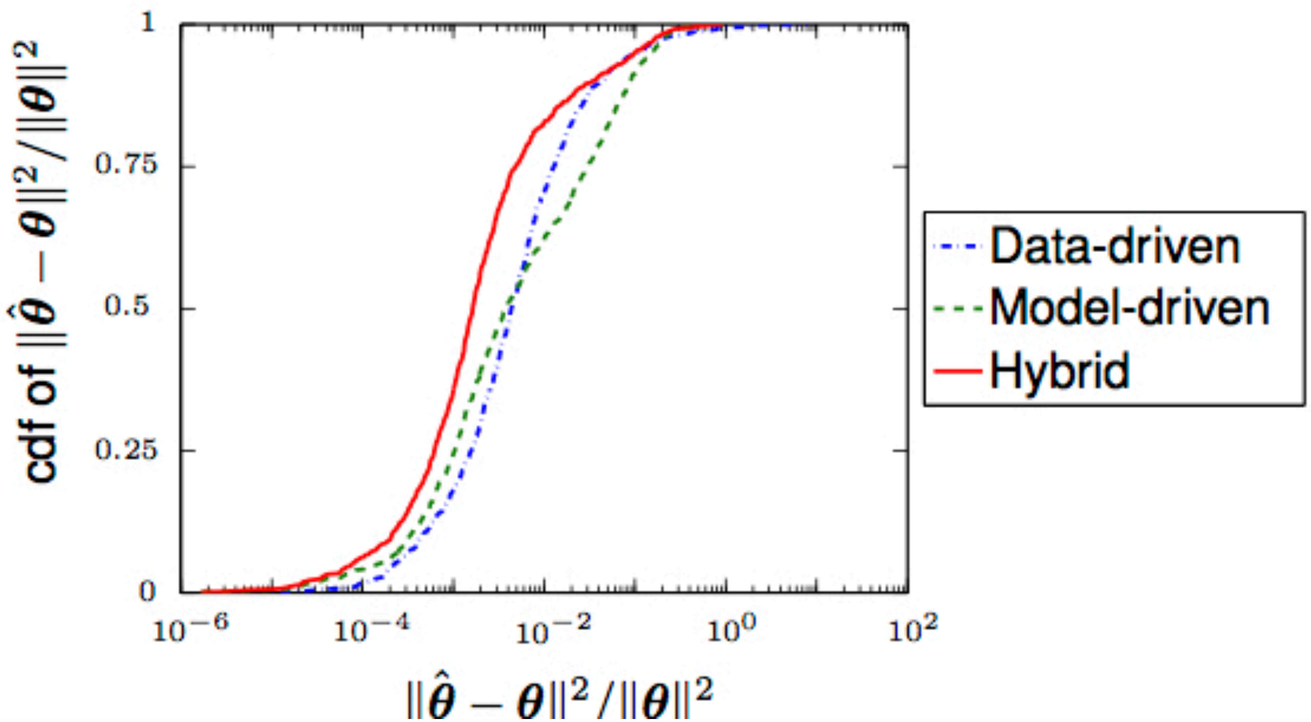
Hybrid model-data-driven

- Jointly optimizes the likelihood function (i.e., residual) and the mean squared error

$$\min_{\mathbf{w} \in \{0,1\}^M} \quad r(\mathbf{w}) + \lambda f(\mathbf{w}) \qquad \text{s.to} \|\boldsymbol{w}\| = K \quad \Leftrightarrow \quad \begin{aligned} &\min_{\mathbf{w} \in \{0,1\}^M, t_1, t_2} \quad t_1 + \lambda t_2 \\ &\text{s.t.} \quad r(\mathbf{w}) \le t_1, f(\mathbf{w}) \le t_2, \\ &\quad \|\boldsymbol{w}\| = K. \end{aligned}$$

- $\lambda \rightarrow 0(\infty)$ results in the related data (model)-driven scheme

Hybrid model-data-driven



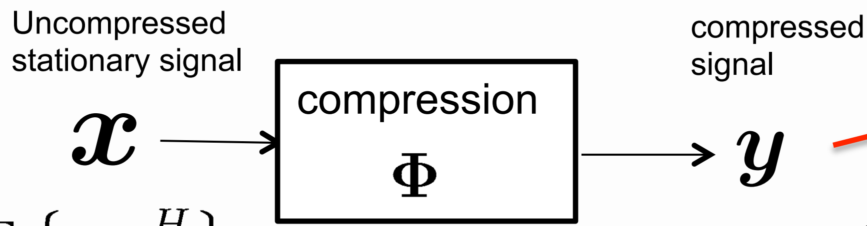
Best performance of the **hybrid** scheme is **not** necessarily **in between** the best performances of the data-driven and model-driven schemes.

Sparse Covariance Sensing

1. D. Romero, D.D. Ariananda, Z. Tian, and G. Leus. **Compressive covariance sensing: Structure-based compressive sensing beyond sparsity**. IEEE Signal Processing Magazine, vol. 33, no. 1, pp.78-93, Jan. 2016.

2. S.P. Chepuri and G. Leus. **Subsampling for Graph Power Spectrum Estimation**. In proc. Of the Ninth IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2016), Rio de Janeiro, Brazil, July 2016.

Compressive covariance sensing

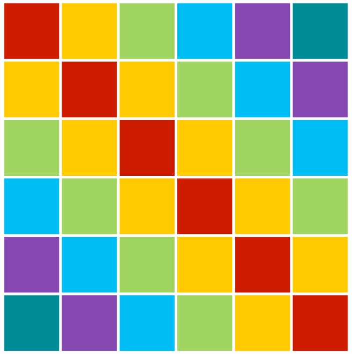


$K \ll M$

$\boldsymbol{R}_x = \mathrm{E}\left\{\boldsymbol{x}\boldsymbol{x}^H\right\}$ $\boldsymbol{R}_y = \mathrm{E}\left\{\boldsymbol{y}\boldsymbol{y}^H\right\}$

$K \times M$

Toeplitz



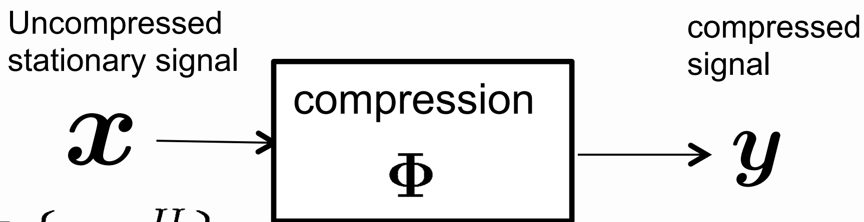
$K^2 \times 1$ $M^2 \times 1$

$$\boldsymbol{r}_y = \mathrm{vec}(\boldsymbol{R}_y) = \mathrm{vec}(\boldsymbol{\Phi}\boldsymbol{R}_x\boldsymbol{\Phi}^H) = (\boldsymbol{\Phi}^* \otimes \boldsymbol{\Phi})\mathrm{vec}(\boldsymbol{R}_x)$$

$$\boldsymbol{r}_y = (\boldsymbol{\Phi}^* \otimes \boldsymbol{\Phi})\boldsymbol{T}\boldsymbol{r}_x \xrightarrow{\text{Least squares}} \boldsymbol{r}_x = [(\boldsymbol{\Phi}^* \otimes \boldsymbol{\Phi})\boldsymbol{T}]^\dagger \boldsymbol{r}_y$$

Design of $\boldsymbol{\Phi}$ crucial for the least-squares solution to be unique

Compressive covariance sensing

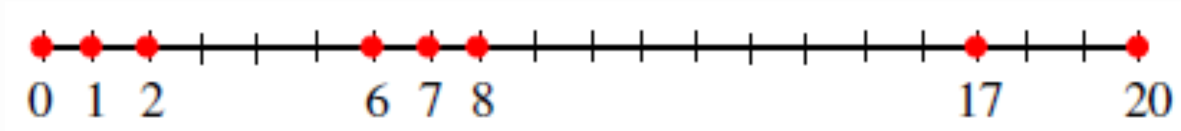


$$R_x = \mathrm{E}\{xx^H\}$$

$$R_y = \mathrm{E}\{yy^H\}$$

- Minimal sparse rulers ensure identifiability

$N = 21:$



[Romero-Ariananda-Tian-Leus-2016]

How about inference performance (mean squared error)?

- D. Romero, D.D. Ariananda, Z. Tian, and G. Leus. “Compressive covariance sensing: Structure-based compressive sensing beyond sparsity,” IEEE Signal Processing Magazine, vol. 33, no. 1, pp.78-93, Jan. 2016.

Sparse covariance sensing

- Quality of the least squares solution

$$\boldsymbol{r}_x = [(\boldsymbol{\Phi}^H \otimes \boldsymbol{\Phi})\boldsymbol{T}]^\dagger \boldsymbol{r}_y$$

depends on the spectrum of

$$\boldsymbol{G}(\boldsymbol{w}) = [(\boldsymbol{\Phi}^H \otimes \boldsymbol{\Phi})\boldsymbol{T}]^H[(\boldsymbol{\Phi}^H \otimes \boldsymbol{\Phi})\boldsymbol{T}] = \boldsymbol{T}^H[\mathrm{diag}(\boldsymbol{w}) \otimes \mathrm{diag}(\boldsymbol{w})]\boldsymbol{T}$$

- We balance the spectrum via (D-optimal design)

$$\arg \max_{\boldsymbol{w}\in\{0,1\}^N} \quad \log\det\{\boldsymbol{G}(\boldsymbol{w})\} \quad \mathrm{s.to} \quad \|\boldsymbol{w}\|_0 = K$$

✓ Submodular or convex relaxation

Some ongoing work

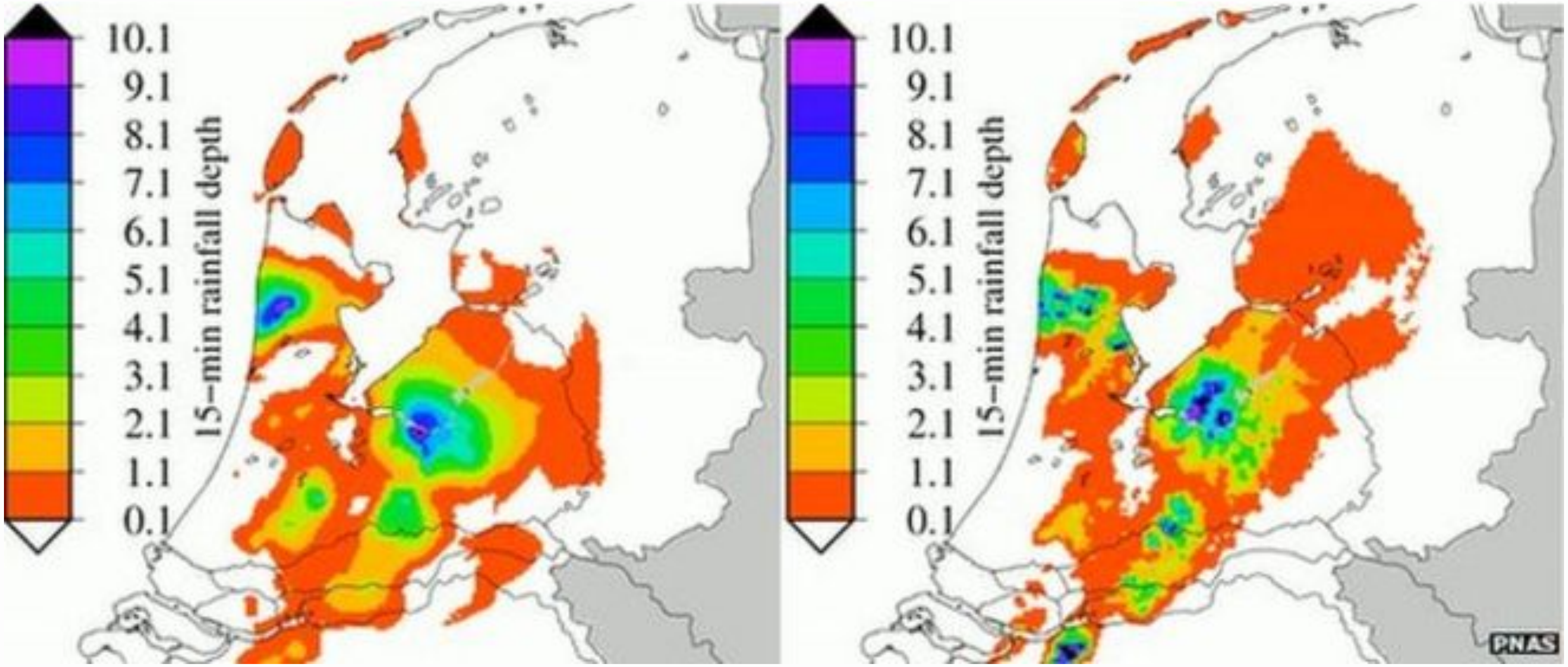
Gas leak detection (field estimation)

- Joint work between TU Delft, MIT, and KAUST
- Communications and energy constraints

Some ongoing work

Rainfall estimation

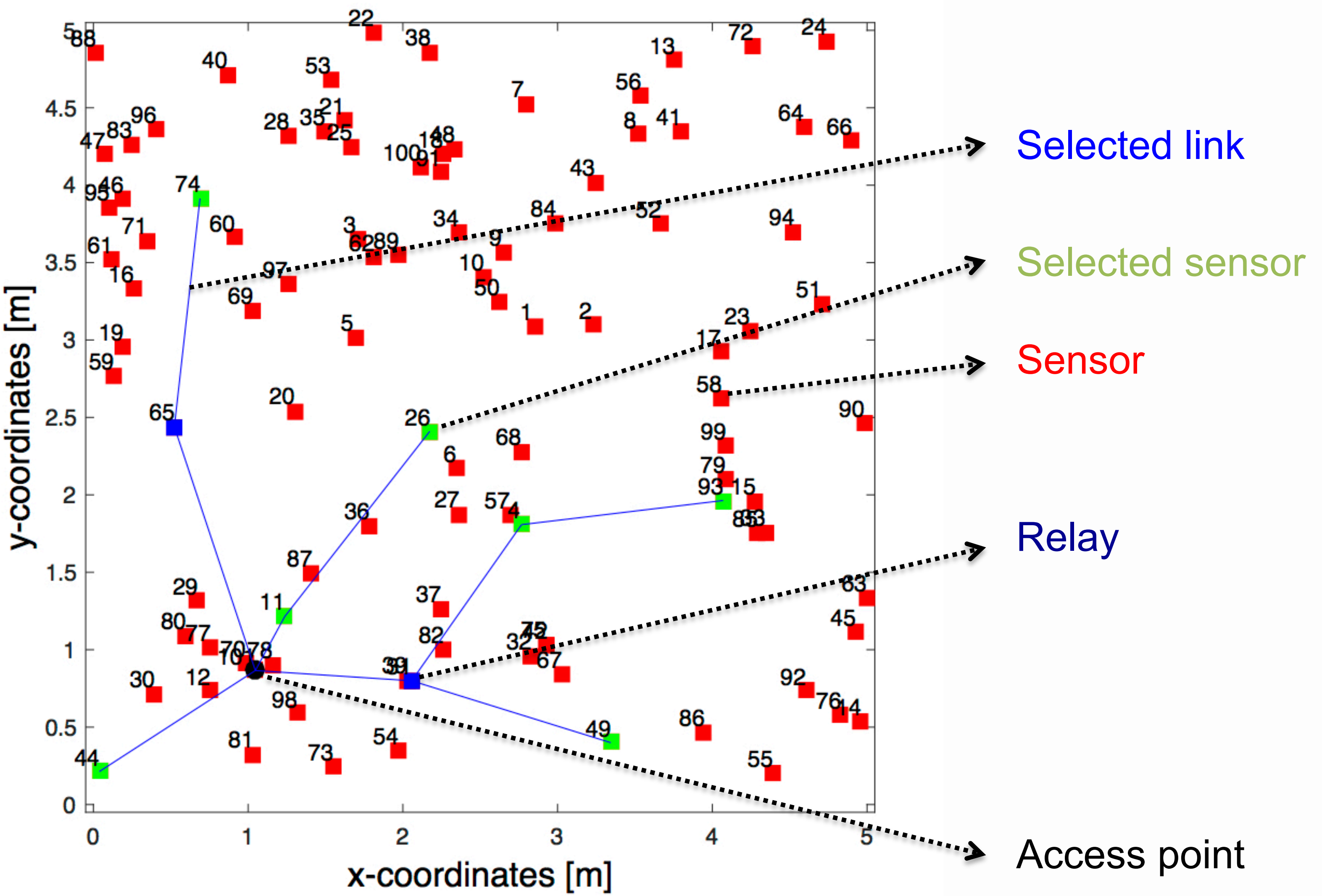
- Joint work between KNMI
- Estimation from selected microwave links



Some ongoing work

Sensor, relay, and link selection

- Towards efficient wireless sensor network design



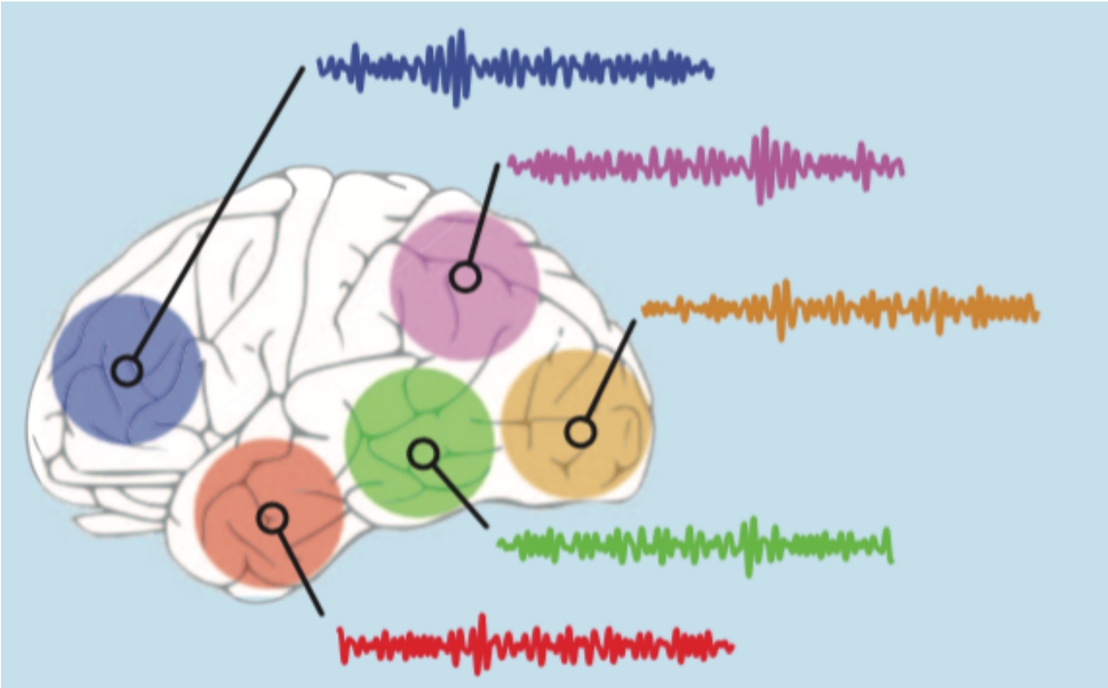
Some ongoing work

Subsampling signals on graphs

- Signal reconstruction/detection
- Power spectral density estimation

Brain networks

- fMRI time series
- EEG signals
- …



Open issues: theoretical

- Tight upper/lower bounds for **Formulations 1 and 2**
- Greedy algorithms for correlated estimation/detection
- Off-the-grid for detection, submodular off-the-grid?
- Specialized inference problems
  - Constrained estimation
  - Composite and multiple hypothesis testing

Open issues: applications

- Medical and computational imaging
- Speech and audio processing
- Big and graph data applications (scalable algorithms)
- Radar applications
- Radio astronomy
- Seismic applications

Conclusions

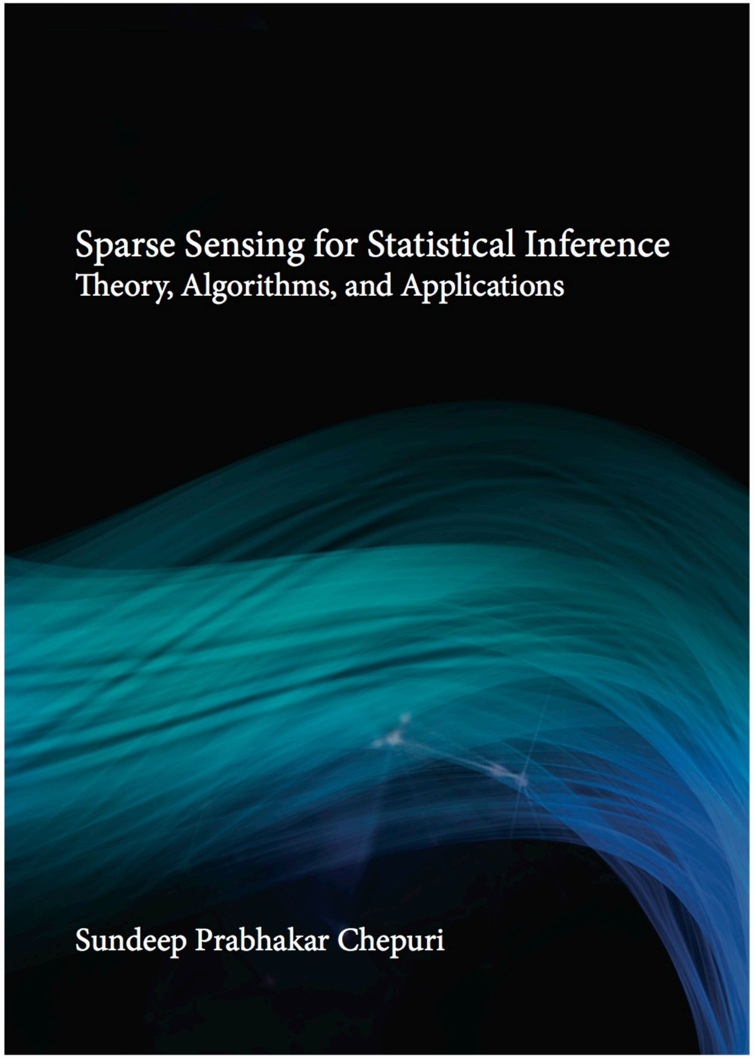
	Sparse sensing
Sparse signal	**Not needed**
Samplers	**Structured and deterministic**
Compression	**practical, controllable**
Signal processing task	**any statistical inference**

- Design space-time sparse samplers
  - Extend Nyquist-based classical sensing techniques
- Basic statistical inference problems
  - Estimation, filtering, and detection
- Related problems
  - Covariance sensing, data analytics

Reference material

PhD thesis available on **EURASIP thesis library**

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



Reference material

Foundations and Trends in Signal processing
Monograph to appear soon

For more papers on sparse sensing see: http://cas.et.tudelft.nl/~sundeep/

or

Send us an email: s.p.chepuri@tudelft.nl

Thank you!
Discussion and Q&A