Sparse Sensing for Statistical Inference

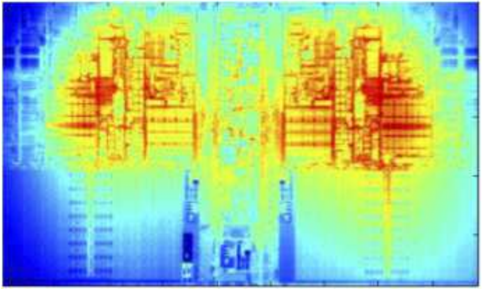
Geert Leus

Sensor Signal Processing for Defence Conference, Edinburgh
Sept. 22, 2016

Acknowledgements:
Sundeep Chepuri, Inna Ivashko, Shilpa Rao, Georg Kail, Venkat Roy, Sijia Liu, Pramod Varshney, Hadi Jamali Rad, Andrea Simonetto, Xiaoli Ma, Yu Zhang, Georgios Giannakis, Alle-Jan van der Veen.

TUDelft Delft University of Technology

How to optimally deploy sensors?



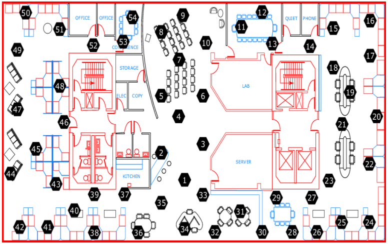
Thermal map of a processor

Example:

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

Radio astronomy (e.g., SKA)

Power networks, PMU placement



Indoor localization (e.g., museum)

Distributed radar (TU Delft campus)

Design sparse space-time samplers

Sparse sensing

- Why sparse sensing?
  - Economical constraints (hardware cost)
  - Limited physical space
  - Limited data storage space
  - Reduce communications bandwidth
  - Reduce processing overhead

Sparse sensing

What is sparse sensing?

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

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

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

Inference tasks can be estimation, filtering, and detection

Sparse sensing vs. compressed sensing

- Compressed sensing – state-of-the-art low-cost sensing scheme

	Compressed sensing	**Sparse sensing**
Sparse $x(t)$	needed	not needed
Samplers	random	structured/deterministic
Compression	robust	practical, controllable
Signal processing task	sparse signal reconstruction	any statistical inference

Discrete Sparse Sensing

Discrete sparse sensing

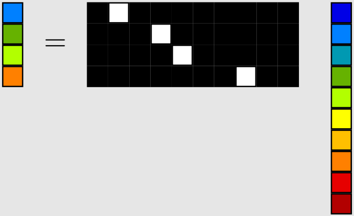
- Assume a set of candidate sampling locations $\{t_1, t_2, \ldots, t_M\}$
- Design the discrete sensing vector

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

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

Discrete sparse sensing

$$\boldsymbol{y} \qquad \boldsymbol{\Phi}(\boldsymbol{w}) = \overbrace{\mathrm{diag_r}(\boldsymbol{w})}^{\{0,1\}^{K \times M}} \qquad \boldsymbol{x}$$



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

"Design a sparsest $\boldsymbol{w}$"

$\boldsymbol{x} = [x(t_1), x(t_2), \ldots, x(t_M)]^T$

$\mathrm{diag_r}(\cdot)$ - diagonal matrix with the argument on its diagonal but with the zero rows removed.

Discrete sparse sensing or sensor selection

What is discrete sparse sensing?

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

- Classic solutions:
  - **convex optimization:** design $\{0,1\}^M$ selection vector
    [Joshi-Boyd-09]
  - greedy methods and heuristics: submodularity
    [Krause-Singh-Guestrin-08], [Ranieri-Chebira-Vetterli-14]
- **Model-driven** vs. data-driven (censoring, outlier rejection)
  [Rago-Willett-Shalom-96], [Msechu-Giannakis-12]

Design problem

Problem 1

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

Problem 2

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

$f(\boldsymbol{w})$	performance measure
λ	accuracy requirement

K	number of selected sensors

Non-convex Boolean problem

Greedy submodular maximization

- If $f(\mathbf{w})$ or $f(\mathcal{X})$ is submodular

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

$\mathcal{X}$: set of selected or not selected sensors, $\mathcal{X} \subseteq \mathcal{Y} \subset \mathcal{M}$

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

Greedy algorithm *[Krause-Singh-Guestrin-08]*

Require: $\mathcal{X} = \emptyset, L$

repeat

$$s^* = \arg\max_{s \notin \mathcal{X}} f(\mathcal{X} \cup \{s\})$$

$$\mathcal{X} \leftarrow \mathcal{X} \cup \{s^*\}$$

until $|\mathcal{X}| = L$

return $\mathcal{X}$

$L = K$ or $M - K$

- linear complexity
- near-optimal: $\sim 63\%$ [Nemhauser et al., 1978]

- 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-logs: $\sum_{m=1}^{M} \ln(w_m + \delta)$ with $\delta > 0$
  - (c.) your favorite approximation

Relaxed problem 1

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

What is <u>convex</u> $f(\boldsymbol{w})$ for estimation, filtering, and detection?

I. Estimation

- S.P. Chepuri and G. Leus. Sparsity-Promoting Sensor Selection for Non-linear Measurement Models. *IEEE Trans. on Signal Processing, Volume 63, Issue 3, pp. 684-698, February 2015*.
- S.P. Chepuri, G. Leus, and A.-J. van der Veen. Sparsity-Exploiting Anchor Placement for Localization in Sensor Networks. *EUSIPCO, September 2013*.

Non-linear inverse problem

- Unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^N$

$$y(t) = w(t)\,\overbrace{h(t;\boldsymbol{\theta}, n(t))}^{x(t)}$$

- e.g., source localization

- Candidate sampling locations $\{t_1, t_2, \ldots, t_M\}$

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

y_m	m-th spatial or temporal sensor measurement;
h_m	(in general) non-linear function;
n_m	white (additive/multiplicative) noise process.

$f(\boldsymbol{w})$ for estimation - Cramér-Rao bound

- Best subset of sensors yields the lowest error

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

$\widehat{\boldsymbol{\theta}}$ unbiased estimate of $\boldsymbol{\theta}$

- Closed-form expression for $\boldsymbol{E}$ is not always available (e.g., non-linear, non-Gaussian)

- Cramér-Rao bound (CRB) as a performance measure
  - well-suited for offline design problems
  - reveals (local) identifiability
  - improves performance of any practical algorithm
  - equal to the MSE for the additive linear Gaussian case

$f(\boldsymbol{w})$ for estimation - Cramér-Rao bound

- Assuming independent observations
  - Fisher information (FIM) is additive

- FIM is linear in w_m:

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

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

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

Select the "most informative sensors"

$f(w)$ for estimation - scalar measures

- Prominent scalar measures (related to the confidence ellipsoid):

1. *A-optimality* (average error):

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

2. *E-optimality* (worst case error):

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

3. *D-optimality* (error volume):

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

Performance measure convex in $\boldsymbol{w}$, but depends on $\boldsymbol{\theta}$

Solver

- SDP problem based on ℓ_1-norm heuristics (E-optimal design):

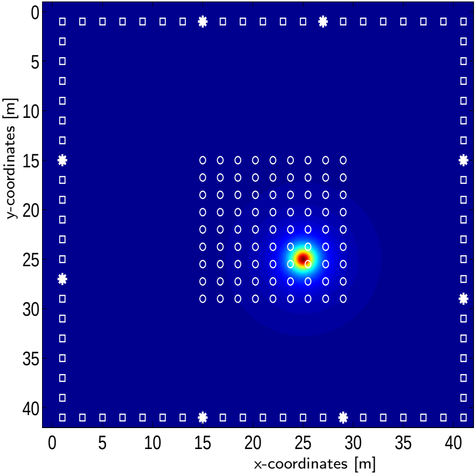
$$\begin{aligned} &\arg\min_{\boldsymbol{w}} \quad \mathbf{1}^T \boldsymbol{w} \\ &\text{s.to} \quad \sum_{m=1}^{M} w_m \boldsymbol{F}_m(\boldsymbol{\theta}) - \lambda \boldsymbol{I}_N \succeq 0, \quad \forall \boldsymbol{\theta} \in \mathcal{T}, \\ &\qquad 0 \leq w_m \leq 1, \quad m = 1, \ldots, M.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

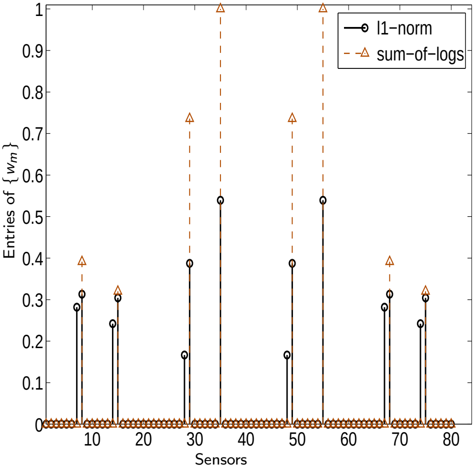
- Prior probability $p(\boldsymbol{\theta})$ is known (e.g., MMSE, MAP):

Bayesian FIM: $\boldsymbol{J}_{\mathrm{p}} + \sum_{m=1}^{M} w_m \mathbb{E}_{\boldsymbol{\theta}}\{\boldsymbol{F}_m(\boldsymbol{\theta})\} \succeq \lambda \boldsymbol{I}_N$

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

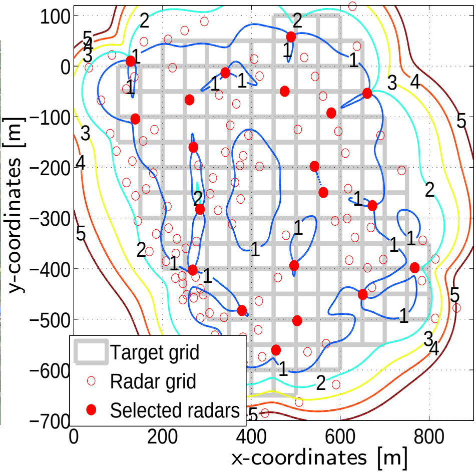
- $\boldsymbol{\theta}$ contains source location.





- Out of M = 80 available sensors (□), 8 sensors indicated by (∗) are selected. The source domain is indicated by (○).

Radar placement — TU Delft campus



- Out of $M = 117$ available radar positions, 20 radar positions are selected. *[Inna et al. 2015]*

Dependent (Gaussian) observations

- Suppose the unknown $\boldsymbol{\theta} \in \mathbb{R}^N$ follows

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

- 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 in $\boldsymbol{w}$.

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

$\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta})$ in its current form is non convex in $\mathbf{w}$

$f(\boldsymbol{w})$ for dependent (Gaussian) observations

- Express

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

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

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

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 $\boldsymbol{S} \succ \boldsymbol{0}$

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

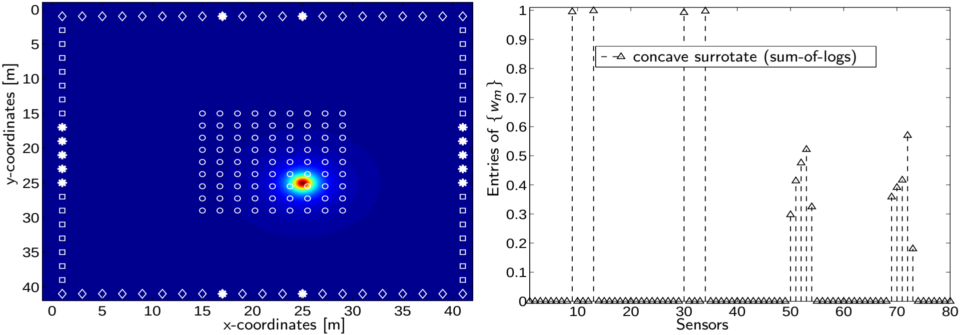
- SDP problem based on ℓ_1-norm heuristics (E-optimal design):

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

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

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

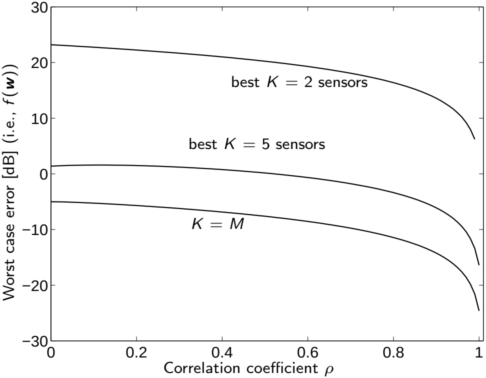
- Sensors along the horizontal edges are equicorrelated (with correlation coefficient = 0.5)
- Sensors along the vertical edges are not correlated



- Out of M = 80 available uncorrelated sensors (□) and correlated sensors (◇), 14 sensors indicated by (∗) are selected. The source domain is indicated by (○).

Is correlation good?

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



- # of sensors required (and MSE, worst case error) reduces as sensors become more coherent

II. Filtering

- S.P. Chepuri, G. Leus. *Sparsity-Promoting Adaptive Sensor Selection for Non-Linear Filtering*. ICASSP, May 2014.
- S.P. Chepuri, G. Leus. *Compression schemes for time-varying sparse signals*. ASILOMAR, November 2014.

Adaptive sparse sensing

- Some applications:

  - target tracking
  - track time-varying fields

  [Masazade-Fardad-Varshney-12], [Chepuri-Leus-14]

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

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

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

- Time-varying selection vector:

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

- Posterior-FIM can be expressed as

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

$$\boldsymbol{F}_{k,m}(\boldsymbol{\theta}_k) = \mathbb{E}\left\{ \left(\frac{\partial \ln p_{k,m}(x;\boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k} \right) \left(\frac{\partial \ln p_{k,m}(x;\boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k} \right)^T \right\} \in \mathbb{R}^{N \times N}$$

- To reduce the computational complexity, the prior Fisher can be simply evaluated at the past estimate.

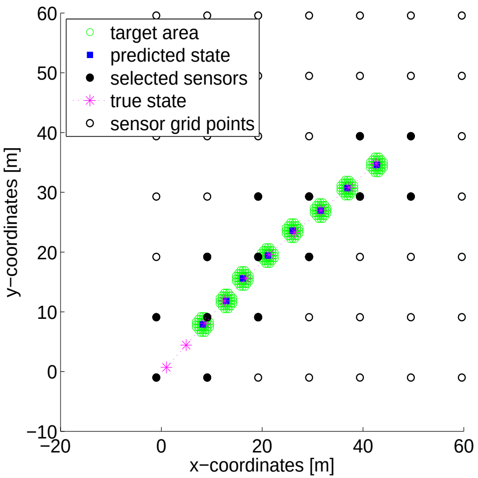
Solver

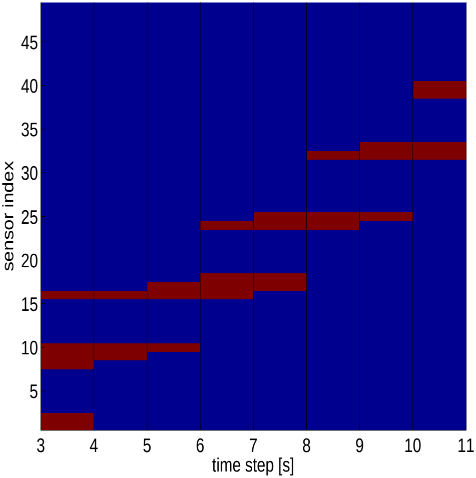
- SDP problem based on ℓ_1-norm heuristics:

$$
\begin{aligned}
\arg \min_{\boldsymbol{w}_k \in [0,1]^M} \quad & \mathbf{1}^T \boldsymbol{w}_k \\
\text{s.to} \quad & \boldsymbol{F}_{\mathrm{p},k-1} + \sum_{m=1}^{M} w_{k,m} \boldsymbol{F}_{k,m}(\boldsymbol{\theta}_k) \succeq \lambda \boldsymbol{I}_N, \forall \boldsymbol{\theta}_k \in \mathcal{T}_k \\
& 0 \leq w_m \leq 1, \quad m = 1, \ldots, M.
\end{aligned}
$$

$\mathcal{T}_k$ around the prediction

Target tracking





- $M = 49$ equally spaced sensor grid points

Structured signals: sparse, joint-sparse, smoothness,...

- Unknown sparse parameter $\boldsymbol{\theta}_k \in \mathbb{R}^N$ obeys

$$
\begin{aligned}
\text{measurements:} \quad & \boldsymbol{y}_k = \mathrm{diag_r}(\boldsymbol{w}_k)\boldsymbol{H}_k\boldsymbol{\theta}_k + \boldsymbol{n}_k \\
\text{dynamics:} \quad & \boldsymbol{\theta}_k = \boldsymbol{A}_k\boldsymbol{\theta}_{k-1} + \boldsymbol{u}_k \\
\text{pseudo-measurement:} \quad & 0 = r(\boldsymbol{\theta}_k) + e_k
\end{aligned}
$$

- $r(\boldsymbol{\theta}_k)$ enforces structure (e.g., sparsity, smoothness,...)
  [Carmi-Gurfil-Kanevsky-10], [Farahmand-Giannakis-Leus-Tian-14]

- Traditional (compressive sensing) samplers
  - Random Gaussian/Bernoulli entries

$f(w)$ for filtering with structured states

- Inverse error covariance

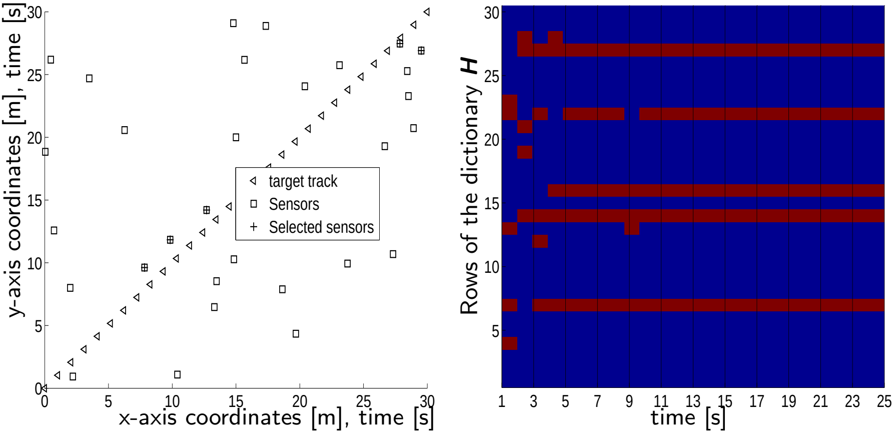
$$\boldsymbol{P}_{k|k}^{-1} = \underbrace{\boldsymbol{P}_{k|k-1}^{-1}}_{\text{dynamics}} + \underbrace{\partial r(\widehat{\boldsymbol{\theta}}_{k|k-1}) \partial r(\widehat{\boldsymbol{\theta}}_{k|k-1})^T}_{\text{sparsity prior/ pseudo-measurement}} + \underbrace{\sum_{m=1}^{M} w_{k,m} \boldsymbol{h}_{k,m}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

$\boldsymbol{h}_{k,m}$: 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 measure $f(\boldsymbol{w}_k) = \mathrm{tr}\{\boldsymbol{P}_{k|k}\}$ depends on $\boldsymbol{\theta}_k$

Target tracking: grid-based model



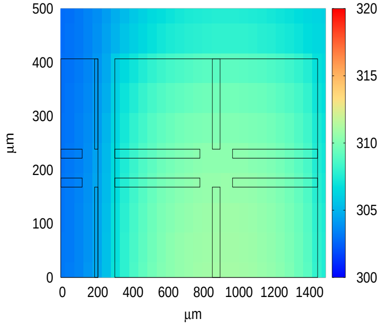
- $M = 30$ sensors; 5 sensors are selected.

III. Detection

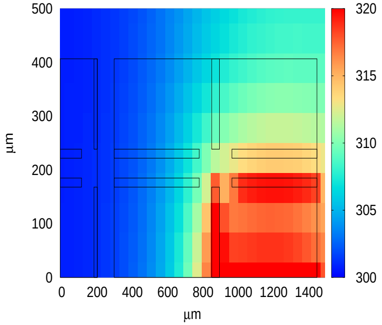
- S.P. Chepuri and G. Leus. *Sparse Sensing for Distributed Detection*. *Trans. on Signal Processing*, Oct 2015.
- S.P. Chepuri and G. Leus. *Sparse Sensing for Distributed Gaussian Detection*. ICASSP, April 2015. **(Best student paper award)**

Distributed detection

- Sensor placement for binary hypothesis testing



$\mathcal{H}_0$: No hot-spot



$\mathcal{H}_1$: Hot-spot

- Other applications
  - spectrum sensing, anomaly detection
  - radar and sonar systems

Distributed detection

- Observations are related to

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

- Binary hypothesis testing:
  - classical setting (Neyman-Pearson detector)
  - Bayesian setting

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

Sparse sensing for distributed detection

Classical setting

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

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

Bayesian setting

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

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

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

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

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.

$f(w)$ for detection

- Weaker measures can be used instead
- Kullback-Liebler distance for the classical setting
  $\rightarrow \mathcal{D}(\mathcal{H}_1 \| \mathcal{H}_0) = \mathbb{E}_{|\mathcal{H}_1}\{\log l(\boldsymbol{x})\}$
  $\rightarrow$ **upper** & lower bounds P_m for fixed P_f
- Bhattacharyya distance (a special case of Chernoff inform.) for the Bayesian setting
  $\rightarrow \mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = -\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l(\boldsymbol{x})}\}$
  $\rightarrow$ **upper** & lower bounds P_e
- These distances are suitable for offline designs

$f(w)$ for detection

- Assuming conditionally independent observations:

$$\text{(KL distance) } \mathcal{D}(\mathcal{H}_1 \| \mathcal{H}_0) = \mathbb{E}_{|\mathcal{H}_1}\{\log l(\boldsymbol{x})\} = \sum_{m=1}^{M} w_m \underbrace{\mathbb{E}_{|\mathcal{H}_1}\{\log l_m(x)\}}_{\mathcal{D}_m}$$

$$\text{(Bhattacharyya distance) } \mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = -\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l(\boldsymbol{x})}\} = -\sum_{m=1}^{M} w_m \underbrace{\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l_m(x)}\}}_{\mathcal{B}_m}$$

$l(\boldsymbol{x}) = \prod_{m=1}^{M} \frac{p_m(x|\mathcal{H}_1)}{p_m(x|\mathcal{H}_0)}$ likelihood ratio

$l_m(x) = \frac{p_m(x|\mathcal{H}_1)}{p_m(x|\mathcal{H}_0)}$ local likelihood ratio

Solver

- Linear program with explicit solution

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

Hint: sorting

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

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

Example: Gaussian detection

Suppose

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

- Kullback-Leibler and Bhattacharyya distance measures are the same up to a constant.

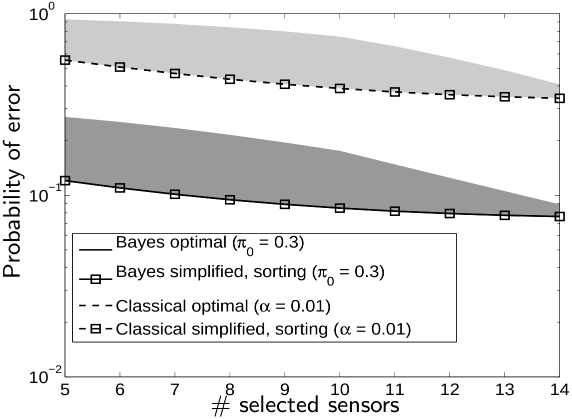
- Distance measure

$$d(\boldsymbol{w}) = \frac{1}{\sigma^2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^T \mathrm{diag}(\boldsymbol{w})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$$

  is simply the scaled signal-to-noise ratio

Example: Gaussian detection

- Sensor selection is optimal in terms of error probabilities



Dependent (Gaussian) observations

Suppose

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

- Distance measure

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

is no more linear in $\boldsymbol{w}$.

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

$f(\boldsymbol{w})$ for dependent (Gaussian) detection

- 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

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

an LMI —linear/convex in $\boldsymbol{w}$.

Choose $a > 0$ and $\boldsymbol{S} \succ \boldsymbol{0}$

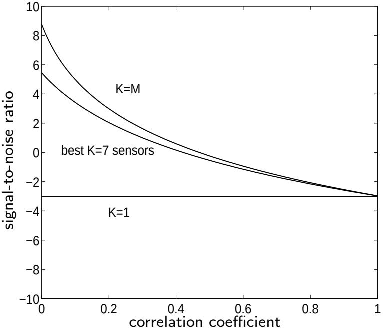
Hint: use matrix inversion lemma and $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \mathrm{diag}(\boldsymbol{w})$

- SDP problem based on ℓ_1-norm heuristics:

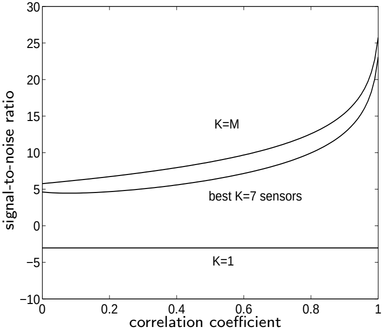
$$\begin{aligned} &\arg\min_{\boldsymbol{w}} \quad \boldsymbol{1}^T \boldsymbol{w} \\ &\text{s.to} \begin{bmatrix} \boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{m} \\ \boldsymbol{m}^T\boldsymbol{S}^{-1} & \boldsymbol{m}^T\boldsymbol{S}^{-1}\boldsymbol{m} - \lambda \end{bmatrix} \succeq \boldsymbol{0}, \\ &\qquad 0 \leq w_m \leq 1, \quad m = 1, \ldots, M. \end{aligned}$$

Is correlation good or bad?

- Equicorrelated Gaussian observations



Identical observations



Non-identical observations

- Required # of sensors reduce significantly as they become more coherent

Continuous Sparse Sensing

- S.P. Chepuri, G. Leus. *Continuous Sensor Placement*. Signal Proc. Letters, Volume 22, Issue 5, 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

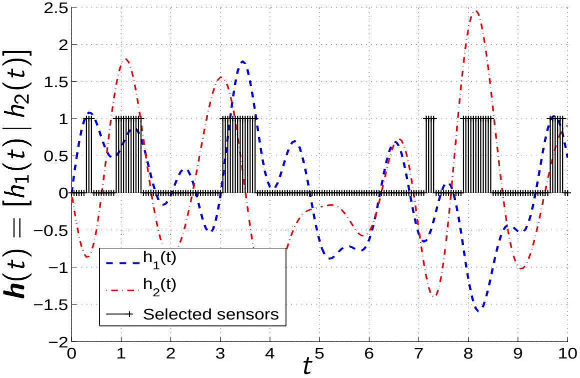
  - candidate set is too small and/or resolution is too coarse

  - desired performance might not be achieved

- Suppose

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

- How about fine gridding?



Continuous sparse sensing

- 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

- Similar to total-least-squares, continuous basis pursuit

[Zhu-Leus-Giannakis-11], [Ekanadham-Tranchina-Simoncelli-11]

- For

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

off-the-grid sample would be

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

$v_m := w_m p_m$

Continuous sparse sensing - estimation

- Mean-squared error of the least-squares estimate

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

- Joint sparse optimization problem

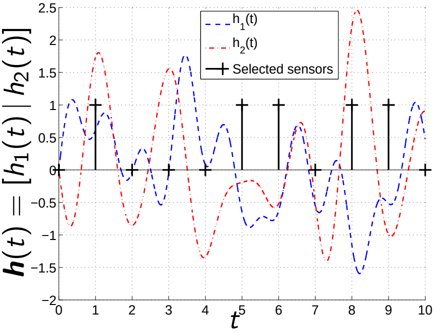
$$\begin{aligned} \arg \min_{\boldsymbol{Z}=[\boldsymbol{w},\boldsymbol{v}]} \quad & \|\boldsymbol{Z}\|_{0,2} \\ \text{s.to} \quad & f(\boldsymbol{w}, \boldsymbol{v}) \leq \lambda, \\ & w_m \in \{0, 1\}, m = 1, 2, \ldots, M, \\ & v_m \in [-r, r], m = 1, 2, \ldots, M. \end{aligned}$$

r: resolution of candidate grid

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

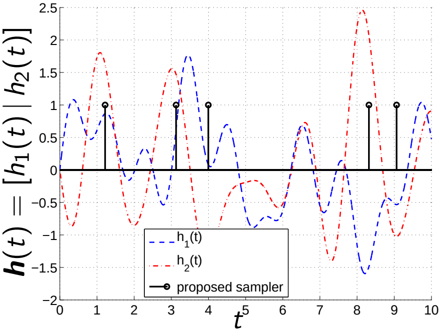
Example: linear inverse problem

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



Discrete sparse sensing

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



Continuous sparse sensing

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

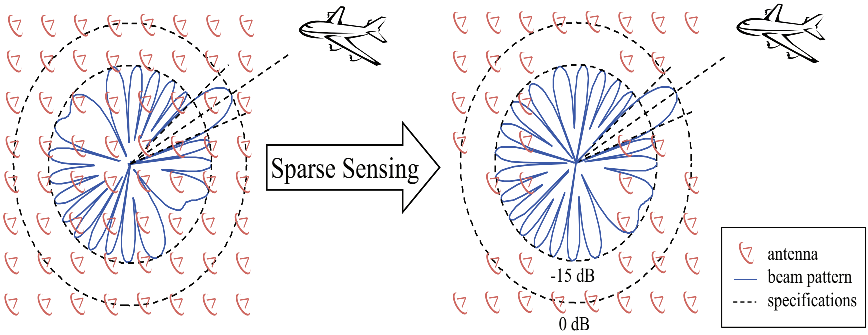
Conclusions and future works

Conclusions:

- Design space-time sparse samplers
  extend Nyquist-based classical sensing techniques
- Fundamental statistical inference problems:
  Estimation, filtering, and detection
- Applications in networks:
  environmental monitoring, location-aware services, spectrum sensing,...

Ongoing and future work:

- Data-driven sparse sensing, model mismatch.
- Continuous sparse sensing
- Clustering and classification



Thank You!!

For more on sparse sensing for statistical inference, see:
http://cas.et.tudelft.nl/∼sundeep