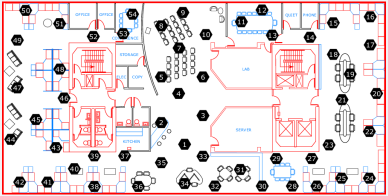
Sparse Sensing
for Estimation with Correlated Observations


Sundeep Chepuri **Geert Leus**

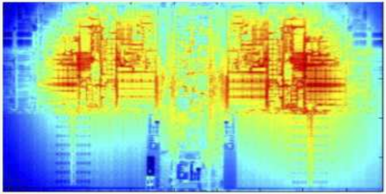
ASILOMAR 2015, Pacific Grove, USA

TU Delft
Delft University of Technology

Radio astronomy (SKA)



Indoor localization, smart buildings



Field estimation/detection

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?

Select the "best" subset of sensors out of the candidate sensors that guarantee a certain desired estimation accuracy.

Sensor selection for estimation – uncorrelated observations:

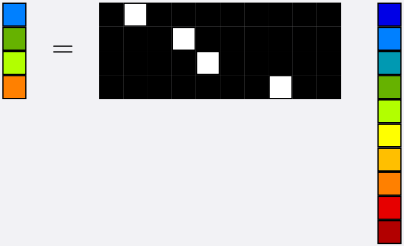
- **convex optimization:** design $\{0,1\}^M$ selection vector
  [Joshi-Boyd-09], [Chepuri-Leus-13]
- greedy methods and heuristics: submodularity
  [Krause-Singh-Guestrin-08], [Ranieri-Chebira-Vetterli-14]

Sparse sensing for estimation

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

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

$$\mathbf{y} \qquad \boldsymbol{\Phi}(\mathbf{w}) = \overbrace{\mathrm{diag_r}(\mathbf{w})}^{\{0,1\}^{K \times M}} \qquad \mathbf{x} \sim \mathcal{N}\left(\mathbf{h}(\boldsymbol{\theta}), \boldsymbol{\Sigma}\right)$$



"Design sparsest **w**"

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

Design problem

Problem 1

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

Problem 2

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

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

K number of selected sensors

Non-convex Boolean problem

Convex relaxation

- 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

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

What is <u>convex</u> $f(\mathbf{w})$ for estimation with correlated observations?

Estimation accuracy $f(\mathbf{w})$ — Cramér-Rao bound

- Best subset of sensors yields the lowest error

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

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

- Closed-form expression for $\mathbf{E}$ 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 linear case

$f(\mathbf{w})$ for estimation - scalar measures

- For Gaussian observations, Fisher information matrix

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

$\mathbf{J}(\boldsymbol{\theta}) = \partial \mathbf{h}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$; $\boldsymbol{\Sigma}(\mathbf{w}) = \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T$

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

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

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

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

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

- These performance metrics
  - in its current form are not convex on $\mathbf{w} \in [0,1]^M$
  - depend on the true parameter

Equivalent convex expression for $f(\mathbf{w})$

- Express

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

- Constraint (E-optimal design)

$$\mathbf{J}^T(\theta)\boldsymbol{\Phi}^T\left(a\mathbf{I} + \boldsymbol{\Phi}\mathbf{S}\boldsymbol{\Phi}^T\right)^{-1}\boldsymbol{\Phi}\mathbf{J}(\theta) \succeq \lambda\mathbf{I}_N$$

is equivalent to

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

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

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

Solver

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

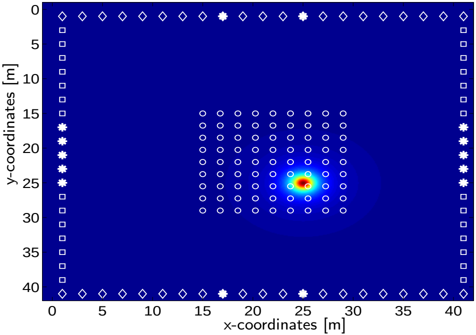
$$\arg\min_{\mathbf{w}} \quad \mathbf{1}^T\mathbf{w}$$

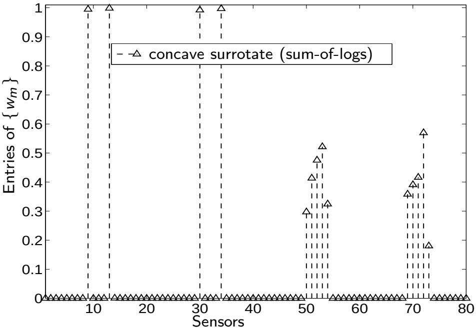
$$\text{s.to} \begin{bmatrix} \mathbf{S}^{-1} + a^{-1}\mathrm{diag}(\mathbf{w}) & \mathbf{S}^{-1}\mathbf{J}(\boldsymbol{\theta}) \\ \mathbf{J}^T(\boldsymbol{\theta})\mathbf{S}^{-1} & \mathbf{J}^T(\boldsymbol{\theta})\mathbf{S}^{-1}\mathbf{J}(\boldsymbol{\theta}) - \lambda\mathbf{I}_N \end{bmatrix} \succeq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

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

Sensor placement for source localization

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

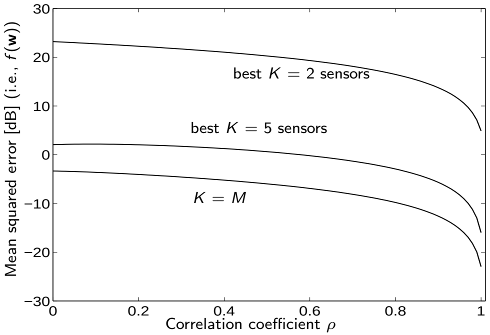




- Out of M = 80 available uncorrelated sensors (□) and correlated sensors (◇), 14 sensors indicated by (∗) are selected. The source domain is indicated by (○).

Is correlation good?

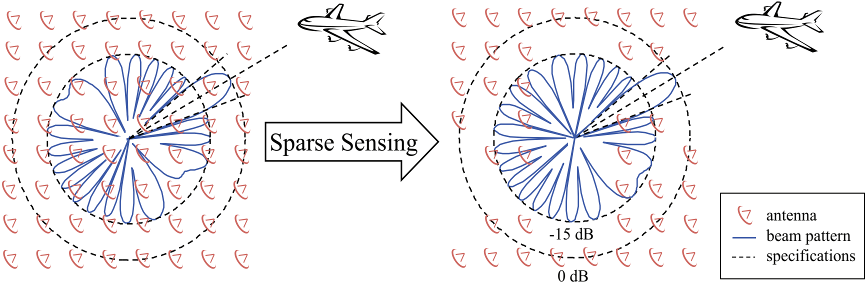
- Linear model, Gaussian regression matrix
- Equicorrelated correlation matrix: $\boldsymbol{\Sigma} = \left[(1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T\right]$



- # of sensors required (and MSE) reduces as sensors become more coherent

Conclusions

- Design space-time sparse samplers
  to reduce sensing and other related costs
- Fundamental statistical inference problems:
  Estimation, filtering, and detection
- Applications in networks:
  environmental monitoring, location-aware services, spectrum sensing,. . .



Thank You!!

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