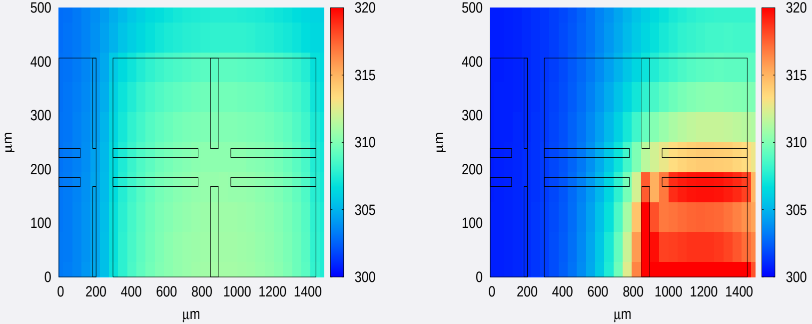
Sparse Sensing for Distributed Gaussian Detection

Sundeep Chepuri **Geert Leus**

ICASSP 2015, Brisbane, Australia

TU Delft Delft University of Technology

Thermal map of a processor



$\mathcal{H}_0$: No hot-spot $\mathcal{H}_1$: Hot-spot

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 global detection probability.

Sensor selection – prior art:

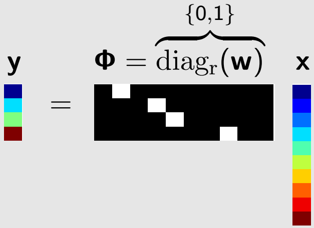
- Estimation
  - **convex optimization:** design $\{0,1\}^M$ selection vector
    [Joshi-Boyd-09], [Chepuri-Leus-13]
- Detection
  - likely to lead to a local optimum
    [Cambanis-Masry-83], [Yu-Varshney-97], [Bajovic-Sinopoli-Xavier-11]

Distributed detection

- Observations are related to

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

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



- Sensor placement
- Antenna selection
- Sample selection
- Data compression

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

Sparse sensing for distributed detection

Classical setting

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

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

Bayesian setting

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

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

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

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.

Performance measures

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

Independent observations

- Assuming conditionally independent observations

KL distance:

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

Bhattacharyya distance:

$$\mathcal{B}(\mathcal{H}_1\|\mathcal{H}_0) = -\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l(\mathbf{y})}\} = \sum_{m=1}^{M} w_m \underbrace{\left(-\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l_m(y)}\}\right)}_{\mathcal{B}_m}$$

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

Solver

- Linear program with explicit solution

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

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.

Gaussian detection

Suppose

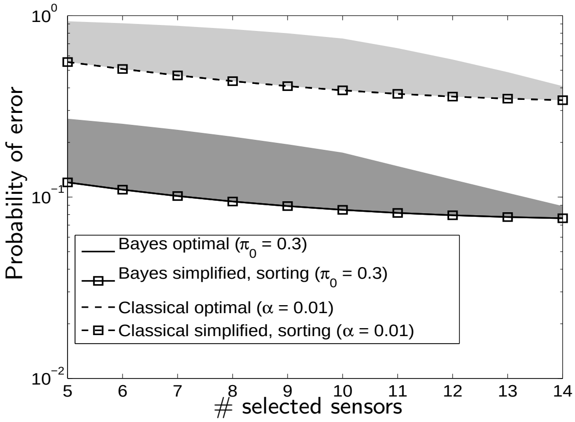
$$\mathcal{H}_0 : \quad \mathbf{x} \sim \mathcal{N}(\boldsymbol{\theta}_0, \sigma^2 \mathbf{I}) \quad \text{vs.} \quad \mathcal{H}_1 : \quad \mathbf{x} \sim \mathcal{N}(\boldsymbol{\theta}_1, \sigma^2 \mathbf{I})$$

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

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

is simply the scaled signal-to-noise ratio

- Sensor selection is optimal in terms of error probabilities



Dependent (Gaussian) observations

Suppose

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

- Distance measure

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

  is no more linear in **w**.

  $\boldsymbol{\Phi} = \mathrm{diag}_{\mathrm{r}}(\mathbf{w})$
  $\mathbf{m} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$
  $\boldsymbol{\Sigma}^{-1}(\mathbf{w}) = \left(\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T\right)^{-1}$

Dependent (Gaussian) detection

- Express

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

- Constraint

$$d(\mathbf{w}) \geq \lambda$$

is equivalent to

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

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

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

Solver – dependent case

- SDP problem based on ℓ_1-norm heuristics:

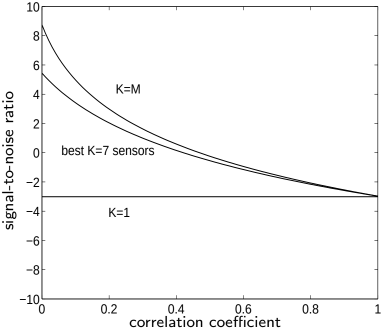
$$\arg\min_{\mathbf{w}} \quad \mathbf{1}^T\mathbf{w}$$

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

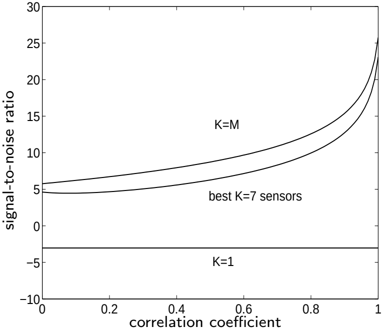
$$0 \leq w_m \leq 1, \quad m = 1, \ldots, M.$$

Is correlation good or bad?

- Equicorrelated Gaussian observations



Identical observations



Non-identical observations

Required # of sensors reduce significantly as they become more coherent

Conclusions

- Design space/time sparse samplers

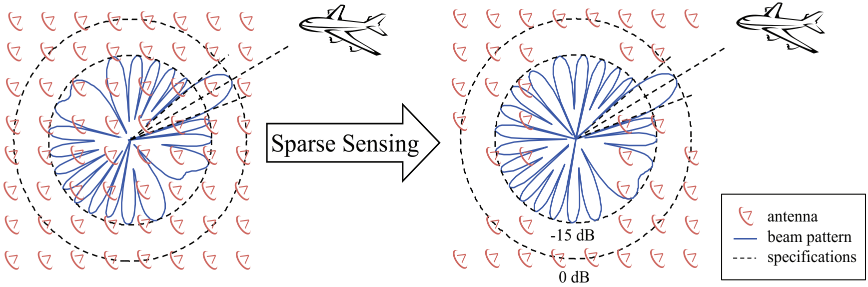
  extend Nyquist-based classical sensing techniques

- Fundamental statistical inference problems:

  Estimation, filtering, and detection

- Applications in networks:

  environmental monitoring, location-aware services, spectrum sensing,...



Thank You!!

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