Subset Selection for Kernel-based Signal Reconstruction

Mario Coutino, Sundeep P. Chepuri, Geert Leus

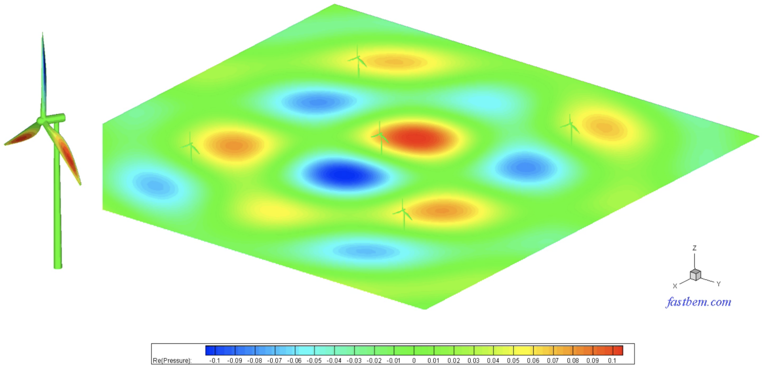
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology, Delft, The Netherlands

ICASSP 2018
Calgary, Alberta, Canada

[m.a.coutinominguez@tudelft.nl]

Motivation

Static Field Estimation

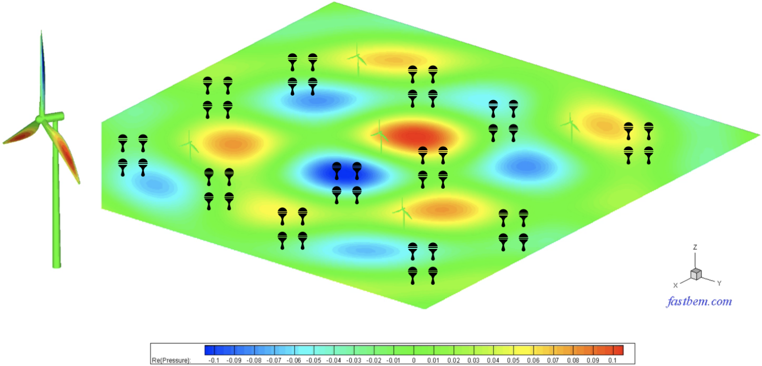


Wind turbine farm noise[1]

Open video

[1] Fast Boundary Element Method (FastBEM) for Solving Large-Scale Engineering Problems, www.fastbem.com

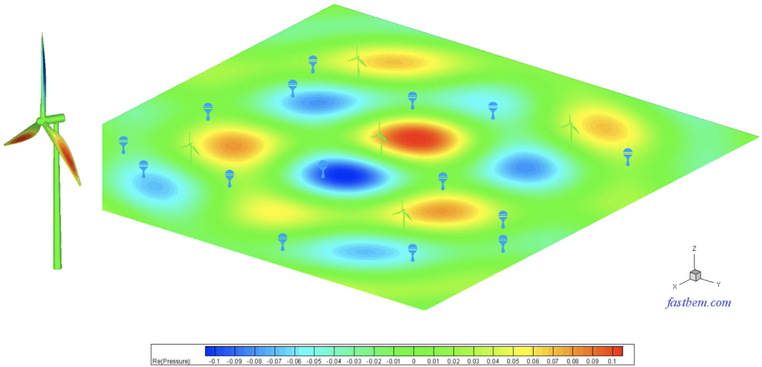
Static Field Estimation



How to select the subset of measurements to provide the best possible reconstruction performance?

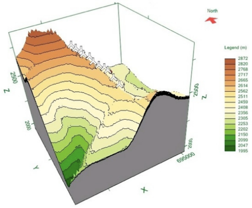
Motivation

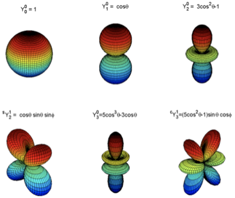
Static Field Estimation



Using information of the topographical relief of the terrain, field signal properties and network topology.

Prior information



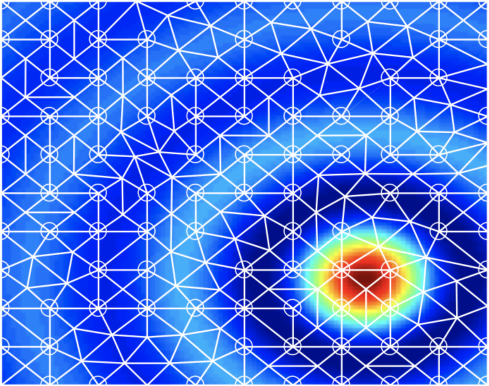


$\boldsymbol{k}(x_i, x)$: kernel function

Kernels leverage structural information to propagate non-linear relations through linear ones.

Kernel-based Signal Reconstruction

Sampling of a static field



Field function:

$f(x) : \mathcal{M} \mapsto \mathbb{R}$

Field measurements:

$\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{n}$

The static field, $f(x)$, is assumed to belong to a reproducing kernel Hilbert space (RKHS), $\mathcal{H}$, defined a kernel map $k : \mathcal{M} \times \mathcal{M} \mapsto \mathbb{R}$.

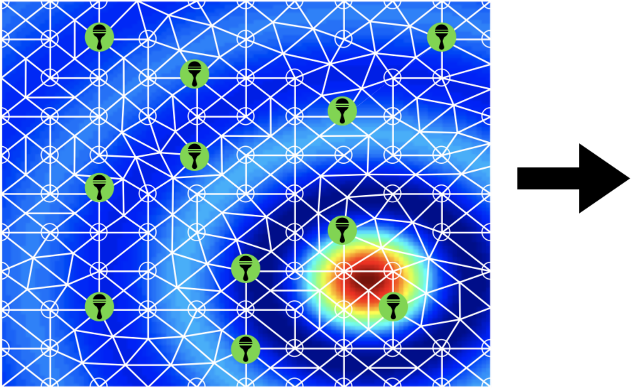
Reproducing kernel Hilber space (RKHS)

$$\mathcal{H} = \left\{ f : f(x) = \sum_{i=1}^{\infty} \alpha_i k(x_i, x),\ \alpha_i \in \mathbb{R} \right\}$$

where $k : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is a symmetric kernel map satisfying $\sup_{x,y} k(x, y) < \infty$.

Kernel-based Signal Reconstruction

Sparse Sampler Design



Field measurements:

$$\mathbf{y}_{\mathcal{S}} = \boldsymbol{\Phi}_{\mathcal{S}}\boldsymbol{f} + \boldsymbol{\Phi}_{\mathcal{S}}\boldsymbol{n} = \boldsymbol{f}_{\mathcal{S}} + \boldsymbol{n}_{\mathcal{S}}$$

Problem: *Given model statistics and a kernel map, find the best subset of sensors $\mathcal{S}$, $|\mathcal{S}| = K$, that provides the best reconstruction of $f(x)$*

Sensor Selection for Sparse Sensing

- Why?
  - possibly many non-informative measurements
  - reduces processing overhead
  - economical or physical constraints
  - sensors might incur different operation costs, e.g., energy requirements
- How?
  - **convex optimization:** through selection vector $\boldsymbol{w} \in \{0, 1\}^M$
    [Joshi-Boyd-09][2], [Chepuri-Leus-16][3]
  - **submodular optimization:** greedy methods and heuristics
    [Krause-Guestrin-07][4],[Ranieri-Chebira-Vetterli-14][5]

In this work, both approaches are considered...
and results of kernel methods leveraged...

---

2 S. Joshi, and S. Boyd. "Sensor selection via convex optimization." *TSP* 2009

3 S.P. Chepuri, and G. Leus, "Sparse Sensing for Statistical Inference," *Foundations and Trends in Sig. Proc.* 2016

4 A. Krause, and C. Guestrin, "Near-optimal observation selection using submodular functions," *AAAI* 2007

5 J. Ranieri, et al., "Near-optimal sensor placement for linear inverse problems," *TSP* 2014

Kernel-based Signal Reconstruction

Estimate of the continuous function $f(x)$:

$$\hat{f} = \underset{f \in \mathcal{H}}{\arg\min} \frac{1}{K} \sum_{x_i \in \mathcal{S}} \mathcal{L}(y(x_i), f(x_i)) + \mu \Omega(\|f\|_{\mathcal{H}}), \tag{1}$$

where $\mathcal{L}(\cdot, \cdot)$ is a loss function and $\Omega(\cdot)$ a smoothness term in $\mathcal{H}$, μ the regularization parameter and $\mathcal{S}$ is the finite sampling set.

The *representer theorem*[6] provides the solution for (1) by the following series:

$$\hat{f}(x) = \sum_{x_i \in \mathcal{S}} \alpha_i k(x_i, x). \tag{2}$$

In this talk, we focus on kernel ridge regression for estimating $f(x)$...

[6] B. Sch olkopf, et al., "A generalized representer theorem" *in Comp. Learn. Theory*, Springer, pp. 416-426, 2001.

Kernel-based Signal Reconstruction

Kernel Ridge Regression (KRR)

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \frac{1}{K} \sum_{x_i \in \mathcal{S}} (y(x_i) - f(x_i))^2 + \mu \|f\|_{\mathcal{H}}^2. \quad (3)$$

Here, $\mathcal{L}(x, y) = (x - y)^2$ and $\Omega(\cdot) = (\cdot)^2$.

Using the representer theorem solution, the following relations hold

$$\boldsymbol{f}_{\mathcal{S}} = \boldsymbol{K}_{\mathcal{S}} \boldsymbol{\alpha}, \quad \|f\|_{\mathcal{H}}^2 = \boldsymbol{\alpha}^T \boldsymbol{K}_{\mathcal{S}} \boldsymbol{\alpha}, \quad (4)$$

Here,
$\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^T \in \mathbb{R}^K$ is the vector with the expansion coefficients
$[\boldsymbol{K}_{\mathcal{S}}]_{ij} = k(x_i, x_j)$, $x_i, x_j \in \mathcal{S}$, is the (i, j)th entry of the kernel matrix.

Kernel-based Signal Reconstruction

Signal Reconstruction Optimization Problem (KRR)

$$\hat{\boldsymbol{\alpha}}_{\mathcal{S}} = \begin{array}{ll} \underset{\boldsymbol{\alpha}\in\mathbb{R}^K}{\arg\min} & \frac{1}{K}\|\boldsymbol{e}\|^2 + \mu\boldsymbol{\alpha}^T\boldsymbol{K}_{\mathcal{S}}\boldsymbol{\alpha} \\ \text{subject to} & \boldsymbol{e} = \boldsymbol{y}_{\mathcal{S}} - \boldsymbol{K}_{\mathcal{S}}\boldsymbol{\alpha} \end{array} \tag{5}$$

Optimal solution

$$\hat{\boldsymbol{\alpha}}_{\mathcal{S}} = [\boldsymbol{K}_{\mathcal{S}} + \gamma \boldsymbol{I}_K]^{-1}\boldsymbol{y}_{\mathcal{S}}.$$

Residuals

$$e(x_j, \mathcal{S}) = y(x_j) - \boldsymbol{k}_{\mathcal{S},j}^T\hat{\boldsymbol{\alpha}}_{\mathcal{S}},$$

where $[\boldsymbol{k}_{\mathcal{S},j}]_i = k(x_i, x_j),\ x_i \in \mathcal{S}$, $\gamma = \mu K$, and $\boldsymbol{I}_K$ is the $K \times K$ identity matrix.

How to use this expression for designing sparse samplers?

Design Problem

Cost function: Mean Squared error

$$\mathrm{MSE}_{\mathcal{M}}(\mathcal{S}) = \int_{x\in\mathcal{M}} |e(x,\mathcal{S})|^2 dx \approx \sum_{x_j\in\mathcal{V}} |e(x_j,\mathcal{S})|^2$$

for $\mathcal{V} \subseteq \mathcal{M}$, $|\mathcal{V}| \neq \infty$.

Convex Problem (SDP) [details in this paper]

$$\begin{aligned} \underset{\boldsymbol{Z},\boldsymbol{w}\in[0,1]^N}{\text{minimize}} \quad & \mathrm{tr}\{\boldsymbol{Z}\} \\ \text{subject to} \quad & \boldsymbol{1}^T\boldsymbol{w} = K, \\ & \boldsymbol{M}^{T/2}\boldsymbol{P}^{-1}(\boldsymbol{w})\boldsymbol{M}^{1/2} \preceq \boldsymbol{Z} \end{aligned}$$

where $\boldsymbol{M} = \boldsymbol{K}^{-1}\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^T]\boldsymbol{K}^{-1}$ and

$$\boldsymbol{P}(\boldsymbol{w}) = \boldsymbol{K}^{-2} + \gamma^{-1}\boldsymbol{K}^{-1}\mathrm{diag}(\boldsymbol{w}) + \gamma^{-1}\mathrm{diag}(\boldsymbol{w})\boldsymbol{K}^{-1} + \gamma^{-2}\mathrm{diag}(\boldsymbol{w}).$$

Design Problem

Cost function: Stable Regressors Selection

$$\underset{\mathcal{S}\subset\mathcal{M},|\mathcal{S}|=K}{\text{minimize}} \quad q(\text{Cov}\{\hat{\boldsymbol{\alpha}}_{\mathcal{S}}\})$$

Submodular Problem (Greedy) [details in this paper]

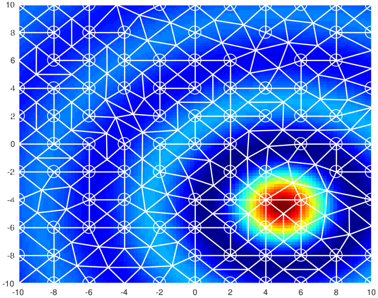
$$\underset{\mathcal{S}\subset\mathcal{M},|\mathcal{S}|=K}{\text{minimize}} \quad \frac{\ln\det\{\boldsymbol{C}_{\mathcal{S}}\}}{2\ln\det\{\boldsymbol{K}_{\mathcal{S}}+\mu K\boldsymbol{I}_K\}}$$

where $\boldsymbol{C}_{\mathcal{S}} = \mathbb{E}[\boldsymbol{y}_{\mathcal{S}}\boldsymbol{y}_{\mathcal{S}}^T]$ and

$$[\boldsymbol{K}_{\mathcal{S}}]_{ij} = k(x_i, x_j),\ x_i, x_j \in \mathcal{S},$$

is the (i,j)th entry of the kernel matrix.

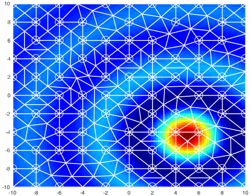
Static field example (0)



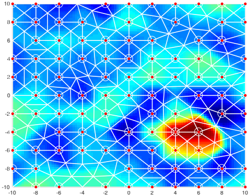
Wave field

- 2-D field estimation
- Rectangular domain of 10×10m
- Source located at coordinates $(x, y) = (5, -4.5)$
- Noise covariance $\mathbf{\Sigma} = \text{Toeplitz}\{1, \rho, \ldots, \rho^{N-1}\}$.
- Gaussian radial basis kernel with $\sigma = 0.8$.

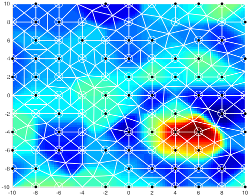
Static field example (1)



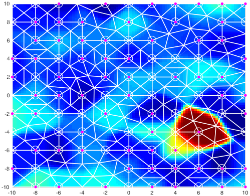
(a) Ground truth



(b) All sensors ($N = 97$)

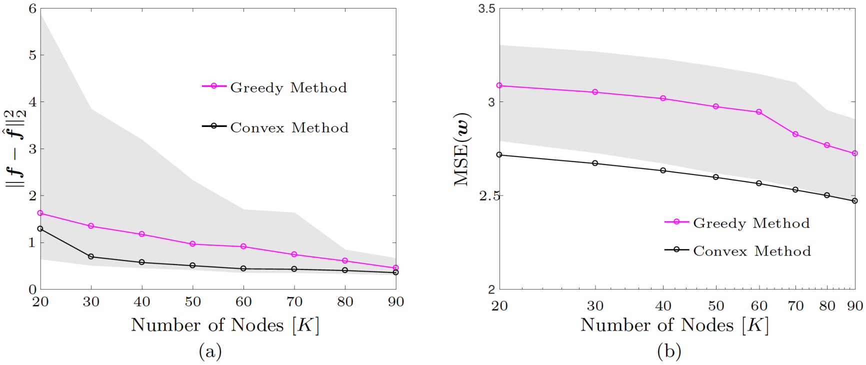


(c) Convex solution ($K = 67$)



(d) Greedy solution ($K = 67$)

Static field example (2)



Comparison of the proposed methods. Shaded gray area shows performance of random samplers. (a) Reconstruction error. (b) MSE($\boldsymbol{w}$).

Conclusions

- Sampling metrics for kernel-based signal reconstruction were proposed.
- Sparse samplers, based on the presented metrics, can be designed efficiently through the convex and submodular machinery.
- The proposed greedy approach provides a tractable alternative for large scale problems without high degradation in performance.

- Outlook
  - Test of sampling strategies with real data and appropriate kernel functions, e.g., harmonics functions for acoustic field reconstruction.
  - Extend results to other kernels methods for statistical inference, e.g., kernel-based detection.

Thank you!

Questions?