Universität Bremen* | Keio University | Fraunhofer Heinrich-Hertz-Institut | ant Dept. of Communications Engineering

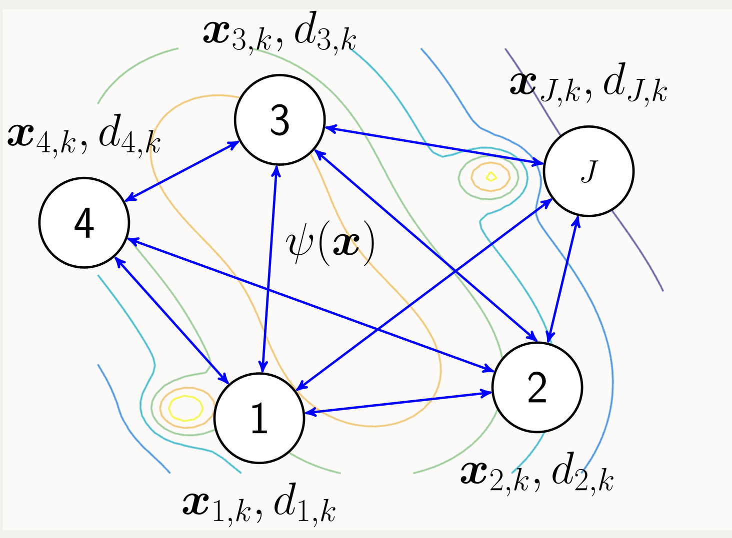
A Hybrid Dictionary Approach for Distributed Kernel Adaptive Filtering in Diffusion Networks

Ban-Sok Shin, Armin Dekorsy
Department of Communications Engineering
University of Bremen, Germany
Email: {shin, dekorsy} @ant.uni-bremen.de

Masahiro Yukawa
Dept. of Electronics and Electrical Engineering
Keio University, Japan
Email: yukawa@elec.keio.ac.jp

Renato L.G. Cavalcante
Fraunhofer Heinrich Hertz Institute
Technical University of Berlin, Germany
Email: renato.cavalcante@hhi.fraunhofer.de

Motivation



Scenario:

- Mobile sensor network of J nodes samples **nonlinear spatial field** $\psi(\boldsymbol{x})$ e.g. spatial distribution of physical quantity
- $\psi(\boldsymbol{x})$ consists of **local part** ψ_{L} with **high frequency** component and **global part** ψ_{G} with **low frequency** component
- Each node j acquires noisy measurement of $\psi(\boldsymbol{x})$ at current Cartesian position $\boldsymbol{x}_{j,k}$:

$$d_{j,k} = \psi_{\mathsf{G}}(\boldsymbol{x}_{j,k}) + \psi_{\mathsf{L}}(\boldsymbol{x}_{j,k}) + n_{j,k}$$

- **Objective:** Distributed nonlinear reconstruction of global ψ_{G} and individual reconstruction of local ψ_{L} in direct vicinity of each node
- Kernel adaptive filter (KAF) suitable for nonlinear approximation of $\psi(\boldsymbol{x})$
- SotA distributed KAFs use one global and fixed dictionary $\mathcal{D}$ for all nodes
- **Problem:** high frequency components in direct vicinity of nodes cannot be recovered with global dictionary

Approach: Use hybrid dictionary with fixed and dynamic part for global and local reconstruction.

Dictionary Strategies for Distributed KAF

SotA global dictionary [1,2]:

- Each node j equipped with KAF $\varphi_{j,k}$ using kernel κ
- All nodes use common and fixed dictionary $\mathcal{D} = \{\kappa(\cdot, \bar{\boldsymbol{x}}_n)\}_{n=1}^N$
- Estimate of KAF for arbitrary input sample $\boldsymbol{x}$ at node j:

$$\varphi_{j,k}(\boldsymbol{x}) := \sum_{n=1}^{N} w_{j,n}^k \kappa(\boldsymbol{x}, \bar{\boldsymbol{x}}_n) = \boldsymbol{w}_{j,k}^\mathsf{T}\boldsymbol{\kappa}(\boldsymbol{x})$$

with $\boldsymbol{w}_{j,k} := [w_{j,1}^k, \dots, w_{j,N}^k]^\mathsf{T}$ and $\boldsymbol{\kappa}(\boldsymbol{x}) := [\kappa(\boldsymbol{x}, \bar{\boldsymbol{x}}_1), \dots, \kappa(\boldsymbol{x}, \bar{\boldsymbol{x}}_N)]^\mathsf{T}$.

Proposed hybrid dictionary:

- Employ **global** κ_{G} and **local** κ_{L} kernel with separate dictionaries
- **Global dictionary** $\mathcal{D}^{\mathsf{G}} := \{\kappa_{\mathsf{G}}(\cdot, \bar{\boldsymbol{x}}_n^{\mathsf{G}})\}_{n=1}^N$:
  - fixed and common to all J nodes
  - initialized a priori
- **Local dictionary** $\mathcal{D}_{j,k}^{\mathsf{L}} := \{\kappa_{\mathsf{L}}(\cdot, \boldsymbol{x}_{j,n})\}_{n \in \mathcal{R}_{j,k}}$:
  - adapted by each node j per time k based on node positions $\boldsymbol{x}_{j,k}$
  - $\mathcal{R}_{j,k}$ indicates which samples $\boldsymbol{x}_{j,k}$ are included into $\mathcal{D}_{j,k}^{\mathsf{L}}$
- Separate node-specific $\boldsymbol{w}_{j,k}$ and $\boldsymbol{\kappa}_j(\boldsymbol{x})$ into global and local parts

$$\boldsymbol{w}_{j,k} := \begin{bmatrix} \boldsymbol{w}_{j,k}^{\mathsf{G}} \\ \boldsymbol{w}_{j,k}^{\mathsf{L}} \end{bmatrix}, \quad \boldsymbol{\kappa}_j(\boldsymbol{x}) := \begin{bmatrix} \boldsymbol{\kappa}_{\mathsf{G}}(\boldsymbol{x}) \\ \boldsymbol{\kappa}_{\mathsf{L},j}(\boldsymbol{x}) \end{bmatrix}$$

- Evaluate local $\boldsymbol{\kappa}_{\mathsf{L},j}(\boldsymbol{x})$ wrt. $\mathcal{D}_{j,k}^{\mathsf{L}}$, global $\boldsymbol{\kappa}_{\mathsf{G}}(\boldsymbol{x})$ wrt. $\mathcal{D}^{\mathsf{G}}$
- Nodes only share global $\boldsymbol{w}_{j,k}^{\mathsf{G}}$ with neighboring nodes
- Each node j able to locally refine reconstruction based on $\mathcal{D}_{j,k}^{\mathsf{L}}$ and $\boldsymbol{w}_{j,k}^{\mathsf{L}}$

Local Dictionary Learning

- Each node j examines current position $\boldsymbol{x}_{j,k}$ for inclusion into local $\mathcal{D}_{j,k}^{\mathsf{L}}$
- Position $\boldsymbol{x}_{j,k}$ needs to pass **coherence criterion**:

$$\max_{n \in \mathcal{R}_{j,k}} |\kappa_{\mathsf{L}}(\boldsymbol{x}_{j,k}, \boldsymbol{x}_{j,n})| \leq \tau, \quad 0 < \tau \leq 1.$$

- **A priori estimation error** $e_{j,k} := d_{j,k} - \boldsymbol{w}_{j,k}^\mathsf{T}\boldsymbol{\kappa}(\boldsymbol{x}_{j,k})$ for $\boldsymbol{x}_{j,k}$ needs to satisfy

$$|e_{j,k}/d_{j,k}| > \varepsilon, \quad \varepsilon \geq 0$$

- If both criteria are fulfilled, $\boldsymbol{x}_{j,k}$ is included via $\mathcal{D}_{j,k+1}^{\mathsf{L}} = \mathcal{D}_{j,k}^{\mathsf{L}} \cup \{\kappa_{\mathsf{L}}(\cdot, \boldsymbol{x}_{j,k})\}$

Diffusion-Based KAF with Hybrid Dictionary

- Define complete kernel Gram matrix

$$\boldsymbol{K}_{j,k} := \begin{bmatrix} \overline{\boldsymbol{K}} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{K}}_{j,k} \end{bmatrix} \in \mathbb{R}^{NR_{j,k} \times NR_{j,k}}.$$

- $\overline{\boldsymbol{K}}$ is global Gram matrix using $\mathcal{D}^{\mathsf{G}}$, **fixed** over time k
- $\widetilde{\boldsymbol{K}}_{j,k}$ is local Gram matrix of node j using $\mathcal{D}_{j,k}^{\mathsf{L}}$, **changing** over time k
- **Two-step diffusion-based KAF [3]:**
  1. **Local adaptive update** per node j on complete $\boldsymbol{w}_{j,k}$

$$\boldsymbol{w}_{j,k+1} = \boldsymbol{w}_{j,k} - \mu\left(\boldsymbol{w}_{j,k} - P_{H_{j,k}}^{\boldsymbol{K}_{j,k}}(\boldsymbol{w}_{j,k})\right)$$

  with $\boldsymbol{K}_{j,k}$-orthogonal projection $P_{H_{j,k}}^{\boldsymbol{K}_{j,k}}(\boldsymbol{w}_{j,k})$ of $\boldsymbol{w}_{j,k}$ onto hyperplane

$$H_{j,k} := \left\{\boldsymbol{w} \in \mathbb{R}^{NR_{j,k}} | \langle \boldsymbol{w}, \boldsymbol{K}_{j,k}^{-1}\boldsymbol{\kappa}(\boldsymbol{x})\rangle_{\boldsymbol{K}_{j,k}} = d_{j,k}\right\}$$

  2. **Consensus averaging** with neighboring nodes only on global $\boldsymbol{w}_{j,k+1}^{\mathsf{G}}$

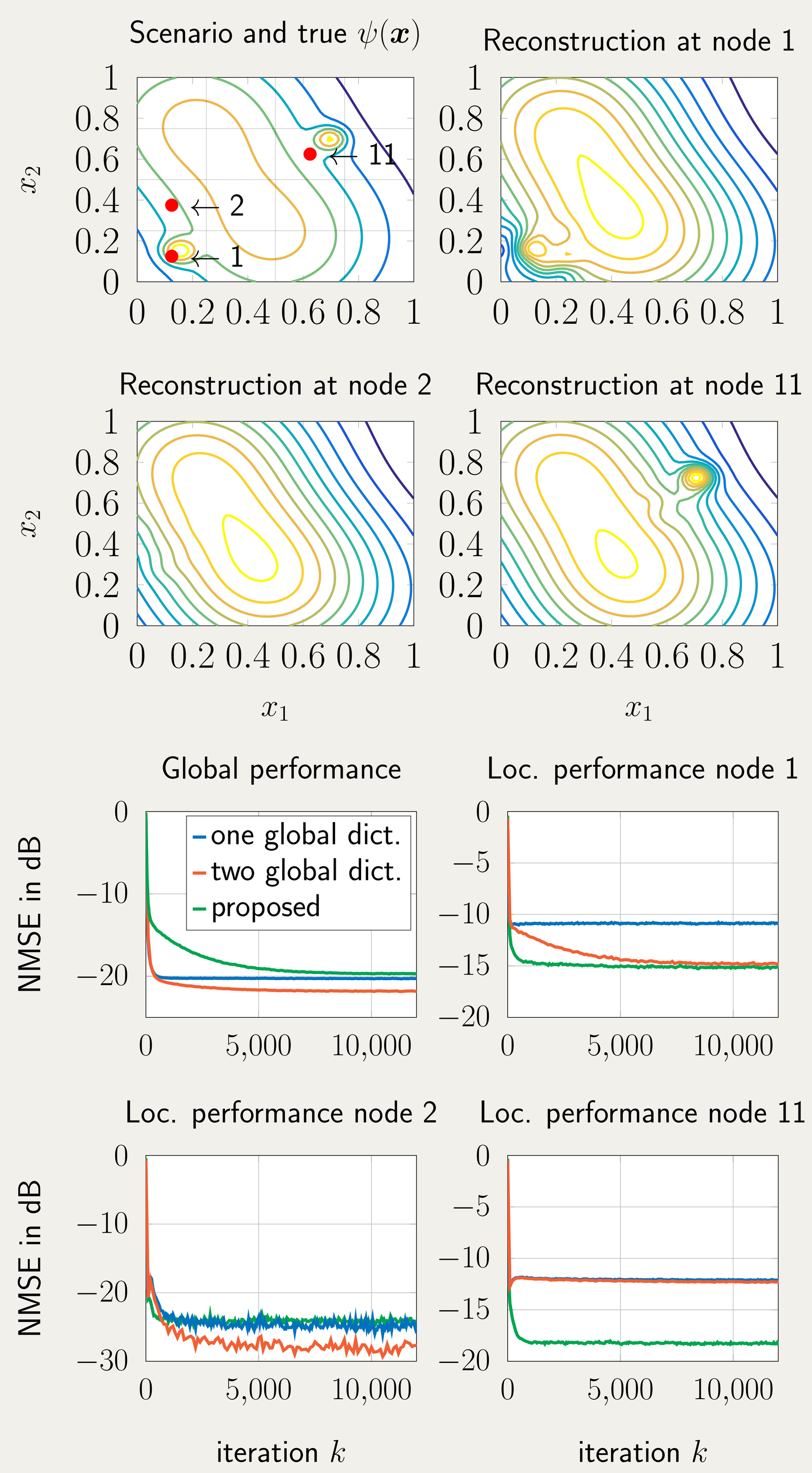
Numerical Evaluation

- Spatial reconstruction of **multiple Gaussian functions**:

$$\psi_{\mathsf{G}}(\boldsymbol{x}) = \sum_{\ell=1}^{2} \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{p}_\ell^{\mathsf{G}}||^2}{2 \cdot 0.3^2}\right), \quad \psi_{\mathsf{L}}(\boldsymbol{x}) = \sum_{\ell=1}^{2} 0.8 \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{p}_\ell^{\mathsf{L}}||^2}{2 \cdot 0.04^2}\right)$$

with $\boldsymbol{p}_1^{\mathsf{G}} = [0.5, 0.2]^\mathsf{T}$, $\boldsymbol{p}_2^{\mathsf{G}} = [0.2, 0.8]^\mathsf{T}$, $\boldsymbol{p}_1^{\mathsf{L}} = [0.7, 0.7]^\mathsf{T}$, $\boldsymbol{p}_2^{\mathsf{L}} = [0.15, 0.15]^\mathsf{T}$

- $J = 16$ nodes, $1/\sigma_n^2 = 10\,\text{dB}$, 100 realizations (noise & topology)
- Separate unit-square into 16 equal square regions, one for each node
- Random movement of nodes in their specific region
- Gaussian kernel with bandwidth ζ: $\kappa(\boldsymbol{x}_j, \boldsymbol{x}_n) = \exp\left(-||\boldsymbol{x}_j - \boldsymbol{x}_n||^2/2\zeta^2\right)$
- **Global** and **local** Gaussian kernel with $\zeta_{\mathsf{G}} = 0.3$ and $\zeta_{\mathsf{L}} = 0.04$
- **Global dictionary** contains all 16 center points of node regions



- **True** $\psi(\boldsymbol{x})$ and **individual reconstructions** at three selected nodes
- All nodes with **good estimate** of global field ψ_{G}
- **Nodes 1** and **11** able to recover local high frequency components
- Reconstruction at **node 2** not tampered by other nodes
- **One global dictionary** with $\zeta_{\mathsf{G}} = 0.04$
- **Two global dictionaries** with $\zeta_{\mathsf{G},1} = 0.3, \zeta_{\mathsf{G},2} = 0.04$
- **Hybrid dictionary** enhances local performance significantly
- Local performance of node 2 similar to other approaches

Conclusion

- **Novel dictionary learning** scheme for distributed KAFs proposed
- **Hybrid dictionary** improves performance of local approximation at nodes
- **Future work:** Extend local dictionary learning by multiple kernels and include selection of kernel with best fit

[1] P. Bouboulis, S. Chouvardas, and S. Theodoridis, "Online distributed learning over networks in RKH spaces using random fourier features," arXiv:1703.08131 [cs.LG], 2017.
[2] S. Chouvardas and M. Draief, "A diffusion kernel LMS algorithm for nonlinear adaptive networks," in IEEE ICASSP, March 2016.
[3] B.-S. Shin, M. Yukawa, R.L.G. Cavalcante, and A. Dekorsy, "Distributed adaptive learning with multiple kernels in diffusion networks," arXiv:1801.07087 [eess.SP], submitted to IEEE Transactions on Signal Processing, 2018.

TZi

Prof. Dr.-Ing. A. Dekorsy
Department of Communications Engineering
Institute for Telecommunication and High-Frequency Techniques
University of Bremen
www.ant.uni-bremen.de

ith University of Bremen