Subset Selection for Kernel-Based Signal Reconstruction

Coutino, Mario; Chepuri, Sundeep Prabhakar; Leus, Geert

DOI
10.1109/ICASSP.2018.8461510

Publication date
2018

Document Version
Final published version

Published in
2018 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2018 - Proceedings

SUBSET SELECTION FOR KERNEL-BASED SIGNAL RECONSTRUCTION

Mario Coutino, Sundeep Prabhakar Chepuri, Geert Leus

Delft University of Technology, Delft, The Netherlands.

ABSTRACT

In this work, we introduce subset selection strategies for signal reconstruction based on kernel methods, particularly for the case of kernel-ridge regression. Typically, these methods are employed for exploiting known prior information about the structure of the signal of interest. We use the mean squared error and a scalar function of the covariance matrix of the kernel regressors to establish metrics for the subset selection problem. Despite the NP-hard nature of the problem, we introduce efficient algorithms for finding approximate solutions for the proposed metrics. Finally, numerical experiments demonstrate the applicability of the proposed strategies.

Index Terms— Kernel regression, kernel-based signal reconstruction, sensor selection, optimal subset selection, submodularity

1. INTRODUCTION

Some of the most successful algorithms in modern machine learning, such as the support vector machine [1], lever structural information through *kernels* [2]. The ability that kernels have to propagate non-linear relations through linear ones offers great flexibility to many regression (signal reconstruction) [3, 4] and classification (detection) [5, 6] tasks.

For example, when a process, e.g., a field evolution or a diffusion, is measured using a sensor network, the following question arises: *how do we use the structural information in the data to improve our estimation performance?* This question is often answered in practice through kernel methods [6]. By using the data structure, e.g., manifold discretization, orography map, correlation map, etc., kernels capable of capturing non-linear relations between measurements can be defined and employed in the regression problem to improve the estimation performance. To illustrate this, let us consider spectrum cartography [7, 8]. Here, kernel methods can be employed to use the information of the network topology or orography map of the terrain to reconstruct spectral maps. Similar applications are found within the field of graph signal processing [9, 11, 10], where kernels can be used to capture the underlying graph structure.

Most of prior works related to kernel methods focus on the reconstruction of the signal or on the estimation of an appropriate kernel for describing the non-linear relations in the data [11, 12, 13]. However, here we mainly focus on the selection problem. That is, given a sensor network or a set of possible acquisition points, a reconstruction algorithm based on kernels, and a sensing budget, *how can we select the subset of measurements, with fixed cardinality, to provide the best possible reconstruction performance?* This problem can be related to the one addressed by the kernelized Lasso [14], which aims to obtain a parsimonious description with a reduced number of non-zero regressors. Differently from this method, instead of using a formulation including the ℓ_1-norm to enforce sparsity in the solution, effectively changing the cost to optimize, we opt for a direct approach through simultaneous measurement and regressor selection by levering the representer theorem [15]. Even though other methods based on *landmark* measurements (Nyström methods) are possible, e.g., [16], they only consider the approximation of the kernel matrix by a CUR-like decomposition. In addition, despite the computational benefits of these methods, they do not deal with missing measurements, i.e., subset selection, which is the main focus of this work.

The rest of this paper is arranged as follows. Section 2 introduces the measurement model and Section 3 discusses signal reconstruction using kernel methods. In Section 4, the problem of sparse sampler design for kernel-based reconstruction and algorithms for its solution are introduced. In Section 5, numerical experiments are presented. Finally, the paper concludes with Section 6.

2. MEASUREMENT MODEL

Let us assume that a continuous function $f : \mathcal{M} \mapsto \mathbb{R}$, defined over a manifold $\mathcal{M}$, is sampled over a finite support $\mathcal{S} = \{x_i : x_i \in \mathcal{M}, \text{for } 1 \leq i \leq K\}$. Furthermore, the set of samples taken at the finite domain $\mathcal{S}$, with $|\mathcal{S}| = K$, i.e., $\{f(x_i)\}_{x_i \in \mathcal{S}}$, are stacked in a single vector $\boldsymbol{f}_{\mathcal{S}} \in \mathbb{R}^K$ and treated as the *signal vector*. Therefore, only the information from a subset of points $\mathcal{S} \subset \mathcal{M}$ is available. This situation arises in many practical applications, as sampling over the entire domain is not feasible or undesirable due to temporal, spatial or economical constraints. For instance, we can con-

This research is supported in part by the ASPIRE project (project 14926 within the STW OTP programme), financed by the Netherlands Organization for Scientific Research (NWO). Mario Coutino is partially supported by CONACYT. The code of this article can be found at https://gitlab.com/fruzti/kernel_point_selection.

sider a tessellation (mesh) performed over a surface and/or terrain. This approach is commonly employed in field estimation through finite-element methods [17]. In this scenario, it is possible to define a stable mesh that covers the whole surface. However, it might be impossible to deploy sensors at all vertices of this tessellation for data acquisition. Therefore, from all possible vertices in the mesh, we can only measure the value of the function f at a certain subset of them.

Mathematically, we can express the acquisition of data, under noise, through the following linear model

$$[\boldsymbol{y}_{\mathcal{S}}]_i \coloneqq y(x_i) = f(x_i) + n(x_i) = [\boldsymbol{f}_{\mathcal{S}}]_i + [\boldsymbol{n}_{\mathcal{S}}]_i \in \mathbb{R}, \quad (1)$$

where $y(x_i)$ is the noise-corrupted measurement at position $x_i \in \mathcal{M}$. The vector $\boldsymbol{n}_{\mathcal{S}} \in \mathbb{R}^K$ is a noise term that is assumed to be zero-mean Gaussian distributed with *known* covariance matrix $\boldsymbol{\Sigma}_{\mathcal{S}}$, i.e., $\boldsymbol{n}_{\mathcal{S}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathcal{S}})$.

3. KERNEL-BASED SIGNAL RECONSTRUCTION

The goal of kernel-based signal reconstruction [4] is to find an estimate $\hat{f}$ of the underlying signal from the available measurements $\boldsymbol{y}_{\mathcal{S}} \in \mathbb{R}^K$ [cf. (1)]. Here, prior knowledge about the underlying structure of the signal of interest is provided in the form of a kernel. For example, in wave field reconstruction the structural information of the data, e.g., shadowing effects, surface tessellation, etc., might be encoded in the kernel to obtain better reconstruction results.

A common assumption, which provides regularity to the space of feasible functions for reconstruction, is that the continuous function f belongs to a reproducing kernel Hilbert space (RKHS), $\mathcal{H}$, defined by a kernel. This assumption implies [2]

$$\mathcal{H} = \left\{ f : f(x) = \sum_{x_i \in \mathcal{M}} \alpha_i k(x_i, x),\ \alpha_i \in \mathbb{R} \right\}, \quad (2)$$

where $k : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is a symmetric kernel map satisfying $\sup_{x,y} k(x, y) < \infty$. To obtain an estimate of the continuous function f, using the RKHS assumption, the following optimization problem can be solved

$$\hat{f} = \underset{f \in \mathcal{H}}{\arg\min} \frac{1}{K} \sum_{x_i \in \mathcal{S}} \mathcal{L}(y(x_i), f(x_i)) + \mu \Omega(\|f\|_{\mathcal{H}}), \quad (3)$$

where $\mathcal{L}(\cdot, \cdot)$ is a loss function that measures the fitness of the estimate and $\Omega(\cdot)$ can be used to promote smoothness in $\mathcal{H}$ controlled by the regularization parameter μ. Here, $\|f\|^2_{\mathcal{H}}$ is the inner product in $\mathcal{H}$ induced by the kernel map. As in this work we focus on kernel-ridge regression (KRR) [6], we employ the *square loss* function, i.e., $\mathcal{L}(x, y) = (x - y)^2$, and the smooth function $\Omega(\cdot) = (\cdot)^2$ as regularizer, i.e.,

$$\hat{f} = \underset{f \in \mathcal{H}}{\arg\min} \frac{1}{K} \sum_{x_i \in \mathcal{S}} (y(x_i) - f(x_i))^2 + \mu \|f\|^2_{\mathcal{H}}. \quad (4)$$

Notice that formulation (4) generalizes other types of algorithms that do not explicitly make use of kernels. Therefore, the metrics here presented can be extended to methods particularizing (4), e.g., gaussian process [6, Ch. 16].

At this point we can formally state the problem of interest. *Given problem* (4)*, the model statistics, and a kernel map $k(\cdot, \cdot)$, can we find the subset $\mathcal{S} \subset \mathcal{M}$, with cardinality $|\mathcal{S}| = K$, that provides the best reconstruction of $\{f(x_i)\}_{x_i \in \mathcal{V}}$, with $\mathcal{S} \subseteq \mathcal{V} \subseteq \mathcal{M}$?*

4. SPARSE SAMPLER DESIGN

When the function f is estimated through the solution of (3) it can be noticed that the solution $\hat{f}$ consists of an expansion over a (possibly *infinite*) number of basis functions [cf. (2)]. Fourtunately, using the representer theorem [15] it is possible to describe the solution $\hat{f}$ for (2) by the following series:

$$\hat{f}(x) = \sum_{x_i \in \mathcal{S}} \alpha_i k(x_i, x). \quad (5)$$

Using vector notation and the above expansion, we can rewrite the optimization problem in (4) as:

$$\hat{\boldsymbol{\alpha}}_{\mathcal{S}} = \begin{array}{ll} \underset{\boldsymbol{\alpha} \in \mathbb{R}^K}{\arg\min} & \frac{1}{K} \|\boldsymbol{e}\|^2 + \mu \boldsymbol{\alpha}^T \boldsymbol{K}_{\mathcal{S}} \boldsymbol{\alpha} \\ \text{subject to} & \boldsymbol{e} = \boldsymbol{y}_{\mathcal{S}} - \boldsymbol{K}_{\mathcal{S}} \boldsymbol{\alpha} \end{array}, \quad (6)$$

where the loss function is equivalently expressed as an equality constraint and the fact that $\|f\|_{\mathcal{H}} = \boldsymbol{\alpha}^T \boldsymbol{K}_{\mathcal{S}} \boldsymbol{\alpha}$ has been used. Here, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^T \in \mathbb{R}^K$ is the vector with the expansion coefficients, 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.

The optimal solution $\hat{\boldsymbol{\alpha}}_{\mathcal{S}}$ for (6), can be written in terms of the set $\mathcal{S}$ as

$$\hat{\boldsymbol{\alpha}}_{\mathcal{S}} = \left[\boldsymbol{K}_{\mathcal{S}} + \gamma \boldsymbol{I}_K \right]^{-1} \boldsymbol{y}_{\mathcal{S}}, \quad (7)$$

where $\gamma = \mu K$, and $\boldsymbol{I}_K$ is the $K \times K$ identity matrix. Therefore, the residual for a given $\mathcal{S}$ with respect to a measurement at $x_j \in \mathcal{M}$ would be given by

$$e(x_j, \mathcal{S}) \quad = \quad y(x_j) - \boldsymbol{k}^T_{\mathcal{S},j} \hat{\boldsymbol{\alpha}}_{\mathcal{S}}, \quad (8)$$

where $[\boldsymbol{k}_{\mathcal{S},j}]_i = k(x_i, x_j),\ x_i \in \mathcal{S}$. In the following, we lever both (7) and (8) to define metrics for optimal subset selection.

4.1. Mean Squared Error

Focusing on the *expected* performance of the sampler, with respect to the statistics of $\{y(x_j)\}_{x_j \in \mathcal{M}}$ (assumed to be known), we can consider the following mean squared error (MSE) cost as a metric for designing the sparse sampler:

$$\text{MSE}_{\mathcal{M}}(\mathcal{S}) = \int_{x \in \mathcal{M}} |e(x, \mathcal{S})|^2 dx. \quad (9)$$

In practice, evaluation and optimization of (9) might lead to a numerically involved task. Therefore, we propose to discretize the manifold $\mathcal{M}$, e.g., surface tessellation, to obtain an approximation of (9).

Consider the set $\mathcal{V} = \{x_j : x_j \in \mathcal{M}, 1 \leq j \leq |\mathcal{V}|\}$, $\mathcal{S} \subseteq \mathcal{V} \subseteq \mathcal{M}$. Then, we can approximate (9) by the following expression

$$\text{MSE}_{\mathcal{V}}(\mathcal{S}) \coloneqq \sum_{x_j \in \mathcal{V}} |e(x_j, \mathcal{S})|^2, \tag{10}$$

where the scaling $1/|\mathcal{V}|$ has been omitted. Now, let us introduce the following notation:

$$\boldsymbol{y}_{\mathcal{S}} = \boldsymbol{\Phi}_{\boldsymbol{w}} \boldsymbol{y}, \tag{11}$$

where $\boldsymbol{\Phi}_{\boldsymbol{w}} \in \{0,1\}^{K \times |\mathcal{V}|}$ is a binary selection matrix defined by the selection vector $\boldsymbol{w} \in \{0,1\}^{|\mathcal{V}|}$ that indicates the available points. In this setup, the nth point is (not) sampled if $[\boldsymbol{w}]_n$ is set to one (zero). Here, $\boldsymbol{y} \in \mathbb{R}^{|\mathcal{V}|}$ results from stacking the elements of $\{y(x_j)\}_{x_j \in \mathcal{V}}$ in a vector.

The error vector $\boldsymbol{e}(\mathcal{S})$, built from stacking the elements of $\{e(x_j, \mathcal{S})\}_{x_j \in \mathcal{V}}$, can be shown to be given in terms of $\boldsymbol{w}$ as

$$\boldsymbol{e}(\boldsymbol{w}) = \left[\boldsymbol{K}^{-1} + \gamma^{-1}\text{diag}(\boldsymbol{w})\right]^{-1} \boldsymbol{K}^{-1} \boldsymbol{y}, \tag{12}$$

where $[\boldsymbol{K}]_{ij} = k(x_i, x_j), x_i, x_j \in \mathcal{V}$. Therefore, the expression (10) can be written in terms of $\boldsymbol{w}$ as

$$\text{MSE}_{\mathcal{V}}(\mathcal{S}) = \text{MSE}_{\mathcal{V}}(\boldsymbol{w}) = \text{tr}\big\{\boldsymbol{P}^{-1}(\boldsymbol{w})\boldsymbol{M}\big\}. \tag{13}$$

Here, we have defined $\boldsymbol{M} = \boldsymbol{K}^{-1}\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^T]\boldsymbol{K}^{-1}$ and

$$\begin{aligned} \boldsymbol{P}(\boldsymbol{w}) \quad = \quad & \boldsymbol{K}^{-2} + \gamma^{-1}\boldsymbol{K}^{-1}\text{diag}(\boldsymbol{w}) + \\ & \gamma^{-1}\text{diag}(\boldsymbol{w})\boldsymbol{K}^{-1} + \gamma^{-2}\text{diag}(\boldsymbol{w}). \end{aligned} \tag{14}$$

As a result, we can formulate the optimization problem

$$\begin{array}{ll} \underset{\boldsymbol{w} \in \{0,1\}^N}{\text{minimize}} & \text{tr}\big\{\boldsymbol{P}^{-1}(\boldsymbol{w})\boldsymbol{M}\big\} \\ \text{subject to} & \|\boldsymbol{w}\|_0 = K \end{array}, \tag{15}$$

to obtain the best subset of K measurements that minimises $\text{MSE}_{\mathcal{V}}(\boldsymbol{w})$. Despite the fact that the problem in (15) is not convex, we can relax both the Boolean and ℓ_0-norm constraints in order to obtain a convex relaxation of (15). Introducing a new variable $\boldsymbol{Z} = \boldsymbol{M}^{T/2}\boldsymbol{P}^{-1}(\boldsymbol{w})\boldsymbol{M}^{1/2}$, and applying the relaxations mentioned above, we obtain the following optimization problem

$$\begin{array}{ll} \underset{\boldsymbol{Z}, \boldsymbol{w} \in [0,1]^N}{\text{minimize}} & \text{tr}\big\{\boldsymbol{Z}\big\} \\ \text{subject to} & \boldsymbol{1}^T\boldsymbol{w} = K, \\ & \boldsymbol{M}^{T/2}\boldsymbol{P}^{-1}(\boldsymbol{w})\boldsymbol{M}^{1/2} \preceq \boldsymbol{Z} \end{array}. \tag{16}$$

The above problem is indeed a convex program as the second constraint can be expressed as a linear matrix inequality (LMI) in $\boldsymbol{w}$, i.e.,

$$\begin{bmatrix} \boldsymbol{Z} & \boldsymbol{M}^{1/2} \\ \boldsymbol{M}^{T/2} & \boldsymbol{P}(\boldsymbol{w}) \end{bmatrix} \succeq 0, \tag{17}$$

where the linearity of (17) with respect to $\boldsymbol{w}$ of (17) is clear from the definition of $\boldsymbol{P}(\boldsymbol{w})$ in (14).

4.2. Stable Regressors Selection

Alternatively, instead of selecting the subset that minimizes the $\text{MSE}_{\mathcal{V}}(\mathcal{S})$ in (10), assume we aim to obtain an estimator that minimizes a scalar function, $q : \mathbb{R}^{K \times K} \to \mathbb{R}$, of the covariance matrix of the regressor, $\hat{\boldsymbol{\alpha}}_{\mathcal{S}}$, e.g., a stable representation of $\hat{f}$ in a feature space given by $\hat{\boldsymbol{\alpha}}$ [6]. That is,

$$\underset{\mathcal{S} \subset \mathcal{M}, |\mathcal{S}| = K}{\text{minimize}} \quad q(\text{Cov}\{\hat{\boldsymbol{\alpha}}_{\mathcal{S}}\}) \quad . \tag{18}$$

Here, the closed form expression for the covariance matrix of the estimate is given by

$$\text{Cov}\{\hat{\boldsymbol{\alpha}}_{\mathcal{S}}\} = (\boldsymbol{K}_{\mathcal{S}} + \gamma \boldsymbol{I}_K)^{-1} \boldsymbol{C}_{\mathcal{S}} (\boldsymbol{K}_{\mathcal{S}} + \gamma \boldsymbol{I}_K)^{-1}, \tag{19}$$

where we have defined $\boldsymbol{C}_{\mathcal{S}} = \mathbb{E}[\boldsymbol{y}_{\mathcal{S}}\boldsymbol{y}_{\mathcal{S}}^T]$.

Among possible options for the scalar function $q(\cdot)$, we can consider the commonly used metrics in experimental design, e.g., trace, determinant, etc. In this work, we focus on minimizing the determinant of (19) as we use the submodular machinery for obtaining a near-optimal solution to (18). Furthermore, similar to the convex method, we restrict ourself to a discretised version $\mathcal{V}$ of $\mathcal{M}$ to solve (18).

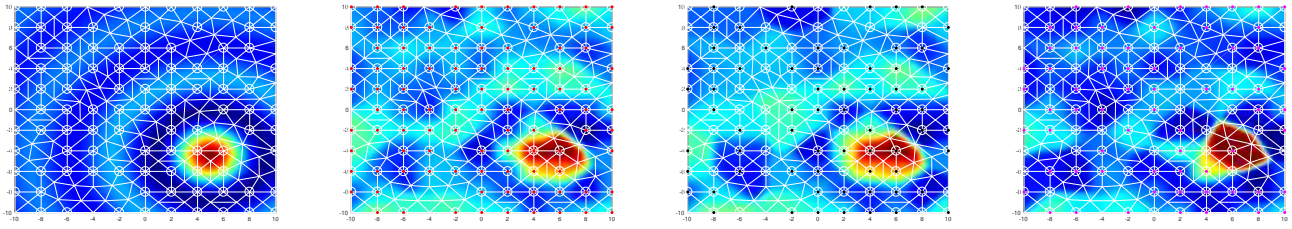
Considering the determinant of the covariance matrix $\text{Cov}\{\hat{\boldsymbol{\alpha}}_{\mathcal{S}}\}$, we propose to use the following objective function for solving the optimization problem in (18):

$$q(\text{Cov}\{\hat{\boldsymbol{\alpha}}_{\mathcal{S}}\}) = \frac{h(\mathcal{S})}{g(\mathcal{S})} = \frac{\ln\det\{\boldsymbol{C}_{\mathcal{S}}\}}{2\ln\det\{\boldsymbol{K}_{\mathcal{S}} + \mu K \boldsymbol{I}_K\}}, \tag{20}$$

where the logarithm of both numerator and denominator in $\det\{\text{Cov}(\hat{\boldsymbol{\alpha}}_{\mathcal{S}})\}$ has been taken. Notice that the function in (20) is a ratio of submodular functions [19], i.e., both $h(\mathcal{S})$ and $g(\mathcal{S})$ are log-determinants of principal submatrices of positive definite matrices. Therefore, Algorithm 1 can be employed to find a near-optimal solution of (18). Even though the results in [19] provide near optimality guarantees for the unconstrained case, i.e., cardinality is not fixed, Algorithm 1 can still be used for minimizing (20) when a fixed number of samples is considered. This is shown in the following section, where results of numerical experiments are presented. A possible alternative is directly consider $\ln\det\{\text{Cov}(\hat{\boldsymbol{\alpha}}_{\mathcal{S}})\}$. As this set function can be written as a difference of submodular functions, strategies similar to the ones described in [20] can be employed for its approximate solution. However, due to the possibility of applying a rank-one update to the determinant evaluations at every step of Algorithm 1, the cost in (20) is considered as a better option in terms of scalability.

5. NUMERICAL EXPERIMENTS

In this section, we show results of the presented sensor selection method for estimating a static 2-D field described by a sinc function on a rectangular domain of 10×10 m. The source is located at the coordinates $(x, y) = (5, -4.5)$ m. A mesh generator [18] is used to define equally spaced points for discrete estimation inside the rectangular domain. From



(a) (b) (c) (d)

Fig. 1: Comparison of the reconstructed field. White circles denote possible sensor locations, and coloured vertices denote active sensors. (a) True field and tesselation. (b) All $N = 97$ available sensors are used. (c) Sensors selected through the convex strategy ($K = 67$). (d) Sensors selected through the (greedy) submodular strategy ($K = 67$).

the stable mesh, a set of $N = 97$ vertices for sensor node deployment are randomly selected. Here, the aim is to estimate the field at all possible locations within the rectangular domain. As it is desired to reconstruct the whole field, a continuous kernel map, $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, given by the Gaussian radial basis (RBF) kernel [6] is considered. For this kernel map, a fixed value $\sigma = 0.8$ is assumed. In addition, we considered $\mathbf{\Sigma} = \text{Toeplitz}(1, \rho, \cdots, \rho^{N-1})$ (typical covariance matrix structure in array processing and field measurements), with $\rho = 0.45$ as noise covariance matrix. Hence, we assume $\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^T] = \mathbf{\Sigma} + \boldsymbol{f}\boldsymbol{f}^T$. The original field, $\boldsymbol{f}$, with the computed mesh is shown in Fig. 1a. The result of the kernel-ridge regression for the full blown sensor network, $N = 97$, and the sampled network using (16) and (18), with $K = 67$, are shown in Fig. 1b-Fig. 1d. In this plot, it can be seen that the degradation of the estimated field, with to respect the full blown solution (Fig. 1b), is not significant. Here, both convex and submodular methods are able to reconstruct an approximation of the source hot spot. However, it is seen that the field structure is better maintained by the convex solution. A comparison of the error $\|\boldsymbol{f} - \hat{\boldsymbol{f}}\|_2^2$, and $\text{MSE}_{\mathcal{V}}(\boldsymbol{w})$ for the different strategies is shown in Fig. 2a and Fig. 2b. In both plots the shaded grey area represents the performance of random samplers. In these plots, it is seen that the convex method outperfoms the greedy and random samplers in terms of $\text{MSE}_{\mathcal{V}}(\boldsymbol{w})$. However, for a given realization of the field measurements, others samplers can outperform the convex-based sampler, as seen in Fig. 2a. Despite that the sampler based on the convex program outperforms the greedy sampler, we require to solve the costly semidefinite program (16) which might not be tractable for large problem instances. In these instances, the proposed greedy method might be employed as its performance follows closely the performance of the convex method (see Fig. 2a). An extension for the methods presented here may rely on online censoring techniques, as the ones presented in [21]. This is left for future research.

Algorithm 1: GREEDYRATIO ALGORITHM

Result: An approximate solution $\mathcal{S}^* \subset \mathcal{V} : |\mathcal{S}^*| = K$
Data: monotone submodular functions h, g
initialization $k = 0, \mathcal{S}_0 = \emptyset, \mathcal{R} = \mathcal{V}$;
while $k \leq K$ **do**
  $v = \arg\min_{v \in \mathcal{R}} \dfrac{h(\mathcal{S}_k \cup \{v\}) - h(\mathcal{S}_k)}{g(\mathcal{S}_k \cup \{v\}) - g(\mathcal{S}_k)}$;
  $\mathcal{S}_{k+1} = \mathcal{S}_k \cup \{v\}$;
  $\mathcal{R} = \{v \in \mathcal{R} | g(\mathcal{S}_{k+1} \cup \{v\}) - g(\mathcal{S}_{k+1}) > 0\}$;
  $k \leftarrow k + 1$;
end
return $\mathcal{S}^* \leftarrow \mathcal{S}_k$;

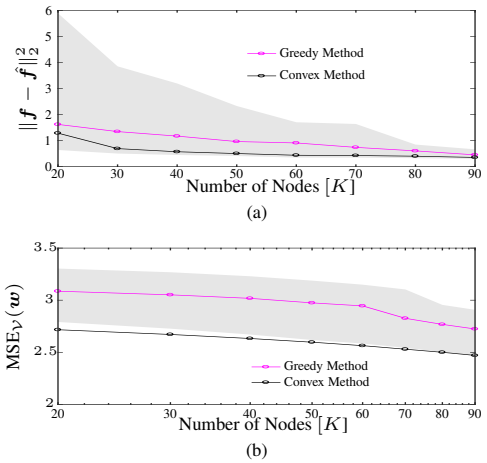


Fig. 2: Comparison of the proposed methods. Shaded grey area shows performance of random samplers. (a) Reconstruction error. (b) $\text{MSE}_{\mathcal{V}}(\boldsymbol{w})$.

6. CONCLUSIONS

In this paper, the problem of optimal sampling for kernel-based signal reconstruction was considered. Expressions based on the mean square error and the covariance matrix of the regressor were proposed as metrics for reducing the sampling density when KRR is employed for signal reconstruction. It was shown that sparse samplers, based on the presented metrics, can be designed efficiently through the convex and submodular machinery. Simulations of field measurements were used to demonstrate the proposed methods. Furthermore, despite that the metrics were designed for KRR, we stressed the fact that they can be extended to methods particularizing the kernel-ridge regression problem.

7. REFERENCES

[1] C. Cortes, and V. Vapnik. "Support-vector networks.", *Machine learning*, vol. 20, no.3, p.p. 273-297, 1995.

[2] N. Aronszajn. "Theory of reproducing kernels.", *Transactions of the American Mathematical Society*, vol. 68, no. 3, p.p. 337-404, 1950.

[3] J.A. Bazerque, and G.B. Giannakis, "Nonparametric basis pursuit via sparse kernel-based learning: A unifying view with advances in blind methods.", *IEEE Signal Proc. Mag.,* vol. 30 no. 4, pp.112-125, 2013.

[4] V. N. Vapnik, et al., "Support vector method for function approximation, regression estimation, and signal processing., *Adv. in Neural Inf. Proc. Systems*, pp. 281-287, 1997.

[5] S. Mika, et al., "Fisher discriminant analysis with kernels.", *Proc. IEEE Signal Process. Society Workshop Neural Netw. for Signal Process.*, pp. 41-48, 1999.

[6] B. Schölkopf, and A.J. Smola, "Learning with kernels: support vector machines, regularization, optimization, and beyond.", *MIT press*, 2002.

[7] J.A. Bazerque, et al., "Group-lasso on splines for spectrum cartography.", *IEEE Trans. on Signal Process.*, vol. 59, no. 10, pp.4648-4663, 2011.

[8] D. Romero, et al., "Spectrum cartography using quantized observations.", *IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 3252-3256, 2015.

[9] A.J., Smola, and R. Kondor. "Kernels and regularization on graphs.", *COLT*, vol. 2777, pp. 144-158, 2003.

[10] D. Romero, et al., "Kernel-based reconstruction of space-time functions on dynamic graphs.", *IEEE J. of Sel. Topics in Signal Process.*, vol. 11, no. 6, pp.856-869, 2017.

[11] D. Romero, et al., "Kernel-based reconstruction of graph signals"., *IEEE Trans. on Signal Process.*, vol. 65, no. 3, pp.764-778, 2017.

[12] S.J. Kim, "Optimal kernel selection in kernel fisher discriminant analysis.", *In ACM Proc. of the 23rd Int. Conf. on Mach. Learn.*, pp. 465-472, 2006.

[13] G.R. Lanckriet, et al., "Learning the kernel matrix with semidefinite programming.", *J. of Mach. Learn. Research*, vol 5, pp. 27-72, 2004.

[14] V. Roth, "The generalized LASSO.", *IEEE Trans. on Neural Netw.*, vol. 15, pp. 16-28, 2004.

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

[16] C. Li, et al., "Fast DPP Sampling for Nyström with Application to Kernel Methods.", *arXiv preprint*, arXiv:1603.06052 2016.

[17] S. C. Brenner and L. R. Scott, "The mathematical theory of finite element methods.", New York, Springer, 2008.

[18] P.O. Persson and G. Strang, "A simple mesh generator in MATLAB," *SIAM Review*, vol. 46, no. 2, pp. 329-345, June 2004.

[19] W. Bai, et al., "Algorithms for optimizing the ratio of submodular functions.", *In Int. Conf. on Mach. Learn.*, pp. 2751-2759. 2016.

[20] M. Coutino, S.P. Chepuri, and G. Leus, "Near-Optimal Sparse Sensing for Gaussian Detection with Correlated Observations.", *arXiv preprint*, arXiv:1710.09676 2017.

[21] E.J. Msechu, and G. B. Giannakis, "Sensor-centric data reduction for estimation with WSNs via censoring and quantization.", *IEEE Trans. on Signal Process.*, vol. 60, no. 1, 400-414, 2012.