Communication Constrained Distributed Spatial Field Estimation Using Mobile Sensor Networks [⋆]

Gennaro Notomista [*] **Magnus Egerstedt** [**]

[*] *School of Mechanical Engineering, Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA 30308 USA (e-mail: g.notomista@gatech.edu).*
[**] *School of Electrical and Computer Engineering, Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA 30308 USA (e-mail: magnus@gatech.edu).*

Abstract: In this paper we address the problem of distributed estimation of spatial fields using mobile sensor networks with communication constraints. These constraints consist of a maximum communication bandwidth which limits the amount of data that can be exchanged between any two nodes of the network at each time instant. An algorithm to select the most significant data to be transferred between neighboring sensor nodes is developed starting from derived analytical error bounds. Moreover, the motion of the network nodes is controlled using a coverage control algorithm with the objective of minimizing the estimation uncertainty of each of the nodes. The resulting communication constrained distributed estimation algorithm is deployed on a team of ground mobile robots in the Robotarium, and its performance is evaluated both in terms of estimation accuracy of a simulated spatial field, and of the amount of data transferred.

Keywords: Autonomous mobile robots, Communication networks, Estimation Algorithms, Mobile robots, Robotics

1. INTRODUCTION

Mobile wireless sensor networks (WSNs) are widely employed in applications of environmental monitoring, which typically involve spatial field estimation tasks (see, e. g., the survey in Akyildiz et al. (2002)). Generally, a mobile sensor network consists of a large number of sensor nodes which, in their basic configuration, are equipped with computation, mobility, sensing, and communication units. These units are responsible for performing the basic tasks for which the sensor nodes are designed: move in the environment and explore it, collect measurement data, process them and communicate them either to a central unit or to their neighboring nodes.

One of the main obstructions to achieving long-term deployment of mobile wireless sensor networks is energy management. While mobility is the main source of energy consumption, in most applications, communication is significantly more energy-consuming than computation, as recognized in Pottie and Kaiser (2000). Thus, in long-term distributed estimation tasks—whereby each node is supposed to build an estimate of an environment field by means of local interactions with its neighbors only—the energy employed for communication needs to be explicitly taken into account.

An approach to reduce the amount of data transferred between sensor nodes in a network has been presented in Kivinen et al. (2004), where a novelty detection algorithm is employed in an online learning framework in order to limit the amount of data that need to be stored and therefore transferred. A similar objective is pursued in Shi et al. (2005), where the authors propose a way of generating small data sets that still keep a high proportion of the information contained in the original, large, dataset. General guidelines for choosing the size of subsets of data based on detailed experimental studies are reported in Das et al. (2015). Finally, with the ultimate goal of decreasing the computational burden of Gaussian Process Regression (GPR) models, massively scalable Gaussian processes are introduced in Wilson et al. (2015).

Communication constraints are not explicitly considered in the approaches reported above, which do not aim at reducing the amount of data that is to be exchanged between the nodes of the sensor network. The amount of data that a node of a mobile WSN has to transfer significantly affects the amount of consumed energy. With the objective of reducing this energy, Tavassolipour et al. (2017) consider information theoretical bounds in order to find the minimum number of bits per symbol to be employed in the communication scheme. Similarly, limited communication capabilities in terms of communication range are considered in Gu and Hu (2012), introducing a Distributed Gaussian Process Regression (DGPR) in which each sensor node only needs to communicate with its

[⋆] This work was supported by the Army Research Lab through ARL DCIST CRA W911NF-17-2-0181.

neighboring nodes. A similar concept has been further explored in Chen et al. (2015), where, however, the problem of keeping the amount of exchanged data between nodes of WSNs bounded has not been specifically investigated. In this paper, we address this gap through the following two contributions:

(i) the errors introduced by the approximations adopted in a distributed estimation framework versus a centralized one are quantified;
(ii) the results of this analysis are leveraged to develop an algorithm to select the most significant data that need to be transferred between neighboring sensor nodes in a WSN.

The remainder of the paper is organized as follows. In the next section, background material is introduced, which will be used in Sections 3 and 4 to develop a communication constrained DGPR algorithm. In Section 5, the proposed distributed estimation framework is implemented on a real multi-robot system.

2. BACKGROUND

In this paper, we use Gaussian process regression as the framework for spatial field estimation Rasmussen and Williams (2006). Gaussian processes are flexible and exhibit good generalization properties thanks to the lack of any underlying model of the process to estimate. In this section, we briefly recall the Gaussian process regression and introduce a way of rendering the estimation process distributed, requiring each node of a WSN to transfer only a fixed amount of data with its neighboring nodes.

2.1 Gaussian Process Regression

From the definition given in Rasmussen and Williams (2006), a Gaussian Process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution. Let $D \subseteq \mathbb{R}^d$ be the input space of a scalar-valued function $f : D \to \mathbb{R}$. Then, f is a GP if, for any index set $J \subset \mathbb{N}$ and $x = \{x_i\}_{i \in J}$, $x_i \in D \; \forall i \in J$, one has that $f(x) = \{f(x_i)\}_{i \in J}$ are Gaussian distributed. A GP is completely specified by its mean function, $\mu : D \to \mathbb{R}$, and its covariance function, also called kernel function, $k : D \times D \to \mathbb{R}$. The value $\mu(x_i)$ is the mean of $f(x_i)$, for $x_i \in D$, whereas $k(x_i, x_j)$ is the covariance between $f(x_i)$ and $f(x_j)$ for $x_i, x_j \in D$. Adopting the same notation used for multivariate Gaussian distributions, we can write $f \sim \mathcal{GP}(\mu, k)$, with which we mean that, for a given $x = [x_1, \ldots, x_n]$, $x_i \in D \; \forall i$, we have $f(x) \sim \mathcal{N}(\mu(x), k(x, x))$. In order to give rise to a valid covariance matrix $k(x, x)$, the function k needs to be symmetric and positive definite.

Spatial field estimations using GP regression are performed using GPs as a prior probability distribution over functions describing the field that is to be estimated. Once the measurements of f at the points $x = [x_1, \ldots, x_M]$, $f(x) = [f(x_1), \ldots, f(x_M)]$, have been performed, one can calculate the mean and the covariance of the posterior probability distribution of the value of the function f at a test point $x^\star \in D$, denoted by $f(x^\star) \mid f(x)$. This calculation has to be performed by each node of a WSN and is carried out by using the marginalization properties of multivariate Gaussian distributions, resulting in the following expression (see, e. g., Rasmussen and Williams (2006)):

$$f(x^\star) \mid f(x) \;\sim\; \mathcal{N}\Big(k(x^\star, x)^\mathsf{T} k(x, x)^{-1} f(x), \\ k(x^\star, x^\star) - k(x^\star, x)^\mathsf{T} k(x, x)^{-1} k(x^\star, x)\Big). \tag{1}$$

2.2 Compactly Supported Kernel Functions and Distributed Gaussian Process Regression

The naive computation of the conditional probability (1) at inference time requires $\mathcal{O}(M^3)$ operations, where M is the number of observations, in order to invert the covariance matrix $k(x, x)$ (Rasmussen and Williams (2006)). For this reason, in practice, the exact implementation cannot handle problems with more than a few thousands observations. To overcome this computational limitation of GPR, a variety of solutions have been proposed such as Kivinen et al. (2004); Shi et al. (2005); Das et al. (2015); Wilson et al. (2015), as discussed in the introduction (see Quinonero-Candela et al. (2007) for a unifying framework for sparse approximations in Gaussian regression models).

The approximation technique we take into consideration in this paper in the context of distributed estimation consists in sparsifying the covariance matrix $k(x, x)$ by making use of compactly supported (CS) kernel functions. This way, GPR can be performed in a distributed fashion. A distributed approach that leverages CS kernel functions to allow spatial estimation using mobile sensor networks is introduced in Gu and Hu (2012). In this work, the authors allow the sensor nodes to transfer all the measurements they have collected to their neighbors. In the next section, we present a way of selecting and communicating the data that are most relevant to the neighbors of a sensor node to improve its estimation. Before that, we give a brief overview of CS covariance functions that are used in GPR and introduce the specific one that is used in this paper.

In Wu (1995), the author provides sufficient conditions for positive definiteness of radial basis functions with compact support. Using the derived conditions, a series of positive definite and CS radial functions, known as Wu's polynomials, can be produced. In the context of spatial estimation for interpolating large datasets, Furrer et al. (2006) show that tapering a covariance matrix with an appropriate CS positive definite function can significantly reduce the computational burden while still leading to asymptotically optimal estimations. The benefits introduced by CS covariance functions in terms of computational efficiency in spatial prediction and data interpolation are recognized also in Gneiting (2002). A constructive way of obtaining CS kernels using functions known as mollifiers—smooth functions with compact support—is presented in Jamshidi and Kirby (2006), where the objective is once again that of significantly reducing the computational complexity inherent in GPR. Finally, the use of CS Radial Basis Fucntion (RBF) kernels for computational improvements and memory reduction in function estimation is investigated in Hamers et al. (2002).

Due to their universality—the property of approximating continuous functions on compact sets with arbitrary accuracy (Steinwart (2001))—in this paper we employ

CS Gaussian RBF kernels (Genton (2001)). These are obtained by mollifying the Gaussian RBF kernel

$$k : (x_1, x_2) \in D \times D \mapsto e^{-\frac{\|x_1 - x_2\|^2}{\sigma^2}} \in \mathbb{R}, \tag{2}$$

where σ is a parameter of the function, by multiplying it by the following CS kernel:

$$k_{\mathrm{c}} : (x_1, x_2) \in D \times D \mapsto \max \left\{ 0, \left(1 - \frac{\|x_1 - x_2\|}{l}\right)^{\nu} \right\} \in \mathbb{R}. \tag{3}$$

In (3), the parameters l and ν need to satisfy the conditions $l > 0$ and $\nu > (d+1)/2$, where d is the dimensionality of the vectors x_i, in order to ensure positive definiteness of k_c. The product of the kernels in (2) and (3),

$$\hat{k} : (x_1, x_2) \in D \times D \mapsto k(x_1, x_2) k_{\mathrm{c}}(x_1, x_2) \in \mathbb{R}, \tag{4}$$

is a CS kernel function and l is a parameter known as the effective range. The meaning of the effective range can be understood observing that $\|x_i - x_j\| \geq l \Rightarrow \hat{k}(x_i, x_j) = 0 \ \forall x_i, x_j \in D$. In the context of WSNs, this implies that two measurements taken at the points x_i and x_j are uncorrelated, namely they do not influence each other. Therefore, measurement points outside the effective range are not required to perform inference using DGPR. Thanks to this property, Gaussian process regression performed using CS kernel functions lends itself to be employed in distributed estimation applications. At the same time, however, the lack of infinite support of the kernel function determines an estimation error when compared to a centralized approach. Quantifying and analyzing this error, with the objective of classifying data to be exchanged between neighboring sensor nodes, is the subject of next section.

3. DISTRIBUTED GAUSSIAN PROCESS REGRESSION ERROR ANALYSIS

The sparsification of the covariance matrix of a GP obtained by using CS kernel functions is exploited to formulate a distributed version of GPR in Gu and Hu (2012). The algorithm proposed by the authors relies on communication between neighboring sensors in order to exchange collected measurements. If the estimation task takes place over long time horizons, or if the mobile sensor networks employed to perform it have energy constraints, reducing the communication burden is an important factor for the successful execution of the estimation task. This is because, as discussed in the Introduction, after mobility, communication is the most energy-consuming task in many mobile sensing applications (Pottie and Kaiser (2000)).

With the objective of deploying mobile sensor networks over long time horizons, in this section, a way of ranking the data to be transferred is presented, which can be used to select only the most significant data that will be transferred between neighboring sensor nodes. In Section 3.1 we obtain bounds on the error that is introduced by approximating a kernel function using a CS version of it, whereas in Section 3.2 the estimation difference due to the fact that each node of the sensor network has a different subset of the measurement data is estimated.

To this end, consider a mobile sensor network with N sensors deployed in a 2D environment $X \subseteq \mathbb{R}^2$ in order to estimate an environment field $f : D \to \mathbb{R}$, with $X \subseteq D \subseteq \mathbb{R}^2$, such as temperature, light intensity, and concentration of a chemical substance. We denote by $x_i \in X$ the position of a sensor nodes and by $y_i \in \mathbb{R}$ the observation made by the sensor at position x_i. Note that the subscript i does not refer to a specific sensor node; instead, x_i and x_j denote just two different points in X where two measurements have been performed. Moreover, we denote by k the Gaussian RBF kernel and by $k(x_i, x_j)$ its value computed as in (2). Similarly, $\hat{k}$ indicates the CS version of k, whose value $\hat{k}(x_i, x_j)$ is computed as in (4).

3.1 Approximation Using Compactly Supported Kernel Functions

Let $\{(x_m, y_m)\}_{m \in \{1,\ldots,M\}}$ be the set of measurements collected by a mobile sensor of the WSN: $\{y_m\}_{m \in \{1,\ldots,M\}}$ are the measured values at locations $\{x_m\}_{m \in \{1,\ldots,M\}}$. We want to analyze the effect of using a CS kernel function instead of a kernel with infinite support. Given a point $x^\star \in X$ and letting $x = [x_1, \ldots, x_M]$ and $y = [y_1, \ldots, y_M] = [f(x_1), \ldots, f(x_M)]$, we denote by $y^\star$ and $\hat{y}^\star$ the means of the conditional probabilities $f(x^\star) \mid y$ obtained using the kernel functions k and $\hat{k}$, respectively.

We now aim at finding a relationship between the estimation difference $|y^\star - \hat{y}^\star|$ and the difference between the kernel functions. To this end, using (1), one can write: $|y^\star - \hat{y}^\star| = \left| k(x^\star, x)^\intercal k(x,x)^{-1} y - \hat{k}(x^\star, x)^\intercal \hat{k}(x,x)^{-1} y \right|$. For sake of notational compactness, we use the following conventions: $K_{xx} = k(x,x)$, $K_{x^\star x} = k(x^\star, x)$, $\hat{K}_{xx} = \hat{k}(x,x)$, $\hat{K}_{x^\star x} = \hat{k}(x^\star, x)$. Using the definitions of vector and matrix l^2-norms, and the fact that $0 < k(x_1, x_2) \leq 1 \ \forall x_1, x_2 \in D$, one can show that $|y^\star - \hat{y}^\star| \leq 2 \left\|K_{xx}^{-1}\right\| \left\|\hat{K}_{xx}^{-1}\right\| \|y\| M^{\frac{3}{2}} \delta$, where $\delta = \sup_{x_i, x_j \in D} \left| k(x_i, x_j) - \hat{k}(x_i, x_j) \right|$.

If the measurements $\{(x_m, y_m)\}_{m \in \{1,\ldots,M\}}$ are linearly independent, the positive definite covariance matrix $k(x,x)$ is non-singular (Ababou et al. (1994)), which, in this case, is equivalent to the fact that its minimum eigenvalue $\lambda_{\min}(k(x,x))$ is strictly positive. Hence, using properties of symmetric and positive definite matrices, one obtains:

$$|y^\star - \hat{y}^\star| \leq \frac{2}{\lambda_{\min}(K_{xx}) \lambda_{\min}(\hat{K}_{xx})} \|y\| M^{\frac{3}{2}} \delta < \infty. \tag{5}$$

Thus $|y^\star - \hat{y}^\star| \to 0$ as $\delta \to 0$. This means that the distributed estimation performed by the mobile sensor nodes of a WSN using CS kernel functions is close to the centralized estimation obtained by using a kernel function with infinite support as long as their effective range is large. This concept is formalized in the following.

Since k and $\hat{k}$ are radial basis functions, we can define the following two functions that depend only on the quantity $r = \|x_i - x_j\| \ \forall x_i, x_j \in D$:

$$\mathcal{K}(r) = e^{-\frac{r^2}{\sigma^2}}, \quad \hat{\mathcal{K}}(r) = e^{-\frac{r^2}{\sigma^2}} \max \left\{ 0, \left(1 - \frac{r}{l}\right)^{\nu} \right\}. \tag{6}$$

$\mathcal{K}$ belongs to the space of continuous functions vanishing at infinity $C_0(\mathbb{R}) = \left\{ f \in C^0(\mathbb{R}) : \lim_{x \to \pm\infty} = 0 \right\}$, whereas $\hat{\mathcal{K}}$ belongs to the space of compactly supported continuous

$C_c(\mathbb{R}) = \{f \in C^0(\mathbb{R}) : f \text{ has compact support}\}$. $C_c(\mathbb{R})$ is a dense proper subspace of $C_0(\mathbb{R})$ with respect to the uniform norm $\|f\|_u = \sup_{x\in\mathbb{R}} |f(x)|$, as shown in Heil (2010). Therefore, $\forall f \in C_0(\mathbb{R})$ and $\forall \varepsilon > 0$, there exists a sequence of functions $\{f_n\}_{n\in\mathbb{N}} \in C_c(\mathbb{R})$ and $N > 0$ such that, $\forall n > N$, $\|f_n - f\|_u < \varepsilon$. This means that any function $f \in C_0(\mathbb{R})$ can be approximated with arbitrary accuracy using compactly supported functions $f_n \in C_c(\mathbb{R})$. The approximation of functions obtained by using series of CS functions can be exploited in the context of DGPR as explained in the following.

Taking the sequence of compactly supported functions

$$\hat{\mathcal{K}}_n(r) = \left\{ e^{-\frac{r^2}{\sigma^2}} \max\left\{0, \left(1-\frac{r}{n}\right)^{\nu}\right\}\right\}_{n\in\mathbb{N}}, \tag{7}$$

we have that, as $n \to \infty$, $\delta = \|\hat{\mathcal{K}}_n - \mathcal{K}\|_u \to 0$. Therefore, in order to minimize the estimation error $|y^\star - \hat{y}^\star|$ due to the use of CS kernel functions, the objective of the sensor nodes of a WSN is that of maximizing the effective range l. Note that for a sensor node, with M collected measurements denoted by $\{(x_m, y_m)\}_{m\in\{1,\dots,M\}}$, the effective range l is bounded by $\max_{r,s\in\{1,\dots,M\}} \|x_r - x_s\|$. In conclusion, for a spatial estimation task, this result means that it is desirable to have measurement locations that are as far apart as possible in space. This way, the effective range, which is a parameter of the CS kernel used for the GPR, can be increased and, consequently, the estimation error will be reduced. The following proposition summarizes what has been derived so far.

Proposition 1. *Let $\mathcal{GP}_1$ be a GPR model that employs the infinitely supported Gaussian RBF kernel function in (2). Define $\mathcal{GP}_2$ as the GPR model that uses as kernel function the compactly supported version of (2), given by (4). Provided that the two models, $\mathcal{GP}_1$ and $\mathcal{GP}_2$, are built using the same dataset $\{(x_m, y_m)\}_{m=1,\dots,M}$, the estimation difference $|y^\star - \hat{y}^\star|$ at a point $x^\star$ is linearly bounded by $\|\hat{\mathcal{K}}_n - \mathcal{K}\|_u$, where $\mathcal{K}$ and $\hat{\mathcal{K}}_n$ are defined as in (6) and (7), respectively.*

The following Corollary shows how DGPR which employs CS kernels generalize to full GPR when the effective range l of the kernel goes to infinity.

Corollary 2. *Under the same conditions of Proposition 1, the estimation difference $|y^\star - \hat{y}^\star| \to 0$ as the effective range $l \to \infty$.*

These results will be employed in Section 4 to develop an algorithm used by the nodes of the sensor network in order to select the data that need to be transferred to the neighboring nodes, given a maximum communication bandwidth.

3.2 Approximation Using Different Sets of Data

The difference between a centralized and a distributed approach for spatial field estimations can be interpreted in terms of different sets of measurements as follows. If all sensor nodes transferred all the data they have collected to a central unit, the entire set of measurements would be available to a single computational unit that would be able to perform a full GPR. In a distributed framework, instead, each sensor node can be seen as a computational unit that has available only a subset of the entire set of measurements. Therefore, we quantify the error introduced by only having available a subset of the measurement data. In order to do that, we proceed as follows. Using the same notation adopted in the previous subsection, we let $\{(x_m, y_m)\}_{m\in\{1,\dots,M\}}$ be the measurement data available to a sensor node. We assume that an additional measurement (x_{M+1}, y_{M+1}) becomes available, and we define $x = [x_1, \dots, x_{M+1}]$, $y = [y_1, \dots, y_{M+1}]$, $\tilde{x} = [x_1, \dots, x_M]$ and $\tilde{y} = [y_1, \dots, y_M]$. We aim at quantifying the difference $|y^\star - \tilde{y}^\star|$ between the estimations at a given point $x^\star \in X$ obtained incorporating or not, respectively, the new measurement. More specifically, we want to find an upper bound for $|y^\star - \tilde{y}^\star|$ with the objective of transferring only those measurements that might lead to a significant change in the estimation.

Adopting the same notational shortcuts introduced in the previous subsection, and denoting by χ the location x_{M+1}, we can proceed as follows.

Formally defining $\bar{K}^{-1}_{\tilde{x}\tilde{x}} = \begin{bmatrix} K^{-1}_{\tilde{x}\tilde{x}} & 0 \\ 0 & 0 \end{bmatrix}$, one has that

$$|y^\star - \tilde{y}^\star| \le \|K^{\mathsf{T}}_{x^\star x}\| \, \|y\| \, \underbrace{\|K^{-1}_{xx} - \bar{K}^{-1}_{\tilde{x}\tilde{x}}\|}_{\Delta K}, \tag{8}$$

where the quantity ΔK can be further simplified as follows:

$$\|K^{-1}_{xx} - \bar{K}^{-1}_{\tilde{x}\tilde{x}}\| = \left\| \begin{bmatrix} \mathcal{X} & \mathcal{Y} \\ \mathcal{Y}^{\mathsf{T}} & \mathcal{Z} \end{bmatrix} \right\| \le \max\{\|\mathcal{X}\|, \|\mathcal{Z}\|\} + \|\mathcal{Y}\| \tag{9}$$

where

$$\begin{aligned} \mathcal{X} &= K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi} \left(K_{\chi\chi} - K^{\mathsf{T}}_{\tilde{x}\chi} K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi}\right)^{-1} K^{\mathsf{T}}_{\tilde{x}\chi} K^{-1}_{\tilde{x}\tilde{x}} \\ \mathcal{Y} &= -K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi} \left(K_{\chi\chi} - K^{\mathsf{T}}_{\tilde{x}\chi} K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi}\right)^{-1} \\ \mathcal{Z} &= \left(K_{\chi\chi} - K^{\mathsf{T}}_{\tilde{x}\chi} K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi}\right)^{-1}, \end{aligned} \tag{10}$$

and their norms satisfy

$$\begin{aligned} \|\mathcal{X}\| &\le \|K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi}\|^2 \|\mathcal{Z}\| \\ \|\mathcal{Y}\| &\le \|K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi}\| \|\mathcal{Z}\| \\ \|\mathcal{Z}\| &= \left| \left(K_{\chi\chi} - K^{\mathsf{T}}_{\tilde{x}\chi} K^{-1}_{\tilde{x}\tilde{x}} K_{\tilde{x}\chi}\right)^{-1} \right|. \end{aligned} \tag{11}$$

Then, using the Woodbury and the Shermann-Morrison formulas, the following upper bound for $\|\mathcal{Z}\|$ can be obtained:

$$\|\mathcal{Z}\| \le \frac{\left|\frac{1}{K_{\chi\chi}}\right| + \left(\frac{\|K_{\tilde{x}\chi}\|}{K_{\chi\chi}}\right)^2 \|K^{-1}_{\tilde{x}\tilde{x}}\|}{1 - \left(\frac{\|K_{\tilde{x}\chi}\|}{K_{\chi\chi}} \|K^{-1}_{\tilde{x}\tilde{x}}\|\right)^2}. \tag{12}$$

In the next section, expression (12) will be used in order to decide whether the datapoint (x_{M+1}, y_{M+1}) should be transferred between neighboring sensor nodes or not.

The result obtained in this section is summarized by the following proposition.

Proposition 3. *Let $\mathcal{GP}_1$ and $\mathcal{GP}_2$ be two GPR models built using the infinitely supported Gaussian RBF kernel function (2). Let $\{(x_m, y_m)\}_{m=1,\dots,M+1}$ be a set of measurement data. Then, the estimation difference $|y^\star - \tilde{y}^\star|$ at a point $x^\star$, defined in (8), is bounded by a monotone increasing function of $\|K_{\tilde{x}\chi}\|$.*

Proof. From (8), $|y^\star - \tilde{y}^\star| \le \alpha\Delta K$, $\alpha \in \mathbb{R}$, $\alpha > 0$. Moreover, from (9), $\Delta K \le \max\{\|\mathcal{X}\|, \|\mathcal{Z}\|\} + \|\mathcal{Y}\|$. As derived above, the functions that bound their norms are all monotone increasing functions of $\|K_{\tilde{x}\chi}\|$. Hence, $\max\{\|\mathcal{X}\|, \|\mathcal{Z}\|\} + \|\mathcal{Y}\|$ is a monotone increasing function of $\|K_{\tilde{x}\chi}\|$, from which the result follows. □

The bound introduced in Proposition 3 allows us to estimate the difference $|y^\star - \tilde{y}^\star|$ by using the scalar quantity $\|K_{\tilde{x}\chi}\|$, without the need of computing the estimate $y^\star$ using the entire data set x. This result, together with the one of Proposition 1, will be leveraged in the next section to define two algorithms required to implement the communication constrained DGPR proposed in this paper.

4. COMMUNICATION CONSTRAINED DISTRIBUTED GAUSSIAN PROCESS REGRESSION

4.1 Maximization of the Effective Range

As stated in Corollary 2, in order to reduce the error due to the use of CS kernel functions, the effective ranges of the sensor nodes have to be increased. From now on, we will need to differentiate between two sensor nodes in a WSN, the receiver and the sender: we use superscripts i and j to refer to the former and the latter, respectively.

Let $\{(x_m^{(i)}, y_m^{(i)})\}_{m\in\{1,\ldots,M\}}$ be the M measurement data points stored by the receiver node i. Because of the previous argument, when exchanging data with its neighbors, it is desirable that a measurement data point (x_{M+1}, y_{M+1}) is received and incorporated if

$$\max_{m\in\{1,\ldots,M\}} \|x_m^{(i)} - x_{M+1}\| > l_i, \tag{13}$$

l_i being the effective range of node i.

Let $j \in N_i$, where $N_i \subset \{1, \ldots, N\}$ is the index set of the neighbors of node i. We require that the sender node j transfers a data point to node i only if (13) holds. Therefore, node j has to know the locations $\{x_m^{(i)}\}_{m\in\{1,\ldots,M\}}$ of the measurement data points of node i. In order for node i to communicate to node j the locations of its data points, these can be compressed by computing the minimum volume ellipsoid that encloses all the data points $\{x_m^{(i)}\}_{m\in\{1,\ldots,M\}}$. This can be done efficiently as shown, for instance, in Gärtner and Schönherr (1997). Moreover, Gärtner and Schönherr (1997) show that scaling the minimum area enclosing ellipse about its center of a factor $1/d$, d being the dimension of the measurement data points, results in an ellipse that is completely inside the convex hull of the data points. In $\mathbb{R}^2$, denoting by $A_e^{(i)} \in \mathbb{R}^{2\times 2}$ the matrix encoding length and directions of the axes of the minimum-area ellipse corresponding to the measurement data points of sensor node i, and by $c_e^{(i)} \in \mathbb{R}^2$ the center of the ellipse, we define

$$\bar{x}_e = \max_{(x-c_e^{(i)})^\intercal A_e^{(i)}(x-c_e^{(i)})=1} \|x - x_{M+1}\|, \tag{14}$$

and

$$\bar{x} = \max_{x\in\{x_m^{(i)}\}_{m\in\{1,\ldots,M\}}} \|x - x_{M+1}\|. \tag{15}$$

Then, we can quantify the accuracy in the approximation of the measurement data points by means of the minimum area enclosing ellipse as follows (Nering and Tucker (1992)):

$$\big|\, \|\bar{x}_e - x_{M+1}\| - \|\bar{x} - x_{M+1}\| \,\big| \leq d^{-1}(d-1)\sqrt{\lambda_{\max}(A_e^{(i)})},$$

where $\lambda_{\max}(A_e^{(i)})$ is the maximum eigenvalue of $A_e^{(i)}$.

After receiving $A_e^{(i)}$, $c_e^{(i)}$ and l_i from node i, node j can decide whether to transfer its data to node i or not. Moreover, in case it needs to transfer data, given a maximum amount of data that can be transferred, it can rank the data to be transferred according to the bound (13) on the resulting l_i, as described in Algorithm 1.

Algorithm 1 Selection of data to transfer — Part 1

```
Require: datasets $D_i$, $D_j$ of neighboring nodes $i$, $j$
Ensure: sorted measurement data of sensor $j$
  procedure NODE $i$
      $[A_e^{(i)}, c_e^{(i)}] \leftarrow$ minimumAreaEnclosingEllipse($D_i$)
      transfer $[A_e^{(i)}, c_e^{(i)}]$ to node $j$
  end procedure
  procedure NODE $j$
      receive $[A_e^{(i)}, c_e^{(i)}]$ from node $i$
      $d = [\ ]$
      for $d_j$ in $D_j$ do
          $d_{\max} \leftarrow$ maximumDistance$\Big(d_j, [A_e^{(i)}, c_e^{(i)}]\Big)$
          $d \leftarrow$ append($d_{\max}$)
      end for
      sort($D_j$)                                  ▷ according to $d$
  end procedure
```

4.2 Maximization of the Novelty of Measurement Data Points

In Section 3.2 a bound for $|y^\star - \tilde{y}^\star|$ has been derived. In the following, we briefly recall the notion of feature space in the context of GPs, which will be used to formulate an algorithm to select the measurement data that each sensor node has to transfer to its neighbors according to the bound (12).

In Section 2 we presented what is known as the function-space view of a GP. In the feature-space view of a GP, the function f to be estimated is expressed as $f = \phi(x)^\intercal w$, where $\phi : D \subseteq \mathbb{R}^d \to D' \subseteq \mathbb{R}^n$ maps the inputs x to an n-dimensional, $n \leq \infty$, inner product space, the feature space. The variable w denotes a vector of weights to be estimated. In this framework, one can show that the the covariance function can be expressed as the inner product $k(x_1, x_2) = \langle \phi(x_1), \phi(x_2)\rangle_{\mathbb{R}^n}$ in $\mathbb{R}^n$ (Rasmussen and Williams (2006)). The matrix $K_{\tilde{x}\chi} = k(\tilde{x}, x_{M+1})$ in (12) can be then expressed as $K_{\tilde{x}\chi} = \langle \phi(\tilde{x}), \phi(x_{M+1})\rangle_{\mathbb{R}^n}$. Since, according to Proposition 3 the function on the right-hand side of (12) is a monotone increasing function of $\|K_{\tilde{x}\chi}\|$, using the Cauchy-Schwarz inequality leads to:

$$\|\mathcal{Z}\| \leq \frac{\left|\frac{1}{K_{\chi\chi}}\right| + \frac{\|K_{\tilde{x}\tilde{x}}^{-1}\|}{K_{\chi\chi}^2}\|\phi(\tilde{x})\|^2\|\phi(x_{M+1})\|^2}{1 - \left(\frac{\|K_{\tilde{x}\tilde{x}}^{-1}\|}{K_{\chi\chi}}\right)^2 \|\phi(\tilde{x})\|^2\|\phi(x_{M+1})\|^2}. \tag{16}$$

Now, as in the case of different kernels described in Section 4.1, we want the sender node j, neighbor of the receiver node i, to transfer only data that significantly influence the prediction $\tilde{y}^\star$. For this reason, node j would have to know the data points that node i has. However, as in the previous section, we do not want node i to transfer all its data points to node j. In view of what has been derived in (16), if node i transfers the values $\|\bar{K}_{\tilde{x}\tilde{x}}^{-1}\|$, $\|\phi(\tilde{x})\|^2 \in \mathbb{R}$, node j can evaluate what are its measurement data points that can more significantly influence the prediction $\tilde{y}^\star$of node i, and transfer them only. Alternatively, as before, if there is a maximum number of data points that can be transferred, node j can rank its data according to the influence that they can have on the prediction $\tilde{y}^\star$. The ranking-based transferring strategy described in this section is summarized in Algorithm 2.

Algorithm 2 Selection of data to transfer — Part 2

```
Require: datasets $D_i$, $D_j$ of neighboring nodes $i$, $j$
Ensure: sorted measurement data of sensor $j$
  procedure NODE $i$
      transfer $\left[\|\phi(\tilde{x})\|, \|K_{\tilde{x}\tilde{x}}^{-1}\|\right]$ to node $j$
  end procedure
  procedure NODE $j$
      receive $\|\phi(\tilde{x})\|$, $\|K_{\tilde{x}\tilde{x}}^{-1}\|$ from node $i$
      $d = [\ ]$
      for $d_j$ in $D_j$ do
          $b \leftarrow$ evaluateBound($d_j$,$\|\phi(\tilde{x})\|$, $\|K_{\tilde{x}\tilde{x}}^{-1}\|$)    ▷ (8), (9), (16)
          $d \leftarrow$ append($b$)
      end for
      sort($D_j$)    ▷ according to $d$
  end procedure
```

4.3 Information-Entropy-Based Sensor Motion Control

Following the approach in Gu and Hu (2012), we employ an area coverage control algorithm to move the mobile sensors in the environment in which they are deployed. The approach presented by Cortes et al. (2004) lends itself to accomplish this objective. The locational cost

$$\mathcal{H}_i(x_i) = \int_X \|x_i - q\|^2 \varphi_i(q)\mathrm{d}q \tag{17}$$

defined for each sensor node i, $i = 1, \dots, N$, can be minimized moving towards the weighted centroid ρ_i of X (see Cortes et al. (2004)). As done in Gu and Hu (2012), the weighting function $\varphi_i(q)$ is set to

$$\varphi_i(q) = \log\det(k(q,q) - k(q,x^{(i)})^\mathsf{T} k(x^{(i)},x^{(i)})^{-1} k(q,x^{(i)})), \tag{18}$$

where $x^{(i)}$ is the data collected by sensor node i. Assuming that it is possible to directly control the velocity of each sensor node, $\dot{x}_i$, the decentralized motion control law

$$\dot{x}_i = \gamma(\rho_i - x_i), \quad i = 1, \dots, N \tag{19}$$

$\gamma > 0$ being a control gain, minimizes the locational cost (17), and lets each node visit regions of the environment where the variance of its estimation is higher. This allows it to collect more data in those regions, which, in turn, has the effect of reducing the variance of its estimation. In Gu and Hu (2012), this strategy is shown to minimize the information entropy of the Gaussian random variable representing the spatial field f to estimate, conditioned on the observations taken by each sensor node in the environment. The combination of motion and communication strategies described in Sections 4.1 and 4.2 is described in Algorithm 3, executed by each sensor node i.

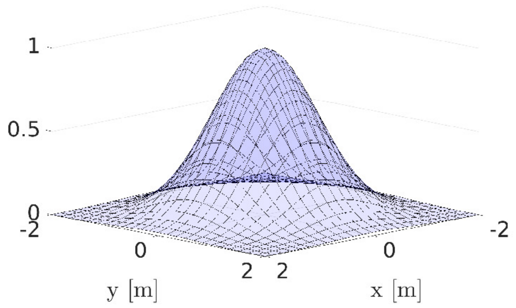

Fig. 1. Surf plot of an example of environment field simulated over the Robotarium testbed, which has to be estimated by the network of mobile sensors.

Algorithm 3 Communication constrained DGPR

```
Require: nodes' positions $x_i$, datasets $D_i$, control gain $\gamma > 0$, maximum number of data points $N_{\max}$
Ensure: communication constrained DGPR
  for $i$ in $\{1, \dots, N\}$ do
      compute $\rho_i$    ▷ Cortes et al. (2004)
      $\dot{x}_i \leftarrow \gamma(\rho_i - x_i)$
      move with velocity $\dot{x}_i$
      for sensor node $j$ neighbor of sensor node $i$ do
          $D_i \leftarrow$ sorted data    ▷ Algorithms 1 and 2
          transfer first $N_{\max}$ data points from $D_i$ to sensor node $j$
      end for
  end for
```

Algorithms 1, 2 and 3 allow a wireless sensor network to perform distributed estimation of a spatial field by exchanging only a limited amount of data between each other at each point in time. This approach lends itself to be employed in long-term distributed estimation applications, where communication requires a non-trivial amount of energy.

5. EXPERIMENTAL RESULTS

The communication constrained DGPR algorithm developed in the previous section has been deployed on a team of 16 ground mobile robots in the Robotarium (Pickem et al. (2017)), a remotely accessible swarm robotics testbed. Here, environment fields consisting in mixtures of Gaussian surfaces (such as the one depicted in Fig. 1) have been simulated, together with the sensor measurements collected by the robots. By varying the environment field to estimate, as well as the initial positions of the robots in the environment, several experiments have been performed. In the following, the results in terms of root mean square (RMS) error are compared to a centralized estimation, in which all robots are able to communicate all collected data to a centralized computational unit.

Figure 2 shows how the RMS error changes over time during the course of one of the experiments performed in the Robotarium. Setting the maximum communication bandwidth—expressed in terms of maximum number of

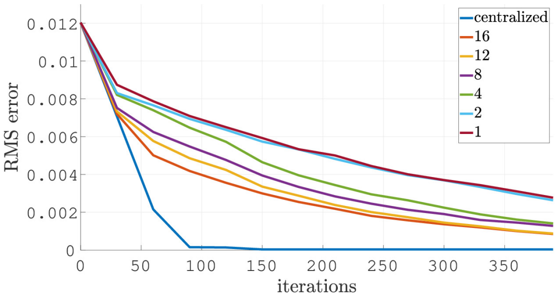

Fig. 2. Plot of RMS error over time for one of the experiments performed in the Robotarium. The different curves show how the RMS error decreases over time as the sensor nodes exchange data between each other, as a function of the maximum number of data points exchanged at each point in time by any two sensor nodes (see legend). The blue curve at the bottom represents the centralized approach where data collected by all robots are gathered by a central computational unit, which is able to perform full GPR. As can be seen, the higher the communication bandwidth—in terms of number of data points exchanged—the faster the decrease of the RMS error towards the centralized lower bound.

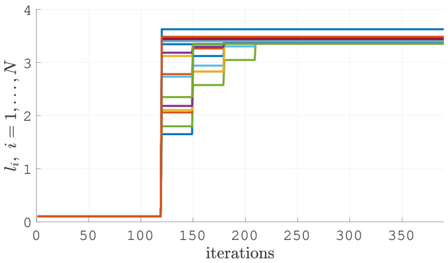

Fig. 3. Effective range l_i of the 16 ground mobile robots recorded over the course of one of the experiment conducted on the Robotarium. The sensor nodes start communicating between each other once enough data has been collected (around 120 iterations). The graphs show how the selection of exchanged data according to Algorithms 1 and 2 allows the sensor nodes to quickly increase their effective range and correspondingly decrease the RMS estimation error (cf. Fig. 1).

data points exchanged between any two neighboring sensor nodes—6 different experiments have been performed for the same environment field. The results show how increasing the communication bandwidth leads to a faster convergence of the estimation to the lower bound defined by the centralized GPR (blue line in Fig. 2). For slowly-varying environment fields that take place over long time horizons, it is thus worthwhile saving energy by decreasing the maximum number of data points that can be communicated between neighbors, at the cost of decreasing the speed of the estimation convergence.

In Fig. 3, the effective range l_i is depicted for 16 ground mobile robots employed in one of the field estimation experiments. At iteration 120, the sensor nodes have collected enough data so they can start exchanging them with the neighboring nodes. The graphs show how Algorithms 1 and 2 are effective at selecting the data points to be exchanged in order to increase l_i for each robot and consequently decrease the RMS estimation error.

Finally, Figures 4a to 4f show snapshots of the video of one of the experiments performed in the Robotarium. The environment field to be estimated is overlaid in blue on the testbed, whereas the estimation of the field performed by one of the sensor nodes is shown in orange. Following Algorithm 3, the sensor nodes (ground mobile robots) move in the environment and exchange data to increase their effective range (yellow circle). As can be seen, during the course of the experiment, the estimated field (orange) approaches the ground truth (blue) as the node collects and exchanges data with its neighbors. This way, an accurate estimate of the environment field is obtained using a distributed and communication constrained algorithm.

6. CONCLUSIONS

In this paper we proposed a solution to communication constrained distributed Gaussian process regression. The main objective is that of enabling long-term deployment of mobile wireless sensor networks for spatial field estimation. Since severe limitations on the battery life of sensor nodes are due to communication, we addressed the problem of estimating spatial fields in a distributed manner while explicitly imposing communication constraints. The proposed approach is based on the derivation of theoretical bounds on the estimation error introduced by distributed algorithms. Given a maximum communication bandwidth, we proposed an algorithm to select the most significant data to be transferred. The performance of this algorithm are demonstrated, both in terms of estimation accuracy and amount of data transferred, on an team of mobile robots.

REFERENCES

Ababou, R., Bagtzoglou, A.C., and Wood, E.F. (1994). On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. *Mathematical Geology*, 26(1), 99–133.

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., and Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer networks*, 38(4), 393–422.

Chen, J., Low, K.H., Yao, Y., and Jaillet, P. (2015). Gaussian process decentralized data fusion and active sensing for spatiotemporal traffic modeling and prediction in mobility-on-demand systems. *IEEE Transactions on Automation Science and Engineering*, 12(3), 901–921.

Cortes, J., Martinez, S., Karatas, T., and Bullo, F. (2004). Coverage control for mobile sensing networks. *IEEE Transactions on robotics and Automation*, 20(2), 243–255.

Das, S., Roy, S., and Sambasivan, R. (2015). Fast gaussian process regression for big data. *arXiv preprint arXiv:1509.05142*.

Furrer, R., Genton, M.G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3), 502–523.

Gärtner, B. and Schönherr, S. (1997). Smallest enclosing ellipses–fast and exact.

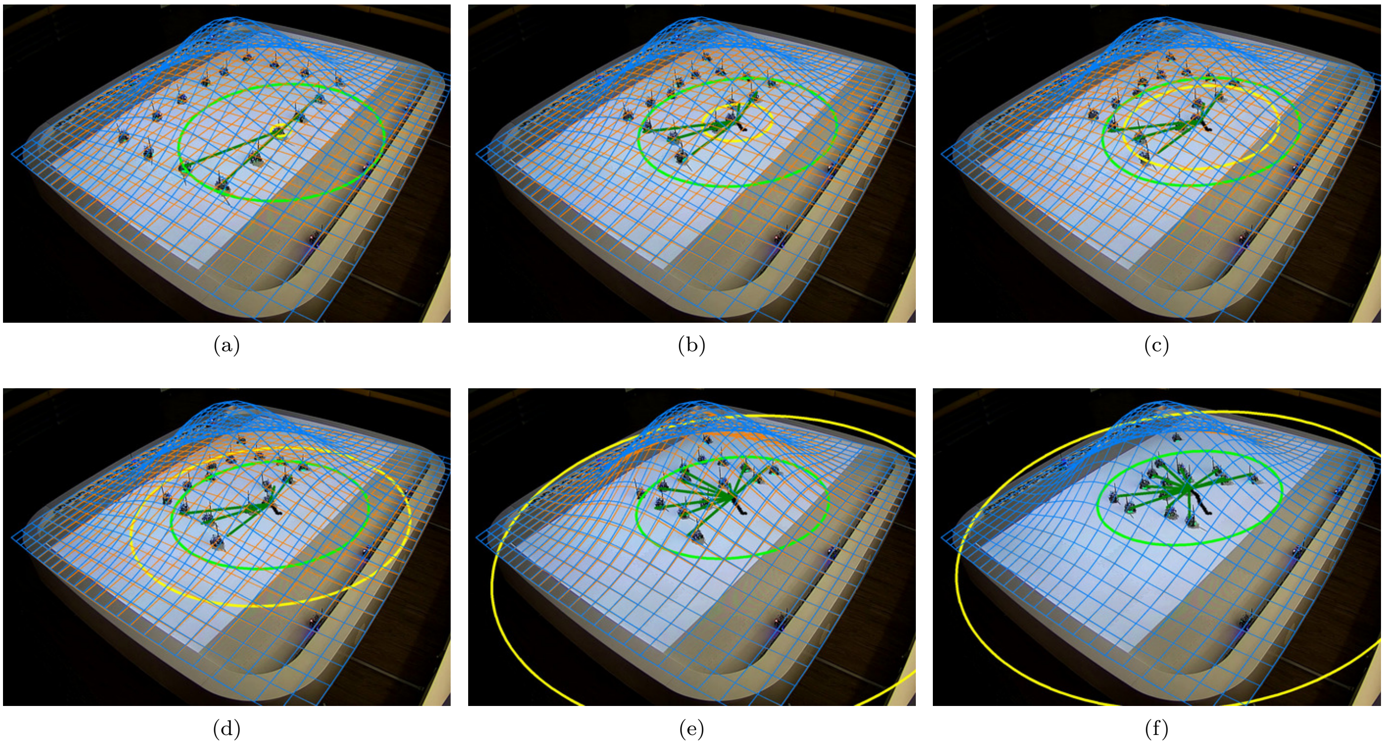

Fig. 4. Snapshots from the video of the experiments performed on the Robotarium. The estimation of a simulated environment field (blue surf plot) performed by one of the 16 mobile robots employed in the experiment is depicted. Its trajectory is shown as a thick black line, its communication and effective ranges are shown in green and yellow, respectively. As can be seen, during the course of the experiment, the error between the true environment field and its estimate (orange surf plot) decreases, as the effective range increases. The video of the experiment is available online at the link `https://youtu.be/6vTcnh4wsZU`.

Genton, M.G. (2001). Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research*, 2(Dec), 299–312.

Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83(2), 493–508.

Gu, D. and Hu, H. (2012). Spatial gaussian process regression with mobile sensor networks. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8), 1279–1290.

Hamers, B., Suykens, J.A., and De Moor, B. (2002). Compactly supported rbf kernels for sparsifying the gram matrix in ls-svm regression models. In *International Conference on Artificial Neural Networks*, 720–726. Springer.

Heil, C. (2010). *A basis theory primer: expanded edition.* Springer Science & Business Media.

Jamshidi, A.A. and Kirby, M.J. (2006). Examples of compactly supported functions for radial basis approximations. In *MLMTA*, 155–160.

Kivinen, J., Smola, A.J., and Williamson, R.C. (2004). Online learning with kernels. *IEEE transactions on signal processing*, 52(8), 2165–2176.

Nering, E.D. and Tucker, A.W. (1992). *Linear Programs & Related Problems: A Volume in the Computer Science and Scientific Computing Series.* Elsevier.

Pickem, D., Glotfelter, P., Wang, L., Mote, M., Ames, A., Feron, E., and Egerstedt, M. (2017). The robotarium: A remotely accessible swarm robotics research testbed. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 1699–1706. IEEE.

Pottie, G.J. and Kaiser, W.J. (2000). Wireless integrated network sensors. *Communications of the ACM*, 43(5), 51–58.

Quinonero-Candela, J., Rasmussen, C.E., and Williams, C.K. (2007). Approximation methods for gaussian process regression. *Large-scale kernel machines*, 203–224.

Rasmussen, C.E. and Williams, C.K. (2006). *Gaussian processes for machine learning*, volume 1. MIT press Cambridge.

Shi, J.Q., Murray-Smith, R., Titterington, D.M., and Pearlmutter, B.A. (2005). Filtered gaussian processes for learning with large data-sets. In *Switching and Learning in Feedback Systems*, 128–139. Springer.

Steinwart, I. (2001). On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov), 67–93.

Tavassolipour, M., Motahari, S.A., and Shalmani, M.T.M. (2017). Learning of gaussian processes in distributed and communication limited systems. *arXiv preprint arXiv:1705.02627.*

Wilson, A.G., Dann, C., and Nickisch, H. (2015). Thoughts on massively scalable gaussian processes. *arXiv preprint arXiv:1511.01870.*

Wu, Z. (1995). Compactly supported positive definite radial functions. *Advances in Computational Mathematics*, 4(1), 283–292.