Asymptotic optimality of multicenter Voronoi configurations for random field estimation

Rishi Graham Jorge Cortés

Abstract

This paper deals with multi-agent networks performing optimal estimation tasks. Consider a network of mobile agents with sensors that can take measurements of a spatial process in an environment of interest. Using the measurements, one can construct a kriging interpolation of the spatial field over the whole environment, with an associated prediction error at each point. We study the continuity properties of the prediction error, and consider as global objective functions the maximum prediction error and the generalized prediction variance. We study the network configurations that give rise to optimal field interpolations. Specifically, we show how, as the correlation between any two different locations vanishes, circumcenter and incenter Voronoi configurations become network configurations that optimize the maximum prediction error and the generalized prediction variance, respectively. The technical approach draws on tools from geostatistics, computational geometry, linear algebra, and dynamical systems.

I. Introduction

Problem statement: Mobile sensor networks are envisioned to perform distributed sensing and data fusion tasks in a wide range of scenarios, including environmental monitoring, oceanographic research, and distributed surveillance of critical infrastructures. This paper considers mobile sensor networks performing optimal estimation of physical processes modeled as spatial random fields. Standard interpolation techniques produce estimates of the spatial field at each point of the environment of interest. When a measure of the accuracy of the estimate is available, a natural objective is then to characterize those network configurations that give rise to optimal estimates of the field. This is the problem that we consider in this paper.

This version: July 17, 2007

Rishi Graham and Jorge Cortés are with the Department of Applied Mathematics and Statistics, School of Engineering, University of California, Santa Cruz, CA 95064, USA, {rishig,jcortes}@soe.ucsc.edu

Literature review: Kriging [1], [2] is a standard technique used in geostatistics to produce estimates of spatial processes based on data collected at a finite number of locations. The main advantage of kriging over other spatial interpolation methods is that it provides a measure of the uncertainty associated to the estimator. The optimal design literature [3], [4] deals with the problem of designing experiments to optimize the resulting statistical estimation. Of particular interest are the notions of G-optimality, minimizing the maximum prediction error, and D-optimality, minimizing the generalized prediction variance.

The work [5] introduces performance metrics for optimal estimation in oceanographic research. The works [6], [7] proposes distributed optimal estimation strategies for deterministic fields, when the measurements taken by individual agents are uncorrelated. In [8], the emphasis is on finding optimal agent trajectories along a given interval of time among a parameterized set of trajectories. Here, instead, we focus on optimal network configurations for the estimation of the random field at a single snapshot. In our technical approach, we have been inspired by [9], which considers the problem of minimizing the maximum uncertainty over a discrete space and shows that minimax configurations are asymptotically optimal as the correlation between any two distinct points vanishes. Minimax configurations minimize the maximum distance to the nearest agent from any point in space. Here, we make the connection to Voronoi partitions of continuous spaces, which are a classical notion in computational geometry [10]. In [11], circumcenter and incenter Voronoi configurations are defined, and various distributed motion coordination algorithms are introduced which are guaranteed to asymptotically bring the network to these desirable configurations.

Statement of contributions: In this paper, we consider two performance metrics for optimal placement of mobile sensor networks based on kriging. We first characterize the continuity properties of the mean-squared error of the simple kriging estimator as a function of the network configuration. In the case of zero measurement error, this is not trivial. Previous results in the optimal design literature have avoided this problem by optimizing over a discrete set of possible configurations, while we consider the continuous space of all agent locations within the region. Next, we define our first optimality criterion, the maximum prediction error of the kriging predictor, and study its critical points asymptotically, as the correlation between any two distinct points vanishes. We define a second optimality criterion as a form of D-optimality, the generalized variance of the kriging predictor within a bounded region, and study the critical points within the

same asymptotic framework. Our main results are showing that, for the simple kriging predictor, circumcenter Voronoi configurations are asymptotically optimal for the maximum prediction error over the environment, while incenter Voronoi configurations are asymptotically optimal for the generalized variance. In general, it is difficult to obtain exactly the configurations that optimize these objective functions. Our results are relevant to the extent that they guarantee that, for scenarios with small enough correlation between distinct points, circumcenter and incenter Voronoi configurations are optimal for appropriate measures of uncertainty. The network can achieve these desirable network configurations by executing simple distributed dynamical systems. Various simulations illustrate our results.

Organization: The paper is organized as follows. Section II introduces basic notions from computational geometry and presents a brief overview of the kriging estimation procedure. Section III states the problem of interest. We present our main results in Section IV on the optimality of circumcenter and incenter Voronoi configurations. Section V presents simulations to illustrate our results. Finally, Section VI gathers our conclusions and ideas for future work.

II. Preliminaries

This section introduces various concepts which will be useful later on. Let us start by introducing some notation for standard geometric objects. Let $\mathbb{R}$, $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geq 0}$ denote the set of reals, positive reals and nonnegative reals, respectively. We are concerned with operations on a compact and connected set $\mathcal{D}$ of Euclidean space $\mathbb{R}^d$, $d \in \mathbb{N}$. For $p, q \in \mathbb{R}^d$, we let $]p, q[= \{\lambda p + (1-\lambda)q \mid \lambda \in]0, 1[\}$ denote the *open segment* with extreme points p and q. For $p \in \mathbb{R}^d$ and $r \in \mathbb{R}_{>0}$, we let $\overline{B}(p, r)$ denote the *closed ball* of radius r centered at p and $B(p, r)$ denote the *open ball* of radius r centered at p. We denote by $|S|$ and ∂S the cardinality and the boundary of a set S, respectively. A *convex polytope* is the convex hull of a finite point set. For a bounded set $S \subset \mathbb{R}^d$, we let $\mathrm{CC}(S)$ and $\mathrm{CR}(S)$ denote the *circumcenter* and *circumradius* of S, respectively, that is, the center and radius of the smallest-radius d-sphere enclosing S. The incenter set of S, denoted by $\mathrm{IC}(S)$, is the set of the centers of maximum-radius d-spheres contained in S. The inradius of S, denoted by $\mathrm{IR}(S)$, is the common radius of these spheres.

We consider tuples or ordered sets of possibly coincident points, $P = (p_1, \ldots, p_n) \in (\mathbb{R}^d)^n$. We will refer to such an element as a *configuration*. Let $\mathfrak{P}(S)$ (respectively $\mathbb{F}(S)$) denote the collection of subsets (respectively, finite subsets) of S. We denote an element of $\mathbb{F}(\mathbb{R}^d)$ by

$\mathcal{P} = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$, where $p_1, \ldots, p_n$ are distinct points in $\mathbb{R}^d$. Let $i_{\mathbb{F}} : (\mathbb{R}^d)^n \to \mathbb{F}(\mathbb{R}^d)$ be the natural immersion, i.e., $i_{\mathbb{F}}(P)$ contains only the distinct points in $P = (p_1, \ldots, p_n)$. Note that the cardinality of $i_{\mathbb{F}}(p_1, \ldots, p_n)$ is in general less than or equal to n. Let S_{coinc} be the set of all tuples in $(\mathbb{R}^d)^n$ which contain at least one coincident pair of points, that is,

$$S_{\text{coinc}} = \{(p_1, \ldots, p_n) \in (\mathbb{R}^d)^n \,|\, p_i = p_j \text{ for some } i, j \in \{1, \ldots, n\}, i \neq j\}.$$

Let $\|\cdot\|$ denote the Euclidean distance function on $\mathbb{R}^d$. Define the distance $\mathrm{d} : \mathbb{R}^d \times \mathfrak{P}(\mathcal{D}) \to \mathbb{R}$ from a point in $\mathbb{R}^d$ to a set of points in $\mathcal{D}$ by $\mathrm{d}(s, \mathcal{P}) = \inf_{p \in \mathcal{P}} \{\|s - p\|\}$, and let $\mathrm{mds} : \mathbb{R}^d \times \mathfrak{P}(\mathcal{D}) \to \mathfrak{P}(\mathcal{D})$ be the *minimum distance set* map defined by $\mathrm{mds}(s, \mathcal{P}) = \{p \in \mathcal{P} \mid \|s - p\| = \mathrm{d}(s, P)\}$.

A. Voronoi partitions and multi-center problems

Here we present some relevant concepts on Voronoi diagrams and refer the reader to [10], [12] for comprehensive treatments. A *partition* of $\mathcal{D}$ is a collection of n polygons $\mathcal{W} = \{W_1, \ldots, W_n\}$ with disjoint interiors whose union is $\mathcal{D}$. The *Voronoi partition* $\mathcal{V}(P) = (V_1(P), \ldots, V_n(P))$ of $\mathcal{D}$ generated by the points $P = (p_1, \ldots, p_n)$ is defined by

$$V_i(P) = \{q \in \mathcal{D} \mid \|q - p_i\| \leq \|q - p_j\|\,,\ \forall j \neq i\}\,.$$

We say that P is a *circumcenter Voronoi configuration* if $p_i = \mathrm{CC}(V_i(P))$, for all $i \in \{1, \ldots, n\}$, and that P is an *incenter Voronoi configuration* if $p_i \in \mathrm{IC}(V_i(P))$, for all $i \in \{1, \ldots, n\}$. An incenter Voronoi configuration is *isolated* if there exists a neighborhood around it in $\mathcal{D}^n$ which does not contain any other incenter Voronoi configuration.

Consider the *disk-covering* and *sphere-packing multi-center* functions defined by

$$\mathcal{H}_{\text{DC}}(P) = \max_{s \in \mathcal{D}} \{\mathrm{d}(s, i_{\mathbb{F}}(P))\}\ ,$$
$$\mathcal{H}_{\text{SP}}(P) = \min_{i \neq j \in \{1, \ldots, n\}} \Big\{\frac{1}{2}\|p_i - p_j\|, \mathrm{d}(p_i, \partial\mathcal{D})\Big\}.$$

We are interested in the configurations that optimize these multi-center functions. The minimization of $\mathcal{H}_{\text{DC}}$ corresponds to minimizing the largest possible distance of any point in $\mathcal{D}$ to one of the agents' locations given by $p_1, \ldots, p_n$. We refer to it as the as the *multi-circumcenter problem*. The maximization of $\mathcal{H}_{\text{SP}}$ corresponds to the situation where we are interested in maximizing the coverage of the area $\mathcal{D}$ in such a way that the radius of the generators do not overlap (in order

not to interfere with each other) or leave the environment. We refer to it as the *multi-incenter problem*. It is useful to define the *index function* $N : \mathcal{D}^n \to \mathbb{N}$ as

$$N(P) = \Big| \operatorname*{argmin}_{p_i \neq p_j} \left\{ \frac{1}{2}\|p_i - p_j\|, \mathrm{d}(p_i, \partial\mathcal{D}) \right\} \Big|.$$

B. Spatial prediction via simple kriging

This section reviews the geostatistical kriging procedure for the estimation of spatial processes, see e.g., [1], [13]. A random process Z is *second-order stationary* if it has constant mean, $E(Z(s)) = \mu$, and its covariance is of the form $\mathrm{Cov}(Z(p_1), Z(p_2)) = C(p_1, p_2)$, where $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is positive-definite and only depends on the difference $p_1 - p_2$. We will focus on *isotropic* covariance functions, which satisfy

$$C(p_1, p_2) = g(\|p_1 - p_2\|),$$

for some decreasing function $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. The covariance matrix of the set of points $p_1, \ldots, p_n \in \mathcal{D}$ is $\boldsymbol{\Sigma} = \Sigma(P) = [C(p_i, p_j)]_{i,j=1}^n \in \mathbb{R}^{n \times n}$. When it is clear from the context, we will use bold face to denote explicit dependence on P. We further define $c : \mathcal{D} \times \mathcal{D}^n \to \mathbb{R}^n$ to be the vector of covariances between a point $s \in \mathcal{D}$ and the locations in P, i.e., $\boldsymbol{c} = c(s, P) = (C(s, p_1), \ldots, C(s, p_n))^T$. Of particular use to us will be the associated correlation function, $\rho : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ defined by

$$\rho(p_1, p_2) = \frac{C(p_1, p_2)}{\sqrt{C(p_1, p_1)}\sqrt{C(p_2, p_2)}} = \frac{g(\|p_1 - p_2\|)}{g(0)}.$$

Throughout the paper, we make the following assumptions on the model for the spatial random process Z of interest. We assume that Z is of the form

$$Z(s) = \mu(s) + \delta(s), \quad s \in \mathcal{D}, \tag{1}$$

and that the mean function μ is known. Also, δ is a zero-mean second-order stationary random process with a known decreasing isotropic covariance function, g. We will further assume that g is everywhere differentiable. Some examples of such functions are the exponential, cubic, spherical, modified Bessel, and rational quadratic covariance functions.

Assume measurement data $\boldsymbol{y} = (Y(p_1), \ldots, Y(p_n))^T$ is corrupted with error such that

$$Y(p_i) = Z(p_i) + \epsilon_i, \qquad \epsilon_i \overset{\text{iid}}{\sim} N\left(0, \tau^2\right), \tag{2}$$

where $\tau^2 \geq 0$. The assumption that the errors ϵ_i, $i \in \{1, \ldots, n\}$ are independent and identically distributed corresponds to the fact that the robotic network is equipped with identical sensors. The no-error scenario is the one most widely studied in the geostatistics literature. In the error case, the covariance between $Y(p_i)$ and $Y(p_j)$ is given by

$$\text{Cov}(Y(p_i), Y(p_j)) = \begin{cases} C(p_i, p_j) + \tau^2, & \text{if } i = j, \\ C(p_i, p_j), & \text{if } i \neq j. \end{cases}$$

Note that the covariance matrix of P with respect to the noisy process Y may be written $\boldsymbol{\Sigma}_\tau = \Sigma_\tau(P) = \boldsymbol{\Sigma} + \tau^2 I_n$, where I_n denotes the $n \times n$ identity matrix. The no-error case is recovered by setting $\tau^2 = 0$.

The *simple kriging predictor* at $s \in \mathcal{D}$ is the predictor that minimizes the mean-squared prediction error $\sigma^2(s; p_1, \ldots, p_n) = E(Z(s) - p(s; Y(p_1), \ldots, Y(p_n)))^2$ among all linear predictors of the form $p(s; Y(p_1), \ldots, Y(p_n)) = \sum_{i=1}^n l_i Y(p_i) + k$. The explicit expression of the simple kriging predictor at $s \in \mathcal{D}$ is

$$\hat{p}_{\text{SK}}(s; Y(p_1), \ldots, Y(p_n)) = \mu(s) + \boldsymbol{c}^T \boldsymbol{\Sigma}_\tau^{-1} (\boldsymbol{y} - \boldsymbol{\mu}), \tag{3}$$

where $\boldsymbol{\mu} = (\mu(p_1), \ldots, \mu(p_n))^T$. The mean-squared prediction error of $\hat{p}_{\text{SK}}$ at $s \in \mathcal{D}$ is

$$\sigma^2(s; p_1, \ldots, p_n) = g(0) - \boldsymbol{c}^T \boldsymbol{\Sigma}_\tau^{-1} \boldsymbol{c}. \tag{4}$$

Note that σ^2 is invariant under permutations of $p_1, \ldots, p_n$.

III. Problem statement

Consider a network of n agents evolving in a convex polytope $\mathcal{D} \subset \mathbb{R}^d$ according to the first-order dynamics $\dot{p}_i = u_i$, $i \in \{1, \ldots, n\}$. Assume each agent is equipped with a point-sized footprint sensor, and can take a noisy measurement $Y(p_i)$ as in (2) of the spatial process Z at its current position p_i. A natural objective is to select locations to take measurements in such a way as to minimize the uncertainty in the estimate of the spatial process. Here, we consider objectives inspired by the notions of G- and D-optimality from optimal design [1], [3].

The *maximum prediction error* is

$$\mathcal{M}(p_1, \ldots, p_n) = \max_{s \in \mathcal{D}} \sigma^2(s; p_1, \ldots, p_n) = g(0) - \min_{s \in \mathcal{D}} \{\boldsymbol{c}^T \boldsymbol{\Sigma}_\tau^{-1} \boldsymbol{c}\}. \tag{5a}$$

Note that $\mathcal{M}$ corresponds to a "worst-case scenario," where we consider locations in the domain with the maximum kriging mean-squared error. Let us make an important observation about the well-posedness of $\mathcal{M}$. Under noisy measurements, i.e., $\tau^2 > 0$, the function σ^2 is well-defined for any $s \in \mathcal{D}$ and $(p_1, \ldots, p_n) \in \mathcal{D}^n$. Indeed, the dependence of σ^2 on the network configuration is continuous, and hence, $\mathcal{M}$ is also well-defined. However, when no measurement noise is present, i.e., $\tau^2 = 0$, then the matrix $\mathbf{\Sigma}_\tau = \mathbf{\Sigma}$ in (4) is not invertible for network configurations that belong to S_{coinc}, and therefore, it is not clear what the value of σ^2 is. This problem is carefully addressed in Appendix A, where it is shown that in the no measurement noise case, σ^2 is a continuous function of the network configuration under suitable technical conditions on the covariance structure of the spatial field.

Before presenting the second objective function, we need to introduce some notions. Note that the variance of the simple kriging predictor is equivalent to the variance of the expression $\boldsymbol{c}^T \mathbf{\Sigma}_\tau^{-1} \boldsymbol{c}$, while the generalized variance [14] is considered to be the determinant of the covariance matrix $\mathbf{\Sigma}_\tau^{-1}$. Minimizing the determinant of $\mathbf{\Sigma}_\tau^{-1}$ is equivalent to minimizing $-|\mathbf{\Sigma}_\tau|$, where $|\cdot|$ denotes the determinant. For discrete state spaces, it can be shown [9] that configurations which maximize the minimum distance between agents asymptotically minimize $-|\mathbf{\Sigma}_\tau|$ in the limit of near independence, but this results in configurations which tends to place agents on the boundary of $\mathcal{D}$. Since we are only interested in predictions over $\mathcal{D}$, we would like a notion of optimality which penalizes agents too close to the boundary as it does agents too close to each other. This can be achieved as follows. Let $\gamma : \mathcal{D} \to \mathbb{R}^d$ map a point in $\mathcal{D}$ to its mirror image reflected across the nearest boundary of $\mathcal{D}$. Formally,

$$\gamma(s) \in s + 2\left(\operatorname*{argmin}_{s^* \in \partial\mathcal{D}} \{\|s^* - s\|\} - s\right).$$

Note that $\gamma(s)$ is in general not unique, and is not a smooth function of s. However, $\|s - \gamma(s)\|$ is smooth, and is the same for all values of $\gamma(s)$. Now consider minimizing the determinant of the simple kriging predictor which would result if we had data from all agents as well as their reflections. The *generalized prediction variance* is then

$$\mathcal{B}(p_1, \ldots, p_n) = -\left|\Sigma_\tau(p_1, \ldots, p_n, \gamma(p_1), \ldots, \gamma(p_n))\right|. \tag{5b}$$

Note that since $\mathcal{B}$ does not require inversion of the covariance matrix, it is always well-posed.

Our main goal is to find network configurations $p_1, \ldots, p_n$ that minimize $\mathcal{M}$ and $\mathcal{B}$.

IV. Optimal network configurations for spatial prediction

In this section, we provide several results that characterize the optimal network configurations of the objective functions $\mathcal{M}$ and $\mathcal{B}$. In Section IV-A, we show that minima of $\mathcal{M}$ cannot be in S_{coinc}. This fact is useful in Section IV-B where we show that, for simple kriging, circumcenter and incenter Voronoi configurations are asymptotically optimal for $\mathcal{M}$ and $\mathcal{B}$, respectively.

A. Coincident configurations are not minima of the maximum prediction error

In this section, we will examine the effect of the location of a subset of agents on the mean-squared error terms. In particular we are interested in comparing $\sigma^2(s;P)$ against $\sigma^2(s;i_{\mathbb{F}}(P))$ for configurations $P \in S_{\text{coinc}}$. The following lemma provides a useful decomposition of σ^2.

Lemma IV.1 *The simple kriging mean-squared error function may be written in the form*

$$\sigma^2(s;P) = \sigma^2(s;\overline{P}) - \frac{\left(\mathcal{N}(s,p_1;\overline{P})\right)^2}{\sigma^2(p_1;\overline{P}) + \tau^2}, \tag{6}$$

with $\mathcal{N}(s,p_1;\overline{P}) = C(s,p_1) - c^T(s,\overline{P})\Sigma_\tau(\overline{P})^{-1}c(p_1,\overline{P})$ *and* $\overline{P} = (p_2,\dots,p_n) \in \mathcal{D}^{n-1}$.

The proof of this fact follows from using [15, Proposition 8.2.4] for the inverse of a partitioned symmetric matrix. Equation (6) may be applied repeatedly to isolate the effects of any subset of locations in P.

Under the assumptions of Proposition A.3, we can extend the mean squared error function by continuity to include configurations in S_{coinc}. With a slight abuse of notation, in the case of no measurement error, we will use $\sigma^2(s;P)$ to denote $\sigma^2(s;i_{\mathbb{F}}(P))$ for $P \in S_{\text{coinc}}$.

Proposition IV.2 (Minima of $\mathcal{M}$ are not in S_{coinc}) *Let $P^\dagger \in \mathcal{D}^n$ be a strict local minimum of the map $P \mapsto \mathcal{M}(P)$. Under the assumptions of Proposition A.3, $P^\dagger \notin S_{coinc}$.*

Proof: We proceed by contradiction. Assume $P^\dagger \in S_{\text{coinc}}$. Consider a configuration $P \in \mathcal{D}^n \setminus S_{\text{coinc}}$ in a neighborhood of $P^\dagger$ such that $i_{\mathbb{F}}(P^\dagger) \subset i_{\mathbb{F}}(P)$. Let $s, s^\dagger \in \mathcal{D}$ such that $\mathcal{M}(P) = \sigma^2(s;P)$ and $\mathcal{M}(P^\dagger) = \sigma^2(s^\dagger;P^\dagger)$. Using Lemma IV.1 and Proposition A.3, one can deduce that $\sigma^2(s;P^\dagger) \geq \sigma^2(s;P)$. By the definition of $\mathcal{M}$, $\sigma^2(s^\dagger;P^\dagger) \geq \sigma^2(s;P^\dagger)$. Therefore $\mathcal{M}(P^\dagger) = \sigma^2(s^\dagger;P^\dagger) \geq \sigma^2(s;P^\dagger) \geq \sigma^2(s;P) = \mathcal{M}(P)$, which is a contradiction. ■

B. Multi-center Voronoi configurations are asymptotically optimal

Let us consider the objective functions $\mathcal{M}$ and $\mathcal{B}$ introduced in Section III but with covariance function C^k, $k \in \mathbb{N}$. This is equivalent to considering the correlation, ρ^k. As k grows, the correlation between distinct points in $\mathcal{D}$ vanishes. To ease the exposition, we denote by $\boldsymbol{c}^{(k)}$, respectively $\boldsymbol{\Sigma}_\tau^{(k)}$, the vector $\boldsymbol{c}$, respectively the matrix $\boldsymbol{\Sigma}_\tau$, with each element raised to the kth power. We will also use $\mathcal{M}^{(k)}$ and $\mathcal{B}^{(k)}$ to denote the objective functions $\mathcal{M}$ and $\mathcal{B}$ with covariance function C^k, that is,

$$\mathcal{M}^{(k)}(p_1, \ldots, p_n) = g^k(0) - \min_{s \in \mathcal{D}} \{(\boldsymbol{c}^{(k)})^T (\boldsymbol{\Sigma}_\tau^{(k)})^{-1} \boldsymbol{c}^{(k)}\},$$

$$\mathcal{B}^{(k)}(p_1, \ldots, p_n) = -\left|\Sigma_\tau^{(k)}\left(p_1, \ldots, p_n, \gamma(p_1), \ldots, \gamma(p_n)\right)\right|.$$

Let us start by establishing a result on the cardinality of the minimum distance set. Let $C_{\mathrm{mds}} : \mathbb{R}^d \times \mathcal{D}^n \to \mathbb{R}$ be defined by

$$C_{\mathrm{mds}}(s, P) = C(s, p) \quad \text{for any } p \in \mathrm{mds}(s, P).$$

Note that C_{mds} is well-defined.

Proposition IV.3 (Cardinality of minimum distance set) *Let the covariance function C be continuous. For $P \in \mathcal{D}^n \setminus S_{coinc}$, one has*

$$\min_{s \in \mathcal{D}} \left\{C_{\mathrm{mds}}(s, P)\, |\mathrm{mds}(s, P)|\right\} = \min_{s \in \mathcal{D}} \left\{C_{\mathrm{mds}}(s, P)\right\}.$$

Proof: We proceed by contradiction. If the statement is false, then there exists $s^\dagger \in \mathcal{D}$ such that $s^\dagger \in \mathrm{argmin}_{s \in \mathcal{D}} \{C_{\mathrm{mds}}(s, P)\, |\mathrm{mds}(s, P)|\}$, and $\left|\mathrm{mds}(s^\dagger, P)\right| > 1$. Let $p^* \in \mathrm{mds}(s^\dagger, P)$, and define $r^\dagger = \|s^\dagger - p^*\|$. Note that

$$\mathrm{mds}(s^\dagger, P) \subset \partial \overline{B}(s^\dagger, r^\dagger). \tag{7}$$

Let $s^* \in]s^\dagger, p^*[$ such that $\|s^* - s^\dagger\| < \epsilon$ for some $\epsilon \in \mathbb{R}_{>0}$ and let $r^* = \|s^* - p^*\|$. By construction, $r^* < r^\dagger$. From (7), we deduce that $\left\{p \in \mathcal{P} \mid p \in \overline{B}(s^*, r^*)\right\} = \{p^*\}$, which leads to $|\mathrm{mds}(s^*, P)| = 1$. Since C changes continuously with the distance between its arguments, it is clear that we may choose ϵ small enough to result in

$$C_{\mathrm{mds}}(s^*, P)\, |\mathrm{mds}(s^*, P)| < C_{\mathrm{mds}}(s^\dagger, P)\left|\mathrm{mds}(s^\dagger, P)\right|,$$

which is a contradiction. ■

We are now ready to prove one of the main results of the paper. The proof follows a similar line of reasoning to [9].

Theorem IV.4 (Minima of $\mathcal{M}$ under near independence) *Let $P_{mcc} \in \mathcal{D}^n$ be a global minimizer of the multi-circumcenter problem. Then, as $k \to \infty$, P_{mcc} asymptotically globally optimizes $\mathcal{M}^{(k)}$, that is, $\mathcal{M}^{(k)}(P_{\mathrm{mcc}})$ approaches a global minimum.*

Proof: Note that minimizing $\mathcal{M}^{(k)}$ is equivalent to finding the tuples P which maximize the function $L^{(k)} : \mathcal{D}^n \to \mathbb{R}$ defined as

$$L^{(k)}(P) = \min_{s \in \mathcal{D}} \left\{ (c^{(k)}(s,P))^T (\Sigma_\tau^{(k)}(P))^{-1} (c^{(k)}(s,P)) \right\}.$$

Let $\lambda_{\min}$ and $\lambda_{\max} : \mathcal{D}^n \times \mathbb{R} \to \mathbb{R}$ be such that $\lambda_{\min}(P,k)$, $\lambda_{\max}(P,k)$ denote, respectively, the minimum and the maximum eigenvalue of $\Sigma_\tau^{(k)}(P)$. We can see that $L^{(k)}(P)$ is bounded above by $\lambda_{\max}(P,k) \sum_{p \in P} C(s,p)^{2k}$ and below by $\lambda_{\min}(P,k) \sum_{p \in P} C(s,p)^{2k}$. For a given s, in terms of the minimum distance set we can write

$$\begin{aligned} \sum_{p \in P} C(s,p)^{2k} &= \sum_{p \in \mathrm{mds}(s,P)} C(s,p)^{2k} + \sum_{p \in P \setminus \mathrm{mds}(s,P)} C(s,p)^{2k} \\ &= |\mathrm{mds}(s,P)|\, C_{\mathrm{mds}}(s,P)^{2k} + \sum_{p \in P \setminus \mathrm{mds}(s,P)} C(s,p)^{2k}. \end{aligned}$$

As $k \to \infty$ the elements in the minimum distance set dominate, so we are left with

$$\sum_{p \in P} C(s,p)^{2k} = |\mathrm{mds}(s,P)|\, C_{\mathrm{mds}}(s,P)^{2k} + o(C_{\mathrm{mds}}(s,P)^{2k}).$$

Note from Proposition IV.3 that

$$\min_{s \in \mathcal{D}} \left\{ |\mathrm{mds}(s,P)|\, C_{\mathrm{mds}}(s,P) \right\} = \min_{s \in \mathcal{D}} \left\{ C_{\mathrm{mds}}(s,P) \right\},$$

so we can write

$$\min_{s \in \mathcal{D}} \Big\{ \sum_{p \in P} C(s,p)^{2k} \Big\} = \min_{s \in \mathcal{D}} \left\{ C_{\mathrm{mds}}(s,P)^{2k} \left(1 + o(1)\right) \right\}.$$

Consider, then, comparing an arbitrary configuration P^* against a global minimizer of $\mathcal{H}_{\mathrm{DC}}$, say P_{mcc}. In the zero measurement error case, by Proposition IV.2, we can assume without loss of generality that $P^* \notin S_{\mathrm{coinc}}$. Therefore, no matter what the value of τ is, we can safely use the eigenvalues of $(\mathbf{\Sigma}_\tau^{(k)})^{-1}$ to provide bounds. Specifically,

$$\frac{L^{(k)}(P^*)}{L^{(k)}(P_{\mathrm{mcc}})} \leq \frac{\lambda_{\max}(P^*,k) \min\limits_{s \in \mathcal{D}} \left\{ C_{\mathrm{mds}}(s,P^*)^{2k} \left(1 + o(1)\right) \right\}}{\lambda_{\min}(P_{\mathrm{mcc}},k) \min\limits_{s \in \mathcal{D}} \left\{ C_{\mathrm{mds}}(s,P_{\mathrm{mcc}})^{2k} \left(1 + o(1)\right) \right\}}. \tag{8}$$

Note that $\lim_{k\to\infty} \Sigma_\tau^{(k)}(P) = (g(0)+\tau^2)\, I_n$, so $\lambda_{\min}(P,k)$ and $\lambda_{\max}(P,k)$ both tend to $g(0)+\tau^2$ for any configuration P. Since P_{mcc} minimizes the maximum distance to any point $s \in \mathcal{D}$, it maximizes the minimum covariance, so for any $P \in \mathcal{D}^n$, $\min_{s\in\mathcal{D}} C_{\mathrm{mds}}(s,P) \leq \min_{s\in\mathcal{D}} C_{\mathrm{mds}}(s,P_{\mathrm{mcc}})$. Thus the ratio (8) is bounded by $1+o(1)$. Therefore, in the limit as $k\to\infty$, minimizing $\mathcal{M}^{(k)}$ is equivalent to solving the multi-circumcenter problem. ■

Note that the proof of the theorem can be reproduced for local minimizers of the multi-circumcenter problem to arrive at the following result.

Corollary IV.5 *Let $P_{\mathrm{mcc}} \in \mathcal{D}^n$ be a local minimizer of the multi-circumcenter problem. Then, as $k\to\infty$, P_{mcc} asymptotically optimizes $\mathcal{M}^{(k)}$, that is, $\mathcal{M}^{(k)}(P_{\mathrm{mcc}})$ approaches a minimum.*

According to [11], under certain technical conditions, solutions to the multi-circumcenter problem are circumcenter Voronoi configurations. Next, let us present a similar asymptotic result for the generalized prediction variance.

Theorem IV.6 (Minima of $\mathcal{B}$ under near independence) *Let $P_{\mathrm{mic}} \in \mathcal{D}^n$ be a global maximizer of the multi-incenter problem with lowest index. Then, as $k\to\infty$, P_{mic} asymptotically globally optimizes $\mathcal{B}^{(k)}$, that is, $\mathcal{B}^{(k)}(P_{\mathrm{mic}})$ approaches a global minimum.*

Proof: Expanding the objective function for asymptotically dominant terms, we may write

$$\mathcal{B}^{(k)}(P) = -(g(0)^k+\tau^2)^{2n} + \left(g(0)^k+\tau^2\right)^{2n-2} J^{(k)}(P) + o\left(\left(g(0)^k+\tau^2\right)^{2n-2} J^{(k)}(P)\right),$$

where

$$J^{(k)}(P) = \sum_{i\neq j} g(\|p_i-p_j\|)^{2k} + \sum_{i,j=1}^{n} g(\|p_i-\gamma(p_j)\|)^{2k} + \sum_{i\neq j} g(\|\gamma(p_i)-\gamma(p_j)\|)^{2k}.$$

Asymptotically all but the largest terms in $J^{(k)}(P)$ will drop out, and minimizing $\mathcal{B}^{(k)}(P)$ becomes equivalent to minimizing those terms. The largest terms in $J^{(k)}(P)$ correspond to the shortest distance between the locations of either the agents or their reflected images. For any two agent locations, $p_i, p_j \in \mathcal{D}$, and any of their reflections $\gamma(p_i), \gamma(p_j)$ the minimum distance between any two of the four points can be reduced to $\min\{\|p_i-p_j\|, \|p_i-\gamma(p_i)\|, \|p_j-\gamma(p_j)\|\}$ (note that this is not in general true for non-convex domains). Thus the shortest distance between agents in P and their reflections may be expressed as $2\mathcal{H}_{\mathrm{SP}}(P)$, though there may be multiple pairs of the form (p_i,p_j) or $(p_i,\gamma(p_i))$ which satisfy the minimum distance criterion, i.e., the index

of P might be larger than 1. Therefore we have $J^{(k)}(P) = N(P)\left(g(2\mathcal{H}_{\text{SP}}(P))^{2k}\right)(1+o(1))$. Consider comparing an arbitrary configuration, $P^* \in \mathcal{D}^n$ against P_{mic}. We have

$$\frac{J^{(k)}(P_{\text{mic}})}{J^{(k)}(P^*)} = \frac{N(P_{\text{mic}})\left(g(2\mathcal{H}_{\text{SP}}(P_{\text{mic}}))^{2k}\right)(1+o(1))}{N(P^*)\left(g(2\mathcal{H}_{\text{SP}}(P^*))^{2k}\right)(1+o(1))}.$$

If P^* is not a global solution of the multi-incenter problem, we have $\mathcal{H}_{\text{SP}}(P_{\text{mic}}) > \mathcal{H}_{\text{SP}}(P^*)$, and since $g(\cdot)$ is decreasing this gives us

$$\lim_{k\to\infty} \frac{J^{(k)}(P_{\text{mic}})}{J^{(k)}(P^*)} = 0.$$

If on the other hand, P^* is a global solution of the multi-incenter problem, then, using the fact that P_{mic} has the lowest index among all of them, we deduce $\frac{J^{(k)}(P_{\text{mic}})}{J^{(k)}(P^*)} \leq 1 + o(1)$. ■

Note that the proof of the theorem can be reproduced for isolated local maximizers of the multi-incenter problem to arrive at the following result.

Corollary IV.7 *Let $P_{\text{mic}} \in \mathcal{D}^n$ be an isolated local maximizer of the multi-incenter problem. Then, as $k \to \infty$, P_{mic} asymptotically optimizes $\mathcal{B}^{(k)}$, that is, $\mathcal{B}^{(k)}(P_{\text{mic}})$ approaches a minimum.*

According to [11], under certain technical conditions, solutions to the multi-incenter problem are incenter Voronoi configurations.

C. Distributed coordination algorithms

In this section, we present coordination algorithms that steer the network towards circumcenter and incenter Voronoi configurations. We do this following the exposition in [11]. In light of the results in Section IV-B, this enables the network to perform a spatial prediction which is asymptotically optimal as $k \to \infty$.

Let us assume each agent can move according to a first-order dynamical model $\dot{p}_i = u_i$, $i \in \{1, \ldots, n\}$. Consider the following coordination algorithms

$$\dot{p}_i = \text{CC}(V_i(P)) - p_i, \tag{9a}$$

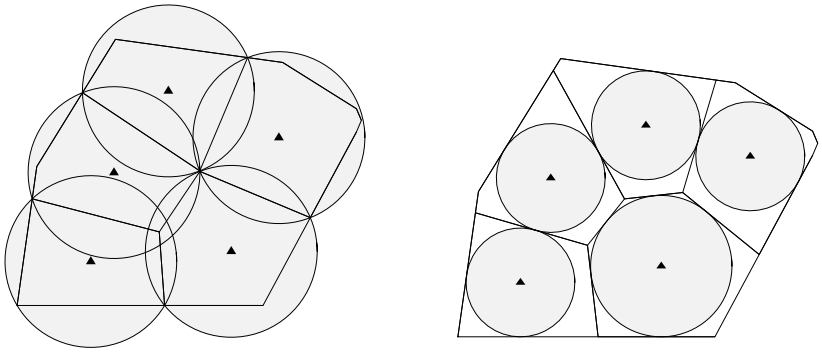
$$\dot{p}_i \in \text{IC}(V_i(P)) - p_i, \tag{9b}$$

for each $i \in \{1, \ldots, n\}$. Note that (9b) is a differential inclusion. We understand its solutions in the Filippov sense [16]. Both coordination algorithms are Voronoi distributed, meaning that each agent only requires information from its Voronoi neighbors in order to execute its control law.

The equilibrium points of the flow (9a) are the circumcenter Voronoi configurations, whereas the equilibrium points of the flow (9b) are incenter Voronoi configurations. Furthermore, the evolution of $\mathcal{H}_{\text{DC}}$ along (9a) is monotonically decreasing, while the evolution of $\mathcal{H}_{\text{SP}}$ along (9b) is monotonically increasing. The convergence properties of these coordination algorithms, as well as alternative flows with similar distributed properties that can also be used to steer the network to center Voronoi configurations, are studied in [11].

V. Simulations

With the aim of illustrating the results presented in Section IV, we performed simulations for both objective functions $\mathcal{M}$ and $\mathcal{B}$ with $n = 5$ agents. In our simulations, we used as domain $\mathcal{D}$ the convex polygon with vertices $\{(0, 0.1), (2.5, 0.1), (3.45, 1.6), (3.5, 1.7), (3.45, 1.8), (2.7, 2.2), (1.0, 2.4), (0.2, 1.3)\}$ and as isotropic covariance the one defined via $C : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, $C(s_1, s_2) = e^{-\frac{1}{5}\|s_1 - s_2\|}$. Note that, in simple kriging, the mean function does not play a role in determining the optimal network configurations. Figure 1 shows the multicenter configurations obtained with the flows (9).



(a) (b)

Fig. 1. Multi-center configurations found (a) using the flow (9a) and (b) using the flow (9b).

A. Analysis of simulations for $\mathcal{M}^{(k)}$

Using $\mathcal{M}^{(1)}$ we ran over 1000 random trials, each time running a gradient descent algorithm, and chose the local minimum configuration with the smallest value of $\mathcal{M}^{(1)}$ to be our approximation of a global minimum. From this configuration P_*, we generated a multi-circumcenter configuration using (9a), depicted in Figure 1(a). For increasing values of k, we ran a gradient

descent of $\mathcal{M}^{(k)}$ to find the best local configuration near P_*. For comparison, we also plotted the performance of a random configuration which was not a local minimum. Figure 2 illustrates the result in Theorem IV.4. At around $k = 15$, the performance of the circumcenter Voronoi

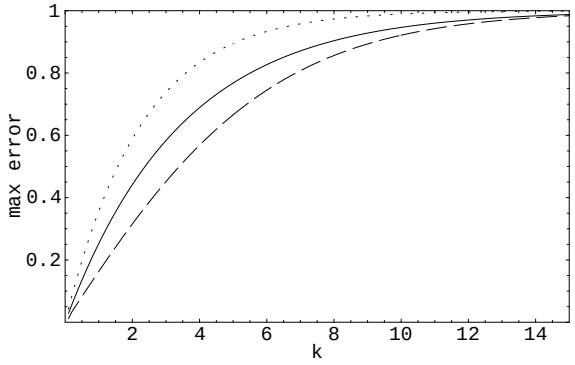


Fig. 2. Value of $\mathcal{M}^{(k)}$ for multi-circumcenter (solid), approximated global minimum (dashed), and random (dotted) configurations of 5 agents for increasing values of k. The covariance function is exponential.

configuration becomes almost identical to the one of the minimizer of $\mathcal{M}^{(k)}$.

B. Analysis of simulations for $\mathcal{B}^{(k)}$

Using $\mathcal{B}^{(1)}$ we ran over 1000 random trials, each time running a gradient descent algorithm, and chose the local minimum configuration with the smallest value of $\mathcal{B}^{(1)}$ to be our approximation of a global minimum. From this configuration P_* we generated the multi-incenter configuration using (9b), depicted in Figure 1(b). For increasing values of k, we ran a gradient descent of $\mathcal{B}^{(k)}$ to find the best local configuration near P_*. For comparison, we also plotted the performance of a random configuration which was not a local minimum. Figure 3 illustrates the result stated in Theorem IV.6. The performance of the minimizer of $\mathcal{B}^{(k)}$ and of the incenter Voronoi configuration are almost identical from the beginning, even though at each k the configurations are different.

VI. CONCLUSIONS

We have used simple kriging as a metric for optimal placement of mobile sensor networks estimating random fields. We have shown that under the assumption of near independence, circumcenter configurations minimize the maximum prediction error and incenter configurations minimize the generalized prediction variance. Under limited time or energy resources, or as a

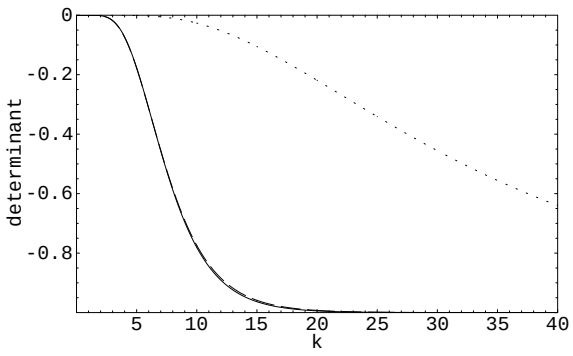


Fig. 3. Value of $\mathcal{B}^{(k)}$ for multi-incenter (solid), approximated global minimum (dashed), and random (dotted) configurations of 5 agents for increasing values of k. The covariance function is exponential. Although the performance of the global and multi-incenter configurations look identical, the configurations are different at each k.

starting point for further exploration, a group of robotic sensors might begin by moving toward these configurations to start the estimation procedure.

Future work will explore: (i) regarding the asymptotic analysis, the determination of lower and upper bounds on the parameter k that guarantee that multicenter Voronoi configurations achieve a given a desired level of performance, (ii) the extension of the results of this paper to similar error metrics for the universal kriging predictor, where the mean function is unknown, and (iii) the characterization of the trajectories (rather than configurations) that provide optimal estimates of the random field as agents reconfigure and take successive measurements over time. Consideration will also be given to fields which change over time, and to distributed methods for estimation and data fusion.

ACKNOWLEDGMENTS

This research was supported in part by NSF CAREER Award ECS-0546871.

REFERENCES

[1] N. A. C. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993. revised edition.

[2] M. L. Stein, *Interpolation of Spatial Data. Some Theory for Kriging*. Springer Series in Statistics, New York: Springer Verlag, 1999.

[3] F. Pukelsheim, *Optimal Design of Experiments*, vol. 50 of *Classics in Applied Mathematics*. Philadelphia, PA: SIAM, 2006.

[4] E. P. Liski, N. K. Mandal, K. R. Shah, and B. K. Sinha, *Topics in Optimal Design*, vol. 163 of *Lecture Notes in Statistics*. New York: Springer Verlag, 2002.

[5] J. S. Willcox, J. G. Bellingham, Y. Zhang, and A. B. Baggeroer, "Performance metrics for oceanographic surveys with autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 711–725, 2001.

[6] K. M. Lynch, I. B. Schwartz, P. Yang, and R. A. Freeman, "Decentralized environmental modeling by mobile sensor networks," *IEEE Transactions on Robotics*, 2006. Submitted.

[7] S. Martínez, "Distributed representation of spatial fields through an adaptive interpolation scheme," in *American Control Conference*, (New York City, NY), pp. 2750–2755, 2007.

[8] N. E. Leonard, D. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. Davis, "Collective motion, sensor networks, and ocean sampling," *Proceedings of the IEEE*, vol. 45, pp. 48–74, Jan. 2007. Special Issue on Networked Control Systems.

[9] M. E. Johnson, L. M. Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, vol. 26, pp. 131–148, 1990.

[10] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley Series in Probability and Statistics, New York: John Wiley, 2 ed., 2000.

[11] J. Cortés and F. Bullo, "Coordination and geometric optimization via distributed dynamical systems," *SIAM Journal on Control and Optimization*, vol. 44, no. 5, pp. 1543–1574, 2005.

[12] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*. New York: Springer Verlag, 2 ed., 2000.

[13] P. Abrahamsen, "A review of Gaussian random fields and correlation functions," Technical Report 917, Norwegian Computing Center, Oslo, Norway, 1997. Electronically available at http://publications.nr.no/917_Rapport.pdf.

[14] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, vol. 24, no. 3/4, pp. 471–494, 1932.

[15] D. S. Bernstein, *Matrix Mathematics*. Princeton, NJ: Princeton University Press, 2005.

[16] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides*, vol. 18 of *Mathematics and Its Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1988.

[17] V. Dem'yanov and V. Malozemov, *Introduction to Minimax*. New York: John Wiley, 1974.

[18] A. Shapiro, "On concepts of directional differentiability," *Journal of Optimization Theory & Applications*, vol. 66, no. 3, pp. 477–487, 1990.

APPENDIX A
CONTINUITY OF THE SIMPLE KRIGING MEAN-SQUARED PREDICTION ERROR

Here we prove our main continuity result for the simple kriging mean squared error function. We will first need some supporting results. Let $h \in \mathbb{R}$ denote the distance between agent locations p_1 and p_2, and we are interested in the behavior of σ^2 as $h \to 0$. For any agent location p_i, $i \in (3, \ldots, n)$ we would like a measure of the distance between p_i and p_1 in terms of h, call it $h_i \in \mathbb{R}$. To find h_i, consider the triangle formed between p_i, p_1, and p_2. Let $\mathbf{u}_{jk} \in \mathbb{R}^d$ denote the unit vector from agent j to agent k, i.e.,

$$\mathbf{u}_{jk} = \frac{p_k - p_j}{\|p_k - p_j\|}.$$

Let $\mathbf{u}_{ib}$ denote the unit vector in the direction which bisects the angle between $\mathbf{u}_{i1}$ and $\mathbf{u}_{i2}$ (see Figure 4). As in the figure, consider a point p_1^* which is a distance of $\|p_1 - p_i\|$ from p_i, but

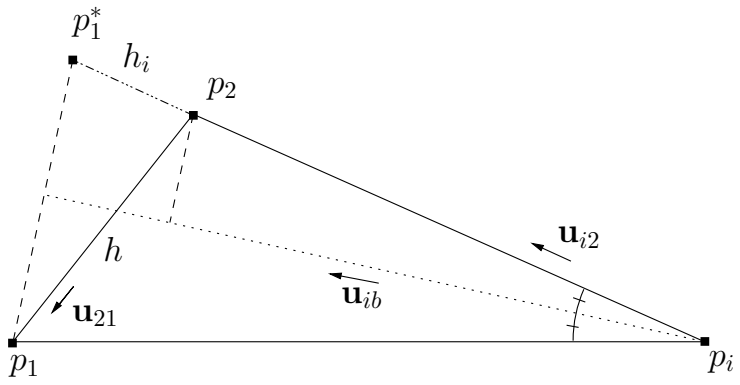


Fig. 4. Triangle between p_i, p_1, and p_2.

in the direction $\mathbf{u}_{i2}$, and note that $h_i = \|p_1^* - p_2\|$. Next, note that the projection of the vector $p_1^* - p_2$ onto $\mathbf{u}_{ib}$ is equal to the projection of the vector $p_1 - p_2$ onto the vector $\mathbf{u}_{ib}$, so that we may write

$$h_i \mathbf{u}_{ib}^T \mathbf{u}_{i2} = \mathbf{u}_{ib}^T \mathbf{u}_{21} h$$
$$h_i = \frac{\mathbf{u}_{ib}^T \mathbf{u}_{21}}{\mathbf{u}_{ib}^T \mathbf{u}_{i2}} h.$$

Note that in the limit as $p_1 \to p_2$, $h_i \to 0$ and $\mathbf{u}_{ib} \to \mathbf{u}_{i2}$.

In the following treatment, we define a function $f : \mathbb{R}^d \to \mathbb{R}$ to be directionally differentiable at a point $a \in \mathbb{R}^d$ if and only if the following limit exists for every $\mathbf{u} \in \mathbb{R}^d$

$$D_{\mathbf{u}} f(a) = \lim_{h \downarrow 0} \frac{f(a + h\mathbf{u}) - f(a)}{h}.$$

This is a common notion in the optimization literature (see for example [17], [18]), where it is sometimes referred to as weak directional differentiability. We will say that a function f is directionally differentiable on D if and only if $D_{\mathbf{u}} f(a)$ exists for all $a \in D$ and all $\mathbf{u} \in \mathbb{R}^d$. Recall from Lemma IV.1 the definition of $\overline{P} = (p_2, \ldots, p_n)$ being the ordered set of agents in P with the first agent removed. Now we are ready to present our results.

Lemma A.1 *Let $f_1, f_2 : \mathbb{R}^d \to \mathbb{R}$ be directionally differentiable on D. Let $F : \mathbb{R}^d \times \mathbb{R}^d \times D^{n-1} \to \mathbb{R}$ be defined as*

$$F(p_1, p_2, \overline{P}) = \frac{(f_1(p_1) - f_1(p_2))(f_2(p_1) - f_2(p_2))}{\sigma^2(p_1, \overline{P})}.$$

Under the assumption that $g'(0) \neq 0$ then

$$\lim_{p_1 \to p_2} F(p_1, p_2, \overline{P}) = 0.$$

Proof: First note that in the limit as $p_1 \to p_2$, both the numerator and denominator of the fraction tend to zero. With a little manipulation, we can rewrite F in terms of h. Remembering that $c(p_2, \overline{P})$ is the first row of the matrix $\Sigma(\overline{P})$, and that $C(p_1, p_1) = C(p_2, p_2)$, we can write

$$F(p_1, p_2, \overline{P}) = \frac{(f_1(p_1) - f_1(p_2))(f_2(p_1) - f_2(p_2))}{\left(c^T(p_1, \overline{P}) + c^T(p_2, \overline{P})\right) \Sigma(\overline{P})^{-1} \left(c(p_2, \overline{P}) - c(p_1, \overline{P})\right)}. \tag{10}$$

In terms of g, we can write

$$\begin{aligned} c(p_2, \overline{P}) - c(p_1, \overline{P}) &= \begin{bmatrix} g(\|p_2 - p_2\|) - g(\|p_1 - p_2\|) \\ g(\|p_2 - p_3\|) - g(\|p_1 - p_3\|) \\ g(\|p_2 - p_4\|) - g(\|p_1 - p_4\|) \\ \vdots \end{bmatrix} \\ &= \begin{bmatrix} \frac{g(0) - g(h)}{h} h \\ \frac{g(\|p_2 - p_3\|) - g(\|p_1 - p_3\|)}{h_3} \frac{\mathbf{u}_{3b}^T \mathbf{u}_{21}}{\mathbf{u}_{3b}^T \mathbf{u}_{32}} h \\ \frac{g(\|p_2 - p_4\|) - g(\|p_1 - p_4\|)}{h_4} \frac{\mathbf{u}_{4b}^T \mathbf{u}_{21}}{\mathbf{u}_{4b}^T \mathbf{u}_{42}} h \\ \vdots \end{bmatrix} \\ &= -\Delta h, \end{aligned}$$

where $\Delta \in \mathbb{R}^{n-1}$ is the vector

$$\Delta = \begin{bmatrix} \frac{g(h) - g(0)}{h} \\ \frac{g(\|p_2 - p_3\| + h_3) - g(\|p_2 - p_3\|)}{h_3} \frac{\mathbf{u}_{3b}^T \mathbf{u}_{21}}{\mathbf{u}_{3b}^T \mathbf{u}_{32}} \\ \frac{g(\|p_2 - p_4\| + h_4) - g(\|p_2 - p_4\|)}{h_4} \frac{\mathbf{u}_{4b}^T \mathbf{u}_{21}}{\mathbf{u}_{4b}^T \mathbf{u}_{42}} \\ \vdots \end{bmatrix}.$$

Note that the limit as $h \to 0$ corresponds to the limit as $p_1 \to p_2$ along the straight line direction $\mathbf{u}_{12}$, and that

$$\lim_{h\to 0} \Delta = \begin{bmatrix} g'(0) \\ \mathbf{u}_{32}^T \mathbf{u}_{21} g'(\|p_2 - p_3\|) \\ \mathbf{u}_{42}^T \mathbf{u}_{21} g'(\|p_2 - p_4\|) \\ \vdots \end{bmatrix}.$$

Plugging this back into F, we can write

$$\begin{aligned} & F(p_1, p_2, \overline{P}) \\ &= \frac{h^2 \left(\frac{(f_1(p_2+h\mathbf{u}_{21}) - f_1(p_2))}{h}\right) \left(\frac{(f_2(p_2+h\mathbf{u}_{21}) - f_2(p_2))}{h}\right)}{-h \left(c^T(p_1, \overline{P}) + c^T(p_2, \overline{P})\right) \Sigma(\overline{P})^{-1} \Delta} \\ &= \frac{h \left(\frac{(f_1(p_2+h\mathbf{u}_{21}) - f_1(p_2))}{h}\right) \left(\frac{(f_2(p_2+h\mathbf{u}_{21}) - f_2(p_2))}{h}\right)}{- \left(c^T(p_1, \overline{P}) + c^T(p_2, \overline{P})\right) \Sigma(\overline{P})^{-1} \Delta}. \end{aligned}$$

Note that for any directional unit vector $\mathbf{u} \in \mathbb{R}^d$,

$$\lim_{\substack{p_1 \to p_2 \\ along \mathbf{u}}} F(p_1, p_2, \overline{P}) \;=\; \left(D_{\mathbf{u}} f_1(p_2)\right) \left(D_{\mathbf{u}} f_2(p_2)\right) \left(\lim_{h\to 0} \frac{h}{- \left(c^T(p_1, \overline{P}) + c^T(p_2, \overline{P})\right) \Sigma(\overline{P})^{-1} \Delta} \right).$$

Regardless of the direction of $\mathbf{u}$, the numerator approaches zero. In the limit, the denominator evaluates to

$$-2c^T(p_2, \overline{P}) \Sigma(\overline{P})^{-1} \begin{bmatrix} g'(0) \\ \mathbf{u}_{32}^T \mathbf{u}_{21} g'(\|p_2 - p_3\|) \\ \mathbf{u}_{42}^T \mathbf{u}_{21} g'(\|p_2 - p_4\|) \\ \vdots \end{bmatrix} = -2g'(0).$$

Since this is constant with respect to the direction of approach, as long as $g'(0) \neq 0$, we have

$$\lim_{p_1 \to p_2} F(p_1, p_2, \overline{P}) = 0.$$

■

Corollary A.2 *Under the assumption of zero measurement error, if $g'(0) \neq 0$ then*

$$\lim_{p_1 \to p_2} \sigma^2(s; P) = \sigma^2(s; \overline{P})$$

Proof: Using Lemma IV.1 we can write

$$\sigma^2(s;P) = \sigma^2(s;\overline{P}) + \frac{\left(C(s,p_1) - c^T(s,\overline{P})\Sigma(\overline{P})^{-1}c(p_1,\overline{P})\right)^2}{\sigma^2(p_1,\overline{P})}$$
$$= \sigma^2(s;\overline{P}) + \frac{\left(C(s,p_1) - c^T(s,\overline{P})\Sigma(\overline{P})^{-1}\delta - C(s,p_2)\right)^2}{\sigma^2(p_1,\overline{P})},$$

where $\delta = c(p_1,\overline{P}) - c(p_2,\overline{P})$. Note that the second term here can be multiplied out as

$$\frac{\left(C(s,p_1) - C(s,p_2)\right)^2}{\sigma^2(p_1,\overline{P})} + 2\frac{\left(C(s,p_1) - C(s,p_2)\right)c^T(s,\overline{P})\Sigma(\overline{P})^{-1}\delta}{\sigma^2(p_1,\overline{P})} + \frac{\left(c^T(s,\overline{P})\Sigma(\overline{P})^{-1}\delta\right)^2}{\sigma^2(p_1,\overline{P})}$$

Since C is directionally differentiable everywhere, the first term fits the criteria of Lemma A.1, and goes to zero in the limit. For the other two, note that

$$c^T(s,\overline{P})\Sigma(\overline{P})^{-1}\delta = \sum_{i=1}^{n}\alpha_i\left(C(p_1,p_i) - C(p_2,p_i)\right),$$

where the α_i's do not depend on p_1. By Lemma A.1, for all i,j in $(1,\ldots,n)$ we can say

$$\lim_{p_1\to p_2}\frac{\left(C(p_1,p_i) - C(p_2,p_i)\right)\left(C(s,p_1) - C(s,p_2)\right)}{\sigma^2(p_1,\overline{P})} = 0,$$
$$\lim_{p_1\to p_2}\frac{\left(C(p_1,p_i) - C(p_2,p_i)\right)\left(C(p_1,p_j) - C(p_2,p_j)\right)}{\sigma^2(p_1,\overline{P})} = 0.$$

Thus all of the parts of our equation which depend on p_1 go to zero in the limit and we are left with

$$\lim_{p_1\to p_2}\sigma^2(s;P) = \sigma^2(s;\overline{P}).$$

■

We are now ready to present our main continuity result.

Proposition A.3 (Continuity of simple kriging predictor error) *Let $C:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ be an isotropic covariance function, $C(p_1,p_2) = g(\|p_1-p_2\|)$, with $g:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ differentiable. Assume $g'(0)\neq 0$, and $\tau^2 = 0$. Then, for each $s\in\mathcal{D}$, the mean-squared simple kriging prediction error $(p_1,\ldots,p_n)\mapsto\sigma^2(s;p_1,\ldots,p_n)$ is continuous. In addition, for $P\in S_{coinc}$ we have $\sigma^2(s;P) = \sigma^2(s;i_{\mathbb{F}}(P))$.*

Proof: Let $s\in\mathcal{D}$. Note that σ^2 is continuous when $P\in\mathcal{D}^n\setminus S_{\text{coinc}}$. Since the covariance function C is differentiable and hence continuous, it follows from (4) that σ^2 is continuous with respect to $P=(p_1,\ldots,p_n)\in\mathcal{D}^n$ except possibly where the matrix $\Sigma_\tau(P)$ is not full rank.

Under the model (2) with $\tau^2 \neq 0$, we have that Σ_τ is always full rank. Therefore, let us consider the ideal sensor case in which $\tau^2 = 0$, $Y(p_i) = Z(p_i)$, and $\Sigma_\tau(P) = \Sigma(P)$. Note that Σ being rank deficient corresponds precisely to the case when $P \in S_{\text{coinc}}$. Let us then take $P^\dagger \in S_{\text{coinc}}$. It suffices to show that

$$\lim_{P \to P^\dagger} \sigma^2(s; P) = \sigma^2(s; i_{\mathbb{F}}(P^\dagger)), \tag{11}$$

where $P \in \mathcal{D}^n \setminus S_{\text{coinc}}$. We begin by considering the case in which only two agents sit at the same location in the configuration $P^\dagger$. Since σ^2 is invariant under permutations of the agents, without loss of generality we can assume that $p_1^\dagger = p_2^\dagger$. Let then $P^\dagger = (p_1^\dagger, \overline{P})$. Since all points in $\overline{P}$ are distinct, we have

$$\lim_{P \to P^\dagger} \sigma^2(s; P) = \lim_{p_1 \to p_1^\dagger = p_2^\dagger} \sigma^2(s; P).$$

Using Corollary A.2, we can write

$$\lim_{p_1 \to p_1^\dagger = p_2^\dagger} \sigma^2(s; P) = \sigma^2(s; \overline{P}). \tag{12}$$

Since $\overline{P}$ is a specific ordering of $i_{\mathbb{F}}(P^\dagger)$, equation result (11) follows.

The case when more than two points in $P^\dagger$ are coincident can be dealt with similarly. If $|i_{\mathbb{F}}(P^\dagger)| = m \leq n-2$, we assume without loss of generality that $i_{\mathbb{F}}(P^\dagger) = \{p_{m+1}^\dagger, \ldots, p_n^\dagger\}$ using the fact that σ^2 is invariant under permutations. Then, we have

$$\lim_{P \to P^\dagger} \sigma^2(s; P) = \lim_{p_1 \to p_1^\dagger} \lim_{p_2 \to p_2^\dagger} \ldots \lim_{p_m \to p_m^\dagger} \sigma^2(s; P).$$

Repeatedly using (12), the limit above is well defined and, moreover, we conclude (11). ■