Double-Opportunity Estimation via Altruism

Nitai Stein and Yaakov Oshman, *Fellow, IEEE*

Abstract

Based on the notion of altruism, we present an approach to cooperative parameter estimation in a system comprising two information-sharing agents. The underlying assumption is that the overall two-agent scheme can reach desired performance level even if only one of the agents performs satisfactorily, hence there exist two independent opportunities to estimate. The notion of altruism motivates a definition of cooperative estimation optimality that generalizes the common definition of minimum mean squared error optimality. Fundamental equations are derived for two types of altruistic cooperative estimation problems, corresponding to heterarchical and hierarchical setups. Although these equations are, generally, hard to solve, their solution in the Gaussian case is straightforward and only entails the computation of the largest eigenvalue of the conditional covariance matrix and its corresponding eigenvector. Moreover, in the Gaussian case the performance improvement of the two altruistic cooperative estimation techniques over the conventional (egoistic) estimation approach is shown to depend on the problem's dimensionality and statistical distribution. In particular, the performance improvement grows with the dispersion of the spectrum of the conditional covariance matrix, rendering the presented estimation approach especially appealing in ill-conditioned problems. The validity of the solution in the Gaussian case is illustrated numerically.

I. Introduction

Complex missions often involve a number of systems, or agents, operating together as a team, to promote flexibility and robustness and to improve overall performance. In such teamwork, it is highly advantageous for the member agents constituting the team to be capable of sharing information among themselves. The need for improving teamwork, as well as the capability to share information among team members, have led to accelerated advances in the research of cooperative estimation. Current cooperative estimation algorithms differ in the way they handle shared information among the nodes of the distributed system, but, in the end, in most cases a *common team estimate* is computed, which is then used by the entire team.

In contradistinction to the prevailing team estimation concept, this paper introduces a nonorthodox paradigm in cooperative parameter estimation, whereby local estimates, that are generated by separate (but information-sharing) agents, are — by design — not identical, and they are not merged to form a unified estimate. Possibly even sub-optimal in the standard, minimum mean-squared error (MMSE) sense, these local estimates are designed, instead, to minimize together a single, global, system-oriented cost.

To illustrate this cooperative estimation paradigm and its applicability, consider the following scenario. An attacker (say, an aircraft equipped with high-precision guided missiles) is tasked with destroying a static high-valued target whose precise location is not deterministically known (say, a well-hidden rocket launcher). We assume that the attacker has two launch opportunities, i.e., it can fire two missiles at the target. We also assume that the attacker knows (either through intelligence sources or via its self-acquired measurements) the target location's distribution function. Clearly, as long as it is eventually successful in destroying the target, the attacker is completely indifferent as to how its mission is actually accomplished, that is, which of its two launch opportunities is successful in hitting the target. So, assuming that the attacker's two missiles are identical, and that they rely on the same positional information, how should they be aimed? If they are aimed at a common point, say, the mean of the target location's distribution function — which would be an optimal aiming point from an estimation perspective — this would amount to wasting one missile (assuming that both missiles operate flawlessly). Indeed, as will be shown later in this paper, the attacker will benefit from dispersing its two shooting opportunities, aiming them at different (possibly suboptimal from an individual missile's estimation standpoint) locations, in order to maximize its overall success probability.

The key assumption underlying the introduced cooperative estimation paradigm is that the system encompasses an inherent redundancy, which, in the scenario illustrated above, is embodied in the attacker having two missiles. Frequently called for to improve the probability of success in critically important missions, redundancy is implemented by using more than the minimal number of agents required to perform a certain, global, task. For example, in the theatre ballistic missile defense world, several (identical) defending interceptors may be launched at a single oncoming threat, if the defended target is so valuable that it must be protected at all cost [1]. While system redundancy obviously contributes to immunity against local, subsystem failures, it is proposed herein to exploit it in a seemingly unrelated manner. Thus, we note that when the system comprises several identical and information-sharing agents, it would be advantageous to put aside the performance of each individual estimator (each agent's estimate) and, instead, to focus on enhancing the estimation performance of the entire system in some global sense. Such design philosophy suggests an altruism-based cooperation, in which each agent forgoes its own (egoistic) estimation

The authors are with the Department of Aerospace Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. (E-mail: nitais@alumni.technion.ac.il; yaakov.oshman@technion.ac.il).
Accepted for publication by *IEEE Transactions on Aerospace and Electronic Systems*.

performance in order to maximize a global estimation performance measure. A well-known term in nature and in sociology, as well as in game theory, altruism means that an individual sacrifices itself for the greater good of its species, or in favor of other individuals, in order to improve the chances of its species to thrive. The use of altruism leads conceptually to a min-min game, that is, a game where all players (belonging to one side) cooperate such that there is one global goal to achieve, and the optimizer aspires to minimize a cost function based on the minimal cost among all players. The underlying notion is that the success of the global mission is determined only by the performance of the most successful individual among all agents. The interested reader is referred to [2], [3] for perfect-information examples of such min-min games in the field of missile guidance, where, however, the estimation problem was not addressed.

Two approaches for cooperative parameter estimation based on the notion of altruism are proposed herein. Called heterarchical altruistic cooperative estimation, the first approach considers two equally-ranked agents that take into account the action of each other, and calculate their estimates fully altruistically such that neither of the two is necessarily optimal (in the conventional sense). Thus, both agents sacrifice their own estimation performance in order to maximize the global mission's probability of success. Termed hierarchical altruistic cooperative estimation, the second approach is more conservative, in that it assumes that one of the agents operates egoistically, as if it were the only estimator present, thus minimizing the conventional mean-squared error (MSE) criterion. The second agent in this approach takes into account the action of the first one, and maximizes the success of the global mission, given the (egoistic) estimate of the first agent. Here the term hierarchy alludes to the fact that the first agent works as if it were ranked higher than the other, and tries to accomplish the mission on its own. Comparing the two approaches in performance, the heterarchical one is superior to the hierarchical one, as the former is globally optimal, whereas the latter results from a constrained optimization. However, in some cases, design conservatism may dictate the use of the hierarchical approach.

Returning to the scenario of the attacker and its two shooting attempts, assume now that the target's location is Gaussian distributed on a line. We will show in this paper that an attacker using the globally-optimal heterarchical approach would shift both of its missiles' aiming points to the sides of the target's expected location (the mode of the distribution), thus maximizing the overall probability of at least one of its missiles striking sufficiently close to the target. Alternatively, the more conservative attacker, that implements the hierarchical approach, would aim one of its missiles at the expected location of the target (maximizing this particular missile's probability of hitting the target), and would shift the aiming point of the second missile aside (reasonably, the second missile's aiming point's shift from the expected target location would be larger than the respective shifts of both heterarchical attempts).

The main contributions of this paper are the following:

1) We generalize the standard MMSE parameter estimation problem to the realm of cooperative estimation, in cases involving two separately-operating but information-sharing agents.
2) We introduce the concept of altruistic cooperative estimation, and, within this concept, we introduce two altruistic estimation approaches (heterarchical and hierarchical estimation), that outperform the standard, egoistic approach in the sense considered herein.
3) We prove the existence of a global solution for each of the two altruistic estimation approaches.
4) In the Gaussian case we provide closed-form, analytical solutions to the two altruistic estimation problems, along with a complete analysis of the solutions' potential performance advantages.

The remainder of this paper is organized as follows. The two altruistic cooperative estimation problems are defined in Section II. In Section III we derive necessary conditions for estimators corresponding to both problems, and prove the existence of optimal estimators satisfying these conditions. In Section IV we address the Gaussian case in detail. Concluding remarks are offered in the last section. Some technical derivations and proofs are deferred to Appendices.

II. Problem Formulation

Consider a random parameter vector $\boldsymbol{\theta}$ defined on the probability space $(\Theta, \mathcal{F}, P)$, where $\Theta \subseteq \mathbb{R}^n$ is the continuous sample space, $\mathcal{F}$ is the set of events (σ-algebra) on Θ, and P is a probability measure. The problem is to find two estimates of $\boldsymbol{\theta}$ based on the random vector of measurements $\boldsymbol{Z}$, which are (possibly nonlinear) functions of $\boldsymbol{\theta}$. The mapping $\boldsymbol{Z}$ induces the sample space $\mathcal{Z} \subseteq \mathbb{R}^m$ (with an appropriate σ-algebra). Both Θ and $\mathcal{Z}$ are Hilbert spaces, equipped with the 2-norm induced by the dot-product. For later purpose, we assume that the (known) joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{Z}$ has finite first two moments.

We consider a scenario where the system tasked with the estimation problem comprises two agents, each of which yields a local estimate of the parameter vector $\boldsymbol{\theta}$ based on the shared measurements $\boldsymbol{Z}$. The system does not merge the two local estimates to a final, single estimate; rather, its overall performance results, in some manner, from the joint performance of the two estimators. Context-depending, we will use the notation $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ to denote both the estimators and the estimates (generated by these estimators) of the two agents, respectively.

A cost function that reflects the idea of altruistic estimation is the following:

$$J(\hat{\boldsymbol{\theta}}^{(1)}, \hat{\boldsymbol{\theta}}^{(2)}) \triangleq \mathbb{E}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 \wedge \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2) \tag{1}$$

where $\mathbb{E}$ is the expectation operator, and $a \wedge b \triangleq \min(a, b)$ for some $a, b \in \mathbb{R}$.

The underlying premise of this work is that the global mission is accomplished even if only one of the agents provides an MSE-acceptable estimate. Thus, the overall system performance is determined by the performance of the better agent among the two.

Remark 1. *Setting the two estimators to be identical in (1) reduces it to the standard MMSE cost, manifesting the fact that the problem defined here is an extension of the MMSE estimation problem to the realm of altruistic cooperative estimation.*

Remark 2. *Somewhat resembling the optimal sub-pattern assignment (OSPA) metric of [4], [5], the cost (1) as defined here is radically different due to the difference between the meanings of both problems. In the OSPA case, the best targets-to-estimates combination is chosen, based on the premise that targets are unlabeled, so that the problem is how to optimally "throw two stones at two indistinguishable birds", aiming at hitting both. In contradistinction, in the present work the problem is "how to throw two stones at a single bird", while maximizing the probability that at least one (unlabeled) stone hits its target.*

We define two altruistic estimation problems. In the heterarchical problem, the estimators $\hat{\boldsymbol{\theta}}_{\text{HT}}^{(1)}$ and $\hat{\boldsymbol{\theta}}_{\text{HT}}^{(2)}$ solve the global minimization problem

$$\min_{\hat{\boldsymbol{\theta}}^{(1)},\hat{\boldsymbol{\theta}}^{(2)}\in L^2(\mathcal{Z})} J(\hat{\boldsymbol{\theta}}^{(1)},\hat{\boldsymbol{\theta}}^{(2)}) \tag{2}$$

where $L^2(\mathcal{Z})$ is the space of all square Lebesgue-integrable (measurable) functions of the measurements.

A constrained version of the heterarchical problem, the hierarchical problem sets one of the estimators, $\hat{\boldsymbol{\theta}}_{\text{HI}}^{(1)}$, identical to the minimum mean squared error estimator (MMSEE), $\hat{\boldsymbol{\theta}}_{\text{MS}}$. The second hierarchical altruistic estimator, $\hat{\boldsymbol{\theta}}_{\text{HI}}^{(2)}$, solves the constrained minimization problem

$$\min_{\hat{\boldsymbol{\theta}}^{(2)}\in L^2(\mathcal{Z})} J(\hat{\boldsymbol{\theta}}^{(1)},\hat{\boldsymbol{\theta}}^{(2)}) \quad \text{such that } \hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}_{\text{MS}}. \tag{3}$$

Problems (2) and (3) are closely related to some well-studied optimization problems appearing in the Voronoi literature [6]–[8], generally called *facility serviceability problems* [7]. In these problems, there exist some points called *facilities* (or, *Voronoi generators*), that are said to supply some necessary resource. The optimization task is to localize these facilities inside a populated region, such that they generate a Voronoi tessellation which is optimal in some sense. The simplest example is the problem of *public mail box localization*: given a city and its population distribution, the problem is to position a certain number of public mail boxes, assuming that every citizen in the city uses the closest public mail box. These problems appear in many scientific domains [8], such as data compression (e.g, in the image processing world), quantization, and distortion problems [9] in the signal compression world. However, to the best of the authors' knowledge, no closed-form solutions have been presented, perhaps because the literature focuses mainly on problems with many facilities, that require efficient numerical solutions, such as Lloyd's algorithm [6]–[8].

III. Estimator Derivation

Applying the smoothing theorem to the cost function (1) yields

$$J = \mathbb{E}[\mathbb{E}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 \wedge \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z})] = \mathbb{E}\, J_Z \tag{4}$$

where

$$J_Z \triangleq \mathbb{E}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 \wedge \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}). \tag{5}$$

Since the outer expectation in (4) does not depend on the choice of the estimators, the global minimizing arguments for J are identical to those of J_Z. We, thus, proceed with minimizing J_Z.

Consider the function $a \wedge b$ for some $a, b \in \mathbb{R}$. Clearly, in the region $a > b$, $a \wedge b = b$, so that the function is not affected by the value of a. Analogously, the space Θ can be divided into two subspaces, in each of which J_Z is affected by only one of the two estimates – the one closer to any value of $\boldsymbol{\theta}$ in this subspace. This observation naturally calls to mind the notion of Voronoi regions [8], giving rise to the following definition.

Definition 1 (Estimates' Voronoi regions)**.** *The Voronoi region of $\hat{\boldsymbol{\theta}}^{(1)}$, denoted $\mathcal{V}_1$, is a set in Θ such that:*

$$\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\| < \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\| \quad \forall \boldsymbol{\theta} \in \mathcal{V}_1. \tag{6}$$

Analogously, $\mathcal{V}_2$ is defined to satisfy

$$\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\| > \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\| \quad \forall \boldsymbol{\theta} \in \mathcal{V}_2. \tag{7}$$

The boundary separating both Voronoi regions (the Voronoi edge), denoted as $\partial\mathcal{V}$, is defined to satisfy:

$$\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\| = \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\| \quad \forall \boldsymbol{\theta} \in \partial\mathcal{V}. \tag{8}$$

Notice that the two estimates, $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$, play the part of Voronoi generators, and that the regions $\mathcal{V}_1$ and $\mathcal{V}_2$ along with $\partial\mathcal{V}$ constitute a Voronoi tessellation [8]. The optimization problems (2) and (3) are Voronoi optimization problems, for, using the law of total probability, we can express J_Z as

$$J_Z = \sum_{i=1}^{2} \mathbb{E}(\|\hat{\boldsymbol{\theta}}^{(i)} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}, \boldsymbol{\theta} \in \mathcal{V}_i) \Pr(\boldsymbol{\theta} \in \mathcal{V}_i \mid \boldsymbol{Z}) \tag{9}$$

where we have used the fact that $\partial\mathcal{V} \subset \Theta$ has measure zero, as $\dim \partial\mathcal{V} = \dim \Theta - 1$ (because $\partial\mathcal{V}$ satisfies a constraint in Θ).

The heterarchical problem stated in (2) is the classical Voronoi facility serviceability problem [7]. For given measurements, the estimates are the facilities, located in $\mathbb{R}^n$; the probability distribution of $\boldsymbol{\theta}$ serves as the population distribution, and each individual (random realization of $\boldsymbol{\theta}$) is associated with the estimate closest to it.

The hierarchical problem stated in (3) is a special case of the problem stated in [6, Section 9.2.4]. To see this, set the number of facilities to two (represented by the two estimates), the number of ranks to two (one egoistic estimate and one altruistic), and the consumption rate for each supply to half, such that both are equally consumed. In that case, it is obvious that the location of the higher ranked facility should be the MMSEE, since it is the only facility supplying this service to the entire population. However, this higher-ranked facility, whose location is already set, supplies also the second service which the other facility supplies as well. Hence, it is obvious that the optimizer should locate the lower-ranked facility according to the global mission of serviceability, taking into consideration the location of the higher-ranked facility.

For these kinds of problems, [8] proves that the optimal solutions lead to centroidal Voronoi tessellations (CVT), in which the facilities are the centroids of their corresponding Voronoi regions. Requiring the domain in which the problem is defined to be a compact subset of $\mathbb{R}^n$, [8, p. 651-652] proves the existence of a globally-optimal solution, that consists of a set of non-identical facilities. The localization problems addressed in the literature are commonly solved numerically, perhaps because most of them involve a large number of facilities [6]–[8]. Related to what would be called the 2-facilities problem in the Voronoi literature, our work extends already known results by proving the existence of a globally-optimal solution in not necessarily compact domains, and by providing a closed-form solution in the Gaussian case.

A. *Preliminary Calculations*

Because we allow unconstrained estimates, the minimizers of J_Z are those for which either the gradient vector vanishes, or the cost function is not differentiable. The function $a \wedge b$ for some $a, b \in \mathbb{R}$ is differentiable everywhere with respect to both a and b, except at $a = b$. Similarly, J_Z is not differentiable with respect to either of the two estimates only in the trivial case $\hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}^{(2)}$, which is not of interest here (one can always do better by dispersing the two estimates; see Lemma 2 in Appendix A). We, thus, seek for optimal solutions rendering $\hat{\boldsymbol{\theta}}^{(1)} \neq \hat{\boldsymbol{\theta}}^{(2)}$. Since, for such solutions, J_Z is differentiable, we will derive the necessary conditions by setting its gradient to zero. Notice that because J_Z is quadratic (assuming $\hat{\boldsymbol{\theta}}^{(1)} \neq \hat{\boldsymbol{\theta}}^{(2)}$), its extremum is necessarily a minimum.

To compute the gradient of J_Z we first rewrite (9) as

$$J_Z(\hat{\boldsymbol{\theta}}^{(1)}, \hat{\boldsymbol{\theta}}^{(2)}) = \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_1}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2)\, \mathbb{P}_{\mathrm{z}}(\mathcal{V}_1) + \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2}(\|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2)\, \mathbb{P}_{\mathrm{z}}(\mathcal{V}_2) \tag{10}$$

where the (probabilistic) measure of Voronoi region i and the local expectation operator associated with that region are defined, respectively, as

$$\mathbb{P}_{\mathrm{z}}(\mathcal{V}_i) \triangleq \Pr(\boldsymbol{\theta} \in \mathcal{V}_i \mid \boldsymbol{Z}), \quad i = 1, 2 \tag{11}$$

and

$$\mathbb{E}_{\mathrm{z}}^{\mathcal{V}_i}(\cdot) \triangleq \mathbb{E}(\cdot \mid \boldsymbol{Z}, \boldsymbol{\theta} \in \mathcal{V}_i), \quad i = 1, 2. \tag{12}$$

Let $\delta(\cdot)$ denote an infinitesimal perturbation of $(\cdot)$. Arbitrarily perturbing the first estimate to $\hat{\boldsymbol{\theta}}^{(1)} + \delta\hat{\boldsymbol{\theta}}^{(1)}$ while keeping the second estimate intact, results in a corresponding infinitesimal change in the Voronoi tessellation. In turn, this results in a perturbation in the cost,

$$\delta J_Z = \delta(\mathbb{E}_{\mathrm{z}}^{\mathcal{V}_1}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2)\, \mathbb{P}_{\mathrm{z}}(\mathcal{V}_1)) + \delta(\mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2}(\|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2)\, \mathbb{P}_{\mathrm{z}}(\mathcal{V}_2)). \tag{13}$$

The infinitesimal change in the tessellation affects both terms on the RHS of (13), but it does so in an antisymmetric manner, as the change in $\partial\mathcal{V}$, the boundary separating the two Voronoi regions, induces oppositely signed changes of both terms. Letting $\delta\hat{\boldsymbol{\theta}}^{(1)} \to \mathbf{0}$ nullifies the change in the tessellation, and its total effect on the perturbation of the cost. The other effect contributing to the perturbation of the cost is the change (integrated over all points $\boldsymbol{\theta} \in \mathcal{V}_1$) in the norm $\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|$ due to the change in $\hat{\boldsymbol{\theta}}^{(1)}$. Thus,

$$\delta J_Z = 2\, \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_1}(\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta})^T \delta\hat{\boldsymbol{\theta}}^{(1)}\, \mathbb{P}_{\mathrm{z}}(\mathcal{V}_1)\,. \tag{14}$$

By symmetry, (14) yields the gradients of J_Z as

$$\nabla_{\hat{\boldsymbol{\theta}}^{(i)}} J_Z = 2\,\mathbb{P}_{\mathrm{z}}(\mathcal{V}_i)\,\mathbb{E}_{\mathrm{z}}^{\mathcal{V}_i}(\hat{\boldsymbol{\theta}}^{(i)} - \boldsymbol{\theta}), \quad i = 1, 2. \tag{15}$$

Remark 3. *Addressing the multi-dimensional case,* (15) *was also derived in [6], [7], albeit in a deterministic setting.*

Next we show that there exists a rotation transformation, that, when applied to the parameter space Θ, maps the Voronoi regions of both estimates to one-dimensional, half-infinite intervals. This transformation will facilitate the ensuing derivation of the altruistic estimators. Defining

$$\Delta\hat{\boldsymbol{\theta}} \triangleq \hat{\boldsymbol{\theta}}^{(2)} - \hat{\boldsymbol{\theta}}^{(1)} \tag{16}$$

the Voronoi edge equation, (8), can be written as

$$\langle \boldsymbol{\theta} - \frac{\hat{\boldsymbol{\theta}}^{(1)} + \hat{\boldsymbol{\theta}}^{(2)}}{2}, \Delta\hat{\boldsymbol{\theta}} \rangle = 0 \tag{17}$$

where $\langle \cdot, \cdot \rangle$ stands for the inner product. Similarly, the definitions (6) and (7) of the estimates' Voronoi regions can be written (for any $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$) as

$$\langle \boldsymbol{\theta} - \frac{\hat{\boldsymbol{\theta}}^{(1)} + \hat{\boldsymbol{\theta}}^{(2)}}{2}, \Delta\hat{\boldsymbol{\theta}} \rangle < 0 \tag{18}$$

and

$$\langle \boldsymbol{\theta} - \frac{\hat{\boldsymbol{\theta}}^{(1)} + \hat{\boldsymbol{\theta}}^{(2)}}{2}, \Delta\hat{\boldsymbol{\theta}} \rangle > 0 \tag{19}$$

respectively. Equation (17) means that the boundary $\partial\mathcal{V}$ is an $(n-1)$-dimensional plane orthogonal to $\Delta\hat{\boldsymbol{\theta}}$, that contains the point $\frac{\hat{\boldsymbol{\theta}}^{(1)} + \hat{\boldsymbol{\theta}}^{(2)}}{2}$, the mid-point between the two estimates. The two Voronoi regions are located on opposite sides of the boundary.

Now let $\boldsymbol{\tau}$ be an $n \times n$ proper orthogonal matrix having $\Delta\hat{\boldsymbol{\theta}}^T / \|\Delta\hat{\boldsymbol{\theta}}\|$ as its first row ($\|\Delta\hat{\boldsymbol{\theta}}\|$ cannot vanish because, as explained earlier, we disregard the case $\hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}^{(2)}$). Define

$$\boldsymbol{u} \triangleq \boldsymbol{\tau}\boldsymbol{\theta} \tag{20}$$

$$\hat{\boldsymbol{u}}^{(i)} \triangleq \boldsymbol{\tau}\hat{\boldsymbol{\theta}}^{(i)}, \quad i = 1, 2 \tag{21}$$

$$\Delta\hat{\boldsymbol{u}} \triangleq \hat{\boldsymbol{u}}^{(2)} - \hat{\boldsymbol{u}}^{(1)}. \tag{22}$$

Rotating the standard basis of the space Θ using the transformation $\boldsymbol{\tau}$, let $\boldsymbol{e}_{\boldsymbol{u}_1}$ be a unit vector along the first basis vector of the rotated space. Using (16) and (21) in (22), and recalling the special construction of the orthogonal matrix $\boldsymbol{\tau}$, yields

$$\Delta\hat{\boldsymbol{u}} = \|\Delta\hat{\boldsymbol{\theta}}\|\boldsymbol{e}_{\boldsymbol{u}_1}. \tag{23}$$

Because $\boldsymbol{e}_{\boldsymbol{u}_1}$ is collinear with $\Delta\hat{\boldsymbol{u}}$, the vector connecting both transformed estimates, we call it the solution-axis.

Using $\boldsymbol{\tau}^T\boldsymbol{\tau} = \boldsymbol{I}$ in (18) along with (23) and the definitions (20) and (21), the Voronoi region of the first estimate can be expressed as

$$\mathcal{V}_1 = \{\boldsymbol{u} \in \Theta : \langle \boldsymbol{u} - \frac{\hat{\boldsymbol{u}}^{(1)} + \hat{\boldsymbol{u}}^{(2)}}{2}, \|\Delta\hat{\boldsymbol{\theta}}\|\boldsymbol{e}_{\boldsymbol{u}_1} \rangle < 0\}. \tag{24}$$

Let $\hat{u}^m$ be the projection of the midpoint between the two estimates in the transformed space on $\boldsymbol{e}_{\boldsymbol{u}_1}$, that is

$$\hat{u}^m \triangleq \frac{1}{2}(\hat{\boldsymbol{u}}^{(1)} + \hat{\boldsymbol{u}}^{(2)})^T\boldsymbol{e}_{\boldsymbol{u}_1}. \tag{25}$$

Using (25) in (24) yields

$$\mathcal{V}_1 = \{\boldsymbol{u} \in \Theta : u_1 < \hat{u}^m\} \tag{26}$$

where u_1 denotes the first component of the vector $\boldsymbol{u}$ (its projection along $\boldsymbol{e}_{\boldsymbol{u}_1}$). Stated in words, in the transformed space, $\mathcal{V}_1$ is the half-infinite open interval $(-\infty, \hat{u}^m)$ along the solution axis. Similarly, the Voronoi region of the second estimate in the transformed space is the half-infinite open interval $(\hat{u}^m, \infty)$ along the solution axis.

Having these preliminary results on hand, we now proceed with the derivation, considering separately each of the altruistic approaches.

B. Heterarchical Altruistic Estimation

Setting the gradient of J_Z [expressed in (15)] to zero yields

$$\hat{\boldsymbol{\theta}}^{(i)} = \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_{\mathrm{i}}} \boldsymbol{\theta}, \quad i = 1, 2 \tag{27}$$

which shows that an optimal heterarchical estimate is locally MMSE-optimal inside its Voronoi region. As such, it inherits the properties of MMSE estimators inside that region. We have thus shown that the two heterarchical estimators yield a CVT of Θ (this has also been shown, for other Voronoi problems of similar nature, in, e.g., [8]).

Equations (27) are coupled and, generally, hard to solve in closed form. Efficient algorithms for their *numerical* solution (even in the general case involving more than two facilities, e.g., Lloyd's algorithm) can be found in [6]–[8].

Remark 4. *Using the law of total probability and (27) yields*

$$\mathbb{E}(\boldsymbol{\theta} \mid \boldsymbol{Z}) = \mathbb{P}_{\mathrm{z}}(\mathcal{V}_1)\, \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_1} \boldsymbol{\theta} + \mathbb{P}_{\mathrm{z}}(\mathcal{V}_2)\, \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2} \boldsymbol{\theta} = \mathbb{P}_{\mathrm{z}}(\mathcal{V}_1)\hat{\boldsymbol{\theta}}^{(1)} + \mathbb{P}_{\mathrm{z}}(\mathcal{V}_2)\hat{\boldsymbol{\theta}}^{(2)} \tag{28}$$

which generalizes the fundamental theorem of MMSE estimation. This should come as no surprise, as the cost function (1) *generalizes the MSE cost function. In fact, as shown in Appendix A, when the norm-difference between the estimates tends to infinity, the Voronoi region of the estimate possessing the larger norm tends to a set of measure zero, whereas the other Voronoi region tends to a set of measure one. In that case, (28) yields that the (single) MMSE estimator is the familiar global conditional mean.*

Sometimes it might be advantageous to calculate first $\hat{u}^m$, the midpoint between estimates along the solution axis. To do that we use (20) in (27) to obtain the following alternative form of (27)

$$\hat{\boldsymbol{\theta}}^{(i)} = \boldsymbol{\tau}^T\, \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_{\mathrm{i}}} \boldsymbol{u}, \quad i = 1, 2. \tag{29}$$

Now using (21) in (25) and substituting (29) results in the following scalar equation

$$\hat{u}^m = \frac{1}{2}(\mathbb{E}_{\mathrm{z}}^{\mathcal{V}_1} u_1 + \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2} u_1) \tag{30}$$

which only depends on the marginal, conditional distribution of u_1.

Hereafter referred to as the heterarchical altruism equation, equation (30) follows naturally from the symmetric definition of the heterarchical estimation problem. Using (26), (30) can be rewritten as

$$\hat{u}^m = \frac{1}{2}[\mathbb{E}(u_1 \mid u_1 < \hat{u}^m, \mid \boldsymbol{Z}) + \mathbb{E}(u_1 \mid u_1 > \hat{u}^m, \mid \boldsymbol{Z})] \tag{31}$$

revealing its dependence on the truncated distribution of u_1 given $\boldsymbol{Z}$ [10, Chapter 22]. We will use this equation in solving the Gaussian case (Section IV).

Finally, although the cost function can have an unbounded domain and is not everywhere differentiable, we can still say something about the existence of globally-optimal solutions. We do this in the next theorem, proven in Appendix A. In passing, we note that a similar theorem, related to the general case albeit assuming a compact parameter domain, is proven in [8, pp. 651-652].

Theorem 1. *There exists at least one globally-optimal heterarchical solution. All such solutions satisfy (27), and their (identical) cost is smaller than the MMSE.*

C. Hierarchical Altruistic Estimation

Recall that, in this case, the first hierarchical estimator is the MMSEE, so that only the second estimator needs to be found. Setting the value of the gradient of J_Z with respect to the second estimate [expressed in (15)] to zero yields

$$\hat{\boldsymbol{\theta}}^{(2)} = \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2} \boldsymbol{\theta} \tag{32}$$

rendering the second hierarchical estimate locally MMSE-optimal with respect to its Voronoi region.

To calculate $\hat{u}^m$ for the hierarchical estimates we use (20) in (32) to yield

$$\hat{\boldsymbol{\theta}}^{(2)} = \boldsymbol{\tau}^T\, \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2} \boldsymbol{u}. \tag{33}$$

Using (21) in (25) and substituting (33) yields

$$\hat{u}^m = \frac{1}{2}[\boldsymbol{\tau}\, \mathbb{E}(\boldsymbol{\theta} \mid \boldsymbol{Z}) + \boldsymbol{\tau}\hat{\boldsymbol{\theta}}^{(2)}]^T \boldsymbol{e}_{\boldsymbol{u_1}} = \frac{1}{2}[\mathbb{E}(u_1 \mid \boldsymbol{Z}) + \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2} u_1]. \tag{34}$$

Equation (34), which is a scalar equation, is referred to as the hierarchical altruism equation. Notice that the first expectation on the RHS of (34) is with respect to the entire sample space Θ (conditioned on Z), which follows naturally from the fact that the first hierarchical estimator is the MMSEE.

Using (26), (34) can be rewritten as

$$\hat{u}^m = \frac{1}{2}[\mathbb{E}(u_1 \mid \boldsymbol{Z}) + \mathbb{E}(u_1 \mid u_1 > \hat{u}^m, \mid \boldsymbol{Z})], \tag{35}$$

revealing its dependence on the truncated conditional distribution of u_1 [10, Chapter 22].

We conclude with the next theorem, the proof of which is deferred to Appendix B.

Theorem 2. *There exists at least one globally-optimal hierarchical solution. All such solutions satisfy (32), and their (identical) cost is smaller than the MMSE.*

IV. The Gaussian Case

In this section we assume that the parameter and the measurements are jointly Gaussian distributed, so that:

$$\boldsymbol{\theta} \mid \boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\theta} \mid \boldsymbol{Z}, \boldsymbol{\mu}(\boldsymbol{Z}), \boldsymbol{R}(\boldsymbol{Z})) \tag{36}$$

where $\boldsymbol{\mu}$, the conditional mean, is the MMSEE. The conditional covariance matrix $\boldsymbol{R}$ is assumed to be positive definite. Let the eigenvalues of $\boldsymbol{R}$ be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and let their corresponding unit-norm eigenvectors be $\boldsymbol{v}_{\lambda_1}, \boldsymbol{v}_{\lambda_2}, \ldots, \boldsymbol{v}_{\lambda_n}$. We further denote the conditional mean of $\boldsymbol{u}$ [as defined in (20)] and its components as

$$\boldsymbol{\mu}_{\boldsymbol{u}} \triangleq \mathbb{E}(\boldsymbol{u} \mid \boldsymbol{Z}) \equiv (\mu_{u_1}, \mu_{u_2}, \ldots \mu_{u_n}). \tag{37}$$

We next derive the optimal solution for each of the altruistic estimation approaches.

A. Heterarchical Altruistic Estimation

Theorem 3. *In the Gaussian case, the optimal altruistic heterarchical estimates are*

$$\hat{\boldsymbol{\theta}}^{(1)}_{HT} = \boldsymbol{\mu} + \sqrt{\frac{2\lambda_1}{\pi}}\boldsymbol{v}_{\lambda_1} \tag{38a}$$

$$\hat{\boldsymbol{\theta}}^{(2)}_{HT} = \boldsymbol{\mu} - \sqrt{\frac{2\lambda_1}{\pi}}\boldsymbol{v}_{\lambda_1} \tag{38b}$$

and their estimation error covariances are identical to that of the MMSEE.

Proof. We begin by stating the following proposition, for the proof of which the reader is referred to Appendix C.

Proposition 1. *In the Gaussian case, the unique solution to the heterarchical altruism equation (30) is*

$$\hat{u}^m = \mu_{u_1}. \tag{39}$$

Using (39) in (26) and noting the symmetry of the Gaussian distribution about its mean yields

$$\mathbb{P}_{\mathrm{z}}(\mathcal{V}_1) = \Pr(u_1 < \mu_{u_1} \mid \boldsymbol{Z}) = \mathbb{P}_{\mathrm{z}}(\mathcal{V}_2) = \frac{1}{2} \tag{40}$$

which is a manifestation of the heterarchy in our problem. Using (29) and the law of total probability yields

$$\frac{\hat{\boldsymbol{\theta}}^{(1)} + \hat{\boldsymbol{\theta}}^{(2)}}{2} = \boldsymbol{\tau}^T(\mathbb{E}^{\mathcal{V}_1}_{\mathrm{z}}\, \boldsymbol{u}\, \mathbb{P}_{\mathrm{z}}(\mathcal{V}_1) + \mathbb{E}^{\mathcal{V}_2}_{\mathrm{z}}\, \boldsymbol{u}\, \mathbb{P}_{\mathrm{z}}(\mathcal{V}_2)) = \boldsymbol{\tau}^T\, \mathbb{E}(\boldsymbol{u} \mid \boldsymbol{Z}) = \boldsymbol{\mu}, \tag{41}$$

identifying the mid-point between the two estimates as the MMSE estimate. It follows that

$$\Delta\hat{\boldsymbol{\theta}} = 2(\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\mu}) = -2(\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\mu}) \tag{42}$$

which facilitates parameterizing the problem in terms of $\Delta\hat{\boldsymbol{\theta}}$, thus reducing the problem's degrees of freedom by half. The cost function can, therefore, be recast as

$$J_Z = \mathbb{E}(\|\boldsymbol{\mu} - \frac{1}{2}\Delta\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \wedge \|\boldsymbol{\mu} + \frac{1}{2}\Delta\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}). \tag{43}$$

Manipulating (43) results in

$$\begin{aligned} J_Z &= \mathbb{E}[\|\boldsymbol{\mu}\|^2 + \frac{1}{4}\|\Delta\hat{\boldsymbol{\theta}}\|^2 + \|\boldsymbol{\theta}\|^2 - 2\langle\boldsymbol{\mu}, \boldsymbol{\theta}\rangle + (\Delta\hat{\boldsymbol{\theta}}^T(\boldsymbol{\theta} - \boldsymbol{\mu}) \wedge \Delta\hat{\boldsymbol{\theta}}^T(\boldsymbol{\mu} - \boldsymbol{\theta})) \mid \boldsymbol{Z}] \\ &= \operatorname{tr} \boldsymbol{R} + \frac{1}{4}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \mathbb{E}(\left|\Delta\hat{\boldsymbol{\theta}}^T(\boldsymbol{\theta} - \boldsymbol{\mu})\right| \mid \boldsymbol{Z}) \end{aligned} \tag{44}$$

where we have used the fact that $\min(-a, a) = -|a|$ for any $a \in \mathbb{R}$. Substituting

$$\Delta\hat{\boldsymbol{\theta}}^T(\boldsymbol{\theta} - \boldsymbol{\mu}) = \Delta\hat{\boldsymbol{\theta}}^T\boldsymbol{\tau}^T\boldsymbol{\tau}(\boldsymbol{\theta} - \boldsymbol{\mu}) = (\boldsymbol{\tau}\Delta\hat{\theta})^T(u - \mu_u) = \|\Delta\hat{\theta}\|(u_1 - \mu_{u_1}) \tag{45}$$

in (44) yields

$$J_Z = \operatorname{tr} \boldsymbol{R} + \frac{1}{4}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \|\Delta\hat{\boldsymbol{\theta}}\|\,\mathbb{E}(|u_1 - \mu_{u_1}| \mid \boldsymbol{Z}). \tag{46}$$

Explicitly expressing the central absolute first moment of the Gaussian variable u_1 in (46) [11] yields

$$J_Z = \operatorname{tr} \boldsymbol{R} + \frac{1}{4}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \|\Delta\hat{\boldsymbol{\theta}}\|\sqrt{\frac{2R_{u_1}}{\pi}} \tag{47}$$

with R_{u_1} being the conditional variance of $u_1 \mid \boldsymbol{Z}$. The term R_{u_1} depends only on the direction of $\Delta\hat{\boldsymbol{\theta}}$ (and not on its norm), because the rotation matrix $\boldsymbol{\tau}$ that maps $\boldsymbol{\theta}$ into $\boldsymbol{u}$ is a function of that direction only. This observation, then, means that (47) is a parametrization of J_Z in terms of the norm and argument of $\Delta\hat{\boldsymbol{\theta}}$. We thus proceed with finding the optimal norm first. Differentiating J_Z with respect to $\|\Delta\hat{\boldsymbol{\theta}}\|$ and setting the derivative to zero yields:

$$\|\Delta\hat{\boldsymbol{\theta}}\| = 2\sqrt{\frac{2R_{u_1}}{\pi}}. \tag{48}$$

Substituting (48) into (47) yields

$$J_Z\Big|_{\|\Delta\hat{\boldsymbol{\theta}}\|=2\sqrt{\frac{2R_{u_1}}{\pi}}} = \operatorname{tr} \boldsymbol{R} - \frac{2R_{u_1}}{\pi}. \tag{49}$$

Therefore, minimizing J_Z is equivalent to solving

$$\max_{\Delta\hat{\boldsymbol{\theta}}} R_{u_1} \quad \text{such that} \quad \|\Delta\hat{\boldsymbol{\theta}}\| = 2\sqrt{\frac{2R_{u_1}}{\pi}}. \tag{50}$$

To do that we write

$$R_{u_1} = \boldsymbol{e}_{\boldsymbol{\theta_1}}^T(\boldsymbol{\tau}\boldsymbol{R}\boldsymbol{\tau}^T)\boldsymbol{e}_{\boldsymbol{\theta_1}} = \frac{\Delta\hat{\boldsymbol{\theta}}^T}{\|\Delta\hat{\boldsymbol{\theta}}\|}\boldsymbol{R}\frac{\Delta\hat{\boldsymbol{\theta}}}{\|\Delta\hat{\boldsymbol{\theta}}\|} \tag{51}$$

where $\boldsymbol{e}_{\theta_1}$ is the unit vector along the first standard basis vector, so that the maximization problem becomes

$$\max_{\Delta\hat{\boldsymbol{\theta}}} \frac{\Delta\hat{\boldsymbol{\theta}}^T\boldsymbol{R}\Delta\hat{\boldsymbol{\theta}}}{\Delta\hat{\boldsymbol{\theta}}^T\Delta\hat{\boldsymbol{\theta}}}. \tag{52}$$

According to the Rayleigh-Ritz theorem [12], the maximum in (52) is λ_1, the largest eigenvalue of $\boldsymbol{R}$, and it is reached for $\Delta\hat{\boldsymbol{\theta}}$ that is collinear with the eigenvector $\boldsymbol{v}_{\lambda_1}$ of $\boldsymbol{R}$ corresponding to λ_1. Thus, using (48), we have

$$\Delta\hat{\boldsymbol{\theta}} = 2\sqrt{\frac{2\lambda_1}{\pi}}\boldsymbol{v}_{\lambda_1} \tag{53}$$

which, with (42), then yields (38). Moreover, using (49), the cost obtained by using the candidate heterarchical estimators is

$$J_{\text{HT}} = \operatorname{tr} \boldsymbol{R} - \frac{2}{\pi}\lambda_1 \tag{54}$$

which is identical among all candidate solutions and independent of $\boldsymbol{v}_{\lambda_1}$. Combining this fact with Theorem 1, which states that the candidate solutions include at least one global solution, we conclude that all candidate solutions are global minimizers.

Finally, because $\hat{\boldsymbol{\theta}}_{\text{HT}}^{(1)}$, $\hat{\boldsymbol{\theta}}_{\text{HT}}^{(2)}$ differ from the MMSEE by a deterministic constant (given the measurements), their estimation error covariances are identical to that of the MMSEE. □

In passing, we observe that, as J_{HT} depends only on $\boldsymbol{R}$, then, for the optimal estimators, $J = J_Z = J_{\text{HT}}$. Also, as the solution requires only the computation of the largest eigenvalue and its corresponding eigenvector, efficient numerical algorithms, such as the power method, can be used in real-time applications involving high dimensionality.

B. Hierarchical Altruistic Estimation

Theorem 4. *In the Gaussian case, letting the first altruistic hierarchical estimate be*

$$\hat{\boldsymbol{\theta}}_{HI}^{(1)} = \hat{\boldsymbol{\theta}}_{MS} = \boldsymbol{\mu} \tag{55}$$

the optimal second estimate is

$$\hat{\boldsymbol{\theta}}_{HI}^{(2)} = \boldsymbol{\mu} + w_{HI}\sqrt{\lambda_1}\boldsymbol{v}_{\lambda_1} \tag{56}$$

where w_{HI} is defined such that $\chi = \frac{1}{2}w_{HI}$ is the unique solution to

$$\frac{\phi(\chi)}{2[1-\Phi(\chi)]} - \chi = 0. \tag{57}$$

In (57), ϕ *and* Φ *are the standard Gaussian probability density and cumulative distribution functions, respectively. The estimation error covariance of* $\hat{\boldsymbol{\theta}}_{HI}^{(2)}$ *is identical to that of the MMSEE.*

Remark 5. *The approximate value of* w_{HI} *is* 1.224 *(see Appendix C). Notice that* w_{HI}*, which can be referred to as the standardized hierarchical shift of* $\hat{\boldsymbol{\theta}}_{HI}^{(2)}$ *from the conditional distribution mode (as it is the shift for the standardized case where* $\lambda_1 = 1$*), is bigger than its analogous heterarchical standardized shift,* $\sqrt{\frac{2}{\pi}}$*, as per Theorem 3. This shift difference can be explained by observing that in the hierarchical approach the first estimate is the mode, rendering a bigger shift from it probabilistically beneficial, compared with the heterarchical approach, where both estimates are already shifted from the mode in opposite directions.*

Remark 6. *In* (56)*, the sign of the second term on the RHS is arbitrary, because the sign of the eigenvector* $\boldsymbol{v}_{\lambda_1}$ *is arbitrary.*

Proof. Using (55) in (16) yields

$$\hat{\boldsymbol{\theta}}^{(2)} = \boldsymbol{\mu} + \Delta\hat{\boldsymbol{\theta}} \tag{58}$$

so that the cost function can be written as

$$J_Z = \mathbb{E}(\|\boldsymbol{\mu} - \boldsymbol{\theta}\|^2 \wedge \|\boldsymbol{\mu} + \Delta\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}). \tag{59}$$

Manipulating (59) yields

$$\begin{aligned} J_Z &= \operatorname{tr}\boldsymbol{R} + \frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \Delta\hat{\boldsymbol{\theta}}^T\,\mathbb{E}(\boldsymbol{\theta} - \boldsymbol{\mu} \mid \boldsymbol{Z}) \\ &\quad + \mathbb{E}\{[-\frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|^2 + \Delta\hat{\boldsymbol{\theta}}^T(\boldsymbol{\theta} - \boldsymbol{\mu})] \wedge [\frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \Delta\hat{\boldsymbol{\theta}}^T(\boldsymbol{\theta} - \boldsymbol{\mu})] \mid \boldsymbol{Z}\} \\ &= \operatorname{tr}\boldsymbol{R} + \frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \mathbb{E}(\left|\Delta\hat{\boldsymbol{\theta}}^T(\boldsymbol{\theta} - \boldsymbol{\mu}) - \frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|^2\right| \mid \boldsymbol{Z}) \end{aligned} \tag{60}$$

where we have used $\mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu}) \mid \boldsymbol{Z}] = 0$. Using (45) in (60) yields

$$\begin{aligned} J_Z &= \operatorname{tr}\boldsymbol{R} + \frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \|\Delta\hat{\boldsymbol{\theta}}\|\,\mathbb{E}(\left|(u_1 - \mu_{u_1}) - \frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|\right| \mid \boldsymbol{Z}) \\ &= \operatorname{tr}\boldsymbol{R} + \frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|^2 - \|\Delta\hat{\boldsymbol{\theta}}\|[-\frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|(1 - 2\Phi(\frac{\|\Delta\hat{\boldsymbol{\theta}}\|}{2\sqrt{R_{u_1}}})) + 2\sqrt{R_{u_1}}\phi(\frac{\|\Delta\hat{\boldsymbol{\theta}}\|}{2\sqrt{R_{u_1}}})]. \end{aligned} \tag{61}$$

Equation (61) is obtained by introducing $\varphi \triangleq u_1 - (\mu_{u_1} + \frac{1}{2}\|\Delta\hat{\boldsymbol{\theta}}\|)$, and calculating the conditional mean of the folded variable $\mathbb{E}(|\varphi| \mid \boldsymbol{Z})$ [11].

To proceed with the minimization of J_Z, we need to calculate $\|\Delta\hat{\boldsymbol{\theta}}\|$. To do that, we need to first solve the hierarchical altruism equation (35). It turns out that this equation has no analytical solution, even in the Gaussian case. Using w_{HI}, which was defined implicitly in the Theorem, yields its unique solution (see Appendix C) as

$$\hat{u}^m = \mu_{u_1} + \frac{1}{2}w_{\mathrm{HI}}\sqrt{R_{u_1}}. \tag{62}$$

To calculate $\|\Delta\hat{\boldsymbol{\theta}}\|$ we use the rotation transformation $\boldsymbol{\tau}$:

$$\|\Delta\hat{\boldsymbol{\theta}}\| = \|\boldsymbol{\tau}\Delta\hat{\boldsymbol{\theta}}\| = \|\Delta\hat{\boldsymbol{u}}\|. \tag{63}$$

In the transformed parameter space, both transformed estimates reside along the solution axis $\boldsymbol{e}_{\boldsymbol{u}_1}$, such that $\hat{u}^m$ is the midpoint between them along that axis. Hence

$$\|\Delta\hat{\boldsymbol{u}}\| = \Delta\hat{u}_1 = 2(\hat{u}^m - \mu_{u_1}) \tag{64}$$

where $\Delta\hat{u}_1$ is the component of $\Delta\hat{\boldsymbol{u}}$ along $\boldsymbol{e}_{\boldsymbol{u}_1}$. Using (62) then yields

$$\|\Delta\hat{\boldsymbol{\theta}}\| = w_{\mathrm{HI}}\sqrt{R_{u_1}}. \tag{65}$$

Using (65) in (61) yields

$$\begin{aligned} J_Z\Big|_{\|\Delta\hat{\boldsymbol{\theta}}\| = w_{\mathrm{HI}}\sqrt{R_{u_1}}} &= \operatorname{tr}\boldsymbol{R} - [2\phi(\frac{w_{\mathrm{HI}}}{2}) - w_{\mathrm{HI}}(1 - \Phi(\frac{w_{\mathrm{HI}}}{2}))]w_{\mathrm{HI}}R_{u_1} \\ &= \operatorname{tr}\boldsymbol{R} - \phi(\frac{w_{\mathrm{HI}}}{2})w_{\mathrm{HI}}R_{u_1} \end{aligned} \tag{66}$$

where the last equality results from using (138) with (139a) (see Appendix C). Moreover, as analytically proved in Appendix C, $\frac{1}{2}w_{\mathrm{HI}} \in (0, \sqrt{3})$, hence $\phi(\frac{w_{\mathrm{HI}}}{2})w_{\mathrm{HI}} > 0$. Thus, (66) leads to a maximization problem identical to (52), obtained in the heterarchical problem. Adopting the solution of that problem (together with the definition of w_{HI}) yields (56). Moreover, the cost obtained from using the candidate hierarchical estimates is

$$J_{\text{HI}} = \operatorname{tr} \boldsymbol{R} - \phi(\frac{w_{\text{HI}}}{2}) w_{\text{HI}} \lambda_1 \tag{67}$$

and that cost is identical among all candidate solutions, and independent of $\boldsymbol{v}_{\lambda_1}$. Combining this observation with Theorem 2, which states that the candidate solutions include at least one global solution, renders all candidate solutions global minimizers.

Finally, because $\hat{\boldsymbol{\theta}}_{\text{HI}}^{(2)}$ differs from the MMSEE by a deterministic constant (given the measurements), its estimation error covariance is identical to that of the MMSEE. □

Similarly to the heterarchical problem, here too only the largest eigenvalue and its corresponding eigenvector need to be calculated.

C. Cost Reduction

To assess the benefit of the altruistic cooperative methodology, we compare its achievable MSE cost (1) to the MMSE baseline cost achieved by using two identical MMSE estimates,

$$J_{\text{MS}} \triangleq \mathbb{E} \| \hat{\boldsymbol{\theta}}_{\text{MS}} - \boldsymbol{\theta} \|^2. \tag{68}$$

In the Gaussian case $J_{\text{MS}} = \operatorname{tr} \boldsymbol{R}$.

For both approaches we define the relative cost reduction as

$$\Upsilon_{\text{MTHD}} \triangleq 1 - \frac{J_{\text{MTHD}}}{J_{\text{MS}}}, \quad \text{MTHD} = \text{HT or HI}. \tag{69}$$

In the heterarchical approach, (54) yields

$$\Upsilon_{\text{HT}} = \frac{\frac{2}{\pi}\lambda_1}{\sum_{i=1}^{n} \lambda_i}. \tag{70}$$

Since $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then, for a given λ_1,

$$\sup_{\lambda_2,\ldots,\lambda_n} \Upsilon_{\text{HT}} = \lim_{\frac{\lambda_2}{\lambda_1} \to 0} \Upsilon_{\text{HT}} = \frac{2}{\pi} \tag{71}$$

and

$$\min_{\lambda_2,\ldots,\lambda_n} \Upsilon_{\text{HT}} = \Upsilon_{\text{HT}} \big|_{\lambda_2 = \cdots = \lambda_n = \lambda_1} = \frac{2}{n\pi} \tag{72}$$

which gives

$$\frac{2}{n\pi} \leq \Upsilon_{\text{HT}} < \frac{2}{\pi}. \tag{73}$$

In the hierarchical case,

$$\Upsilon_{\text{HI}} = \frac{w_{\text{HI}} \phi(\frac{w_{\text{HI}}}{2}) \lambda_1}{\sum_{i=1}^{n} \lambda_i} \tag{74}$$

whence

$$\frac{w_{\text{HI}} \phi(\frac{w_{\text{HI}}}{2})}{n} \leq \Upsilon_{\text{HI}} < w_{\text{HI}} \phi(\frac{w_{\text{HI}}}{2}). \tag{75}$$

Notice that in both approaches the best achievable relative reduction corresponds to $\frac{\lambda_1}{\lambda_2} \to \infty$, whereas the worst achievable reduction corresponds to $\lambda_1 = \lambda_2 = \ldots = \lambda_n$. This is so because in both approaches the two estimates are dispersed along the eigenvector that corresponds to λ_1. Thus, the benefit gained from dispersing the estimates is biggest when the variance in that direction is largest compared with the other variances. When the variances in all directions are equal, the benefit assumes its smallest possible value. It is also noted that the benefit shrinks when the dimension of the system increases, because there are only two estimates, distributed along one direction. Nevertheless, even in high dimensional cases, if one direction dominates the others in terms of its variance, still the reduction can be significant, which means that the altruistic approaches become appealing in cases involving ill-conditioned covariance matrices (characterized by large condition numbers).

To demonstrate the effect of the problem's dimensionality on the cost function reduction, the upper and lower cost reduction bounds for each approach are depicted in Fig. 1. In a scalar problem ($n = 1$) the lower and upper bounds coincide, yielding a unique value for the reduction. At higher dimensions the best achievable gains are identical to those obtained for the scalar problem, whereas the worst achievable reductions diminish with the increasing dimension.

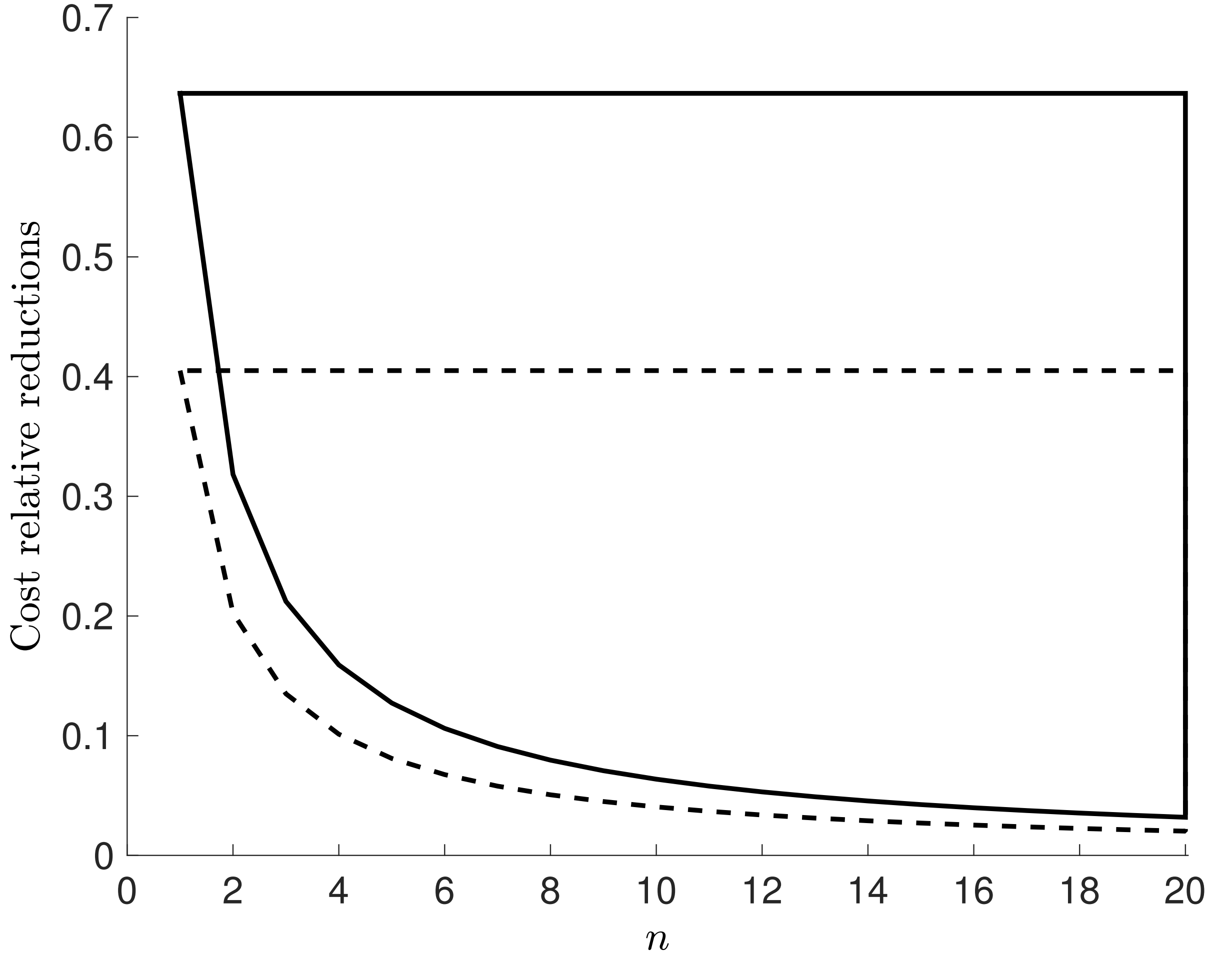

Fig. 1: Relative cost reduction bounds vs problem dimension. Solid lines: heterarchical (Υ_{HT}). Dashed lines: hierarchical (Υ_{HI}).

D. Numerical Illustration

We now numerically illustrate the behavior of the cost function J_Z, in order to demonstrate the validity of the solutions in the Gaussian case, and highlight some properties of altruistic estimation problems and their solutions. We focus on J_Z, rather than on J, because, as we have previously shown, the choice of the optimal estimators is independent of the measurements.

Let $\theta \mid \boldsymbol{Z} \sim \mathcal{N}(0, 100)$. The heterarchical and hierarchical estimates are

$$\hat{\theta}^{(1)}_{\mathrm{HT}} = 10\sqrt{\frac{2}{\pi}} \approx 7.979, \qquad \hat{\theta}^{(2)}_{\mathrm{HT}} = -\hat{\theta}^{(1)}_{\mathrm{HT}} \tag{76}$$

and

$$\hat{\theta}^{(1)}_{\mathrm{HI}} = \mu = 0, \qquad \hat{\theta}^{(2)}_{\mathrm{HI}} = \pm 10 w_{\mathrm{HI}} \approx \pm 12.240. \tag{77}$$

The costs for the MMSE (two egoistic estimates), heterarchical, and hierarchical approaches are: $J_{\mathrm{MS}} = 100$, $J_{\mathrm{HT}} \approx 36.338$, and $J_{\mathrm{HI}} \approx 59.5$, respectively.

Figure 2 shows equilevel contour lines of the cost function, computed at each node $(\hat{\theta}^{(1)}, \hat{\theta}^{(2)})$ of a 500×500 grid of the two estimates. The cost is approximated as the mean of 10^5 samples drawn from the given parameter conditional distribution. The contours are distributed logarithmically, so that they are denser around lower values of J_Z. The optimal heterarchical and hierarchical estimates, (76) and (77), respectively, are superimposed on the contour plot as squares and circles, respectively. The figure exhibits the tendency of J_Z to infinity when both estimates tend to infinity in absolute values (as per Lemma 1 in A). On the other hand, when one of the estimates tends to infinity in absolute value and the other estimate remains finite (which renders the infinite estimate irrelevant in the computation of the cost), the lowest value of J_Z is the MMSE estimate, which is achieved when the finite estimate is the MMSEE.

Figure 2 exhibits two reflection symmetries of the function J_Z, 1) about the mirror line $\hat{\theta}^{(1)} = \hat{\theta}^{(2)}$, and 2) about the normal to the line $\hat{\theta}^{(1)} = \hat{\theta}^{(2)}$ through the origin. The first symmetry expresses the symmetric nature of J_Z with respect to

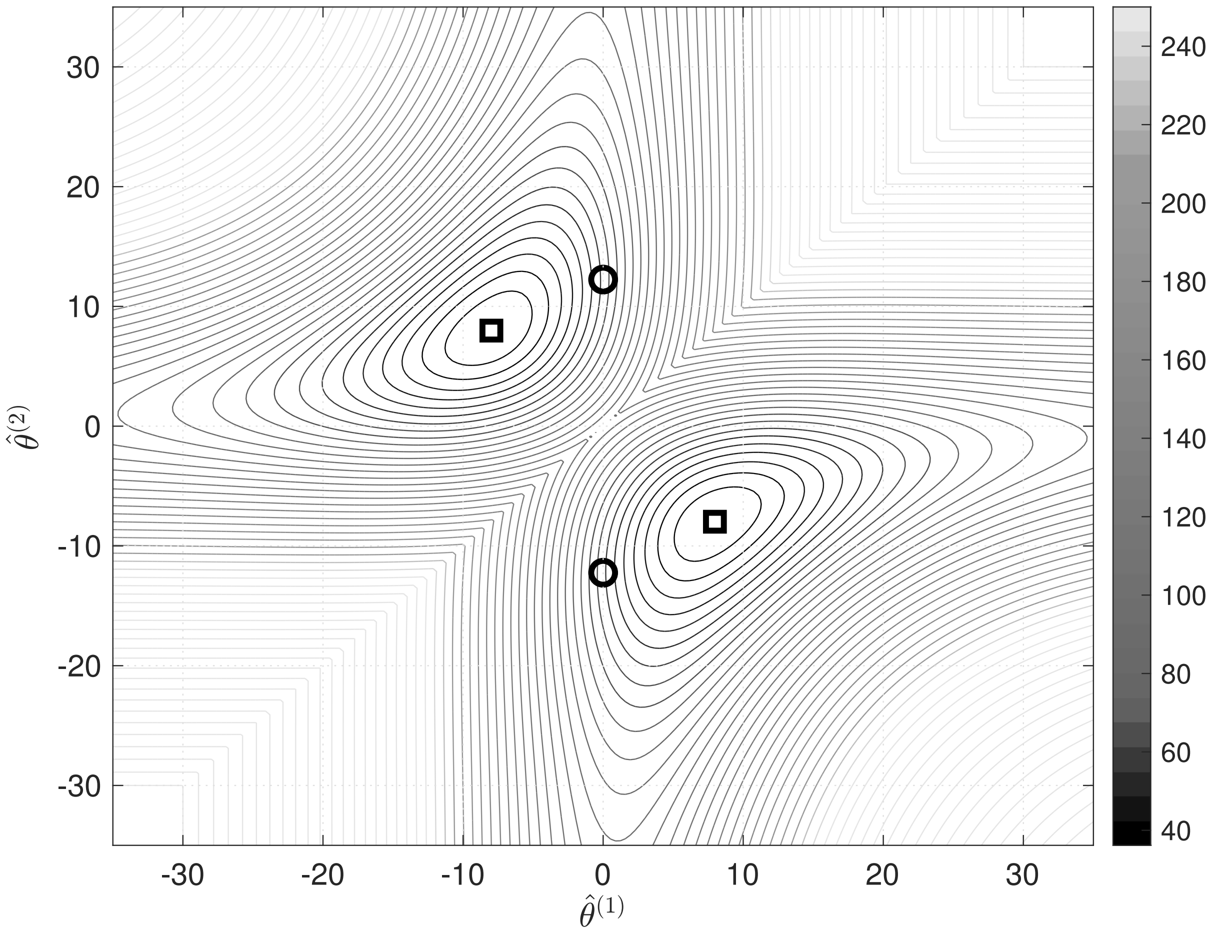

Fig. 2: Equilevel cost contours of J_Z in the 1-D Gaussian example of subsection IV-D. Square markers: heterarchical solutions. Circle markers: hierarchical solutions.

its arguments (the two estimates). The second expresses the symmetric nature of the Gaussian distribution. The figure also clearly demonstrates the non-differentiability of J_Z on the reflection axis $\hat{\theta}^{(1)} = \hat{\theta}^{(2)}$. The figure exhibits two (reflective) global minima of the function J_Z, precisely at the analytically calculated heterarchical estimates (76), with cost agreeing with the analytically computed cost. The two (reflective) optimal hierarchical estimates of the second hierarchical estimator are positioned at the minima of J_Z along the constraint line $\hat{\theta}^{(1)} = 0$, which coincide with the analytically calculated hierarchical estimates (77), with cost agreeing with the analytically computed cost. As could be expected, at both minima the J_Z contour lines are tangent to the constraint.

V. Conclusions

We have proposed an estimation methodology for optimal cooperation between two information-sharing agents, that is based on the notion of altruism. The methodology is suited for scenarios that can benefit from the existence of two opportunities to estimate, such as scenarios involving two cooperating agents that have one global mission that is accomplished even if only one of the agents provides a satisfactory estimate (using its own, local estimator). In the proposed approach, the two agents do not yield an identical optimal estimate, but, rather, at least one of them sacrifices its own estimation performance by providing a sub-optimal estimate. The benefit of the proposed scheme is an improvement in the overall estimation performance, measured by a global mean squared error criterion.

Two approaches of altruistic cooperation are proposed. In the heterarchical approach, both estimators are altruistic, which yields two sub-optimal estimates that are different than the (egoistic) MMSE estimate. In the hierarchical approach the first agent maximizes its performance egoistically without considering the other estimator, thus computing the MMSE estimate; the second agent maximizes the global performance measure while taking into account the presence of the first (MMSE-optimal) agent. Implicit and coupled equations are derived for the design of the estimators in both approaches. In the Gaussian case, explicit optimal solutions are provided, that require, in both approaches, only the calculation of the largest eigenvalue of the conditional covariance matrix of the parameter and its corresponding eigenvector. These results can also be viewed as analytical solutions to two well-known Voronoi serviceability problems in the two-facility case.

In the Gaussian case, it is shown that the improvement in the overall performance (relative to naive MMSE estimation) depends on the dimension of the problem and on the spread of the spectrum of the conditional covariance matrix. In general, the larger the dimension of the problem, the smaller the improvement that can be expected using the proposed cooperative estimation approach. On the other hand, the proposed altruistic approaches are especially appealing in (even high-dimensional) ill-conditioned estimation problems.

Appendix A
Proof of Theorem 1

Proof. Following the reasoning of the estimator derivation in Section III, the proof follows from minimizing the measurement-conditioned cost, J_Z, defined in (5). Let $\hat{\boldsymbol{\theta}} \in (\Theta \times \Theta)$ and $J_Z(\hat{\boldsymbol{\theta}})$ denote the augmented estimate vector $[(\hat{\boldsymbol{\theta}}^{(1)})^T, (\hat{\boldsymbol{\theta}}^{(2)})^T]^T$ and the value of the cost (5) computed with the pair of estimates comprising $\hat{\boldsymbol{\theta}}$, respectively.

For the reader's convenience, we first provide an overview of the proof's main stages:

1) We tessellate the augmented estimate vector space $(\Theta \times \Theta)$ into two parts: an internal part (where $\hat{\boldsymbol{\theta}}$ is bounded), and its complementary external part.
2) We prove Lemma 1, showing that choosing the internal part to be large enough guarantees that for any point $\hat{\boldsymbol{\theta}}$ in the external part, the cost $J_Z(\hat{\boldsymbol{\theta}})$ is arbitrarily close to the MMSE.
3) We prove Lemma 2, showing that choosing the internal part to be sufficiently large guarantees that there is at least one minimum point (satisfying (27)) in the internal part, for which the cost $J_Z(\hat{\boldsymbol{\theta}})$ is strictly smaller than the MMSE.
4) We build a final internal part, which is sufficiently large to guarantee that any point in the external part yields a cost strictly higher than that of any minimum of the internal part.

Moving on to the proof itself, define

$$\mathcal{S}^a_{\hat{\boldsymbol{\theta}}} \triangleq \{\hat{\boldsymbol{\theta}} \mid \|\hat{\boldsymbol{\theta}}^{(1)}\| \vee \|\hat{\boldsymbol{\theta}}^{(2)}\| \leq a\}, \quad a \in \mathbb{R}_{>0} \tag{78}$$

where $\alpha \vee \beta \triangleq \max(\alpha, \beta)$ for some $\alpha, \beta \in \mathbb{R}$. We begin by proving the following two lemmas.

Lemma 1. *For any $\epsilon \in \mathbb{R}_{>0}$ there exists $a(\epsilon) \in \mathbb{R}_{>0}$ such that if $\hat{\boldsymbol{\theta}} \notin \mathcal{S}^{a(\epsilon)}_{\hat{\boldsymbol{\theta}}}$ then $J_Z(\hat{\boldsymbol{\theta}}) > J_{MS} - \epsilon$.*

Proof. We begin with a brief overview of the proof of this lemma:

1) Choosing some (temporary) edge $a_0(\epsilon)$ to the internal part $\mathcal{S}^{a(\epsilon)}_{\hat{\boldsymbol{\theta}}}$, we show in Proposition 2 that when the norms of both estimates are beyond $a_0(\epsilon)$, the lemma is satisfied.
2) Next, we show in Proposition 3 that if the distance between the norms of both estimates is large enough, the lemma is satisfied.
3) Finally, we choose a (bigger) edge $a(\epsilon)$ for which, if $\hat{\boldsymbol{\theta}} \notin \mathcal{S}^{a(\epsilon)}_{\hat{\boldsymbol{\theta}}}$, then either of the conditions of Proposition 2 or Proposition 3 must be satisfied.

Using the law of total probability and the triangle inequality in (5) yields, for any $r \in \mathbb{R}_{>0}$,

$$\begin{aligned} J_Z(\hat{\boldsymbol{\theta}}) &= \mathbb{E}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 \wedge \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}, \|\boldsymbol{\theta}\| \leq r) \Pr(\|\boldsymbol{\theta}\| \leq r \mid \boldsymbol{Z}) \\ &\quad + \mathbb{E}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 \wedge \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}, \|\boldsymbol{\theta}\| > r) \Pr(\|\boldsymbol{\theta}\| > r \mid \boldsymbol{Z}) \\ &\geq \mathbb{E}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 \wedge \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}, \|\boldsymbol{\theta}\| \leq r) \Pr(\|\boldsymbol{\theta}\| \leq r \mid \boldsymbol{Z}) \\ &\geq \mathbb{E}[(\|\hat{\boldsymbol{\theta}}^{(1)}\| - \|\boldsymbol{\theta}\|)^2 \wedge (\|\hat{\boldsymbol{\theta}}^{(2)}\| - \|\boldsymbol{\theta}\|)^2 \mid \boldsymbol{Z}, \|\boldsymbol{\theta}\| \leq r] \\ &\quad \times \Pr(\|\boldsymbol{\theta}\| \leq r \mid \boldsymbol{Z}). \end{aligned} \tag{79}$$

We now define

$$r \triangleq \underset{\rho \in \mathbb{R}_{>0}}{\arg} \{\Pr(\|\boldsymbol{\theta}\| \leq \rho \mid \boldsymbol{Z}) = \frac{1}{2}\}. \tag{80}$$

Also, observing that $\epsilon < J_{\text{MS}}$ (as otherwise the lemma holds trivially since $J_Z > 0$), and using the assumption that the second moment of the joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{Z}$ is finite, we set

$$a_0(\epsilon) \triangleq 1 + r + \sqrt{2(J_{\text{MS}} - \epsilon)}. \tag{81}$$

We first state and prove the following propositions.

Proposition 2. *If both $\|\hat{\boldsymbol{\theta}}^{(1)}\| > a_0(\epsilon)$ and $\|\hat{\boldsymbol{\theta}}^{(2)}\| > a_0(\epsilon)$, then $J_Z(\hat{\boldsymbol{\theta}}) > J_{MS} - \epsilon$.*

Proof. Employing definitions (80) and (81) in (79) yields

$$\begin{aligned} J_Z(\hat{\boldsymbol{\theta}}) &\geq [(\|\hat{\boldsymbol{\theta}}^{(1)}\| - r)^2 \wedge (\|\hat{\boldsymbol{\theta}}^{(2)}\| - r)^2] \Pr(\|\boldsymbol{\theta}\| \leq r \mid \boldsymbol{Z}) \\ &> [a_0(\epsilon) - r]^2 \Pr(\|\boldsymbol{\theta}\| \leq r \mid \boldsymbol{Z}) \\ &= \frac{1}{2}[1 + \sqrt{2(J_{\text{MS}} - \epsilon)}]^2 > J_{\text{MS}} - \epsilon. \end{aligned} \tag{82}$$

□

Prior to presenting the next proposition we now assume that $\|\hat{\boldsymbol{\theta}}^{(2)}\| \geq \|\hat{\boldsymbol{\theta}}^{(1)}\|$ and define $\mathcal{D} \triangleq \|\hat{\boldsymbol{\theta}}^{(2)}\| - \|\hat{\boldsymbol{\theta}}^{(1)}\|$.

Proposition 3. *For any number $\epsilon \in \mathbb{R}_{>0}$ there exists a number $L(\epsilon) \in \mathbb{R}_{>0}$ such that if $\mathcal{D} > L(\epsilon)$ then $J_Z > J_{MS} - \epsilon$.*

Proof. Define $\varphi(\hat{\boldsymbol{\theta}}^{(1)}) \triangleq \mathbb{E}[\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 \mid \boldsymbol{Z}]$. Then, since $\partial\mathcal{V}$ is a set of measure zero, the law of total probability yields

$$\varphi(\hat{\boldsymbol{\theta}}^{(1)}) = \mathbb{E}_{\text{z}}^{\mathcal{V}_1}\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2\,\mathbb{P}_{\text{z}}(\mathcal{V}_1) + \mathbb{E}_{\text{z}}^{\mathcal{V}_2}\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2\,\mathbb{P}_{\text{z}}(\mathcal{V}_2). \tag{83}$$

Equation (83) yields

$$\mathbb{E}_{\text{z}}^{\mathcal{V}_1}\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2\,\mathbb{P}_{\text{z}}(\mathcal{V}_1) = \varphi(\hat{\boldsymbol{\theta}}^{(1)}) - \mathbb{E}_{\text{z}}^{\mathcal{V}_2}\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2\,\mathbb{P}_{\text{z}}(\mathcal{V}_2). \tag{84}$$

Expressing J_Z using the law of total probability and using (84) yields

$$J_Z = \varphi(\hat{\boldsymbol{\theta}}^{(1)}) - f_1 + f_2 \tag{85}$$

where we have defined

$$f_1 \triangleq \mathbb{E}_{\text{z}}^{\mathcal{V}_2}\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2\,\mathbb{P}_{\text{z}}(\mathcal{V}_2) \tag{86}$$

$$f_2 \triangleq \mathbb{E}_{\text{z}}^{\mathcal{V}_2}\|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2\,\mathbb{P}_{\text{z}}(\mathcal{V}_2). \tag{87}$$

Clearly, $f_1 \geq 0$ and $f_2 \geq 0$. Using (7) and the monotonicity of the probability measure, we have

$$\begin{aligned} \mathbb{P}_{\text{z}}(\mathcal{V}_2) &= \Pr(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2 - \|\hat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\|^2 > 0 \mid \boldsymbol{Z}) \\ &= \Pr\{\|\hat{\boldsymbol{\theta}}^{(1)}\|^2 - \|\hat{\boldsymbol{\theta}}^{(2)}\|^2 + 2\boldsymbol{\theta}^T(\hat{\boldsymbol{\theta}}^{(2)} - \hat{\boldsymbol{\theta}}^{(1)}) > 0 \mid \boldsymbol{Z}\} \\ &\leq \Pr\{\|\hat{\boldsymbol{\theta}}^{(1)}\|^2 - \|\hat{\boldsymbol{\theta}}^{(2)}\|^2 + 2\left|\boldsymbol{\theta}^T(\hat{\boldsymbol{\theta}}^{(2)} - \hat{\boldsymbol{\theta}}^{(1)})\right| > 0 \mid \boldsymbol{Z}\} \\ &\leq \Pr\{\|\hat{\boldsymbol{\theta}}^{(1)}\|^2 - \|\hat{\boldsymbol{\theta}}^{(2)}\|^2 + 2\|\boldsymbol{\theta}\|\|\hat{\boldsymbol{\theta}}^{(2)} - \hat{\boldsymbol{\theta}}^{(1)}\| > 0 \mid \boldsymbol{Z}\} \end{aligned} \tag{88}$$

where the last inequality follows from the Cauchy-Schwarz inequality. Using the triangle inequality yields

$$\begin{aligned} \mathbb{P}_{\text{z}}(\mathcal{V}_2) \leq \Pr\{&\|\hat{\boldsymbol{\theta}}^{(1)}\|^2 - \|\hat{\boldsymbol{\theta}}^{(2)}\|^2 \\ &+ 2\|\boldsymbol{\theta}\|(\|\hat{\boldsymbol{\theta}}^{(1)}\| + \|\hat{\boldsymbol{\theta}}^{(2)}\|) > 0 \mid \boldsymbol{Z}\} \\ = \Pr\{&\|\hat{\boldsymbol{\theta}}^{(1)}\| - \|\hat{\boldsymbol{\theta}}^{(2)}\| + 2\|\boldsymbol{\theta}\| > 0 \mid \boldsymbol{Z}\} = \Pr\{\|\boldsymbol{\theta}\| > \frac{1}{2}\mathcal{D} \mid \boldsymbol{Z}\} \end{aligned} \tag{89}$$

showing that when $\mathcal{D} \to \infty$, $\mathcal{V}_2$ becomes a set of measure zero. Thus $\mathbb{P}_{\text{z}}(\mathcal{V}_1) \to 1$ and, hence, the LHS of (84) satisfies

$$\lim_{\mathcal{D}\to\infty} \mathbb{E}_{\text{z}}^{\mathcal{V}_1}(\|\hat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}\|^2)\,\mathbb{P}_{\text{z}}(\mathcal{V}_1) = \varphi(\hat{\boldsymbol{\theta}}^{(1)}) \tag{90}$$

yielding $\lim_{\mathcal{D}\to\infty} f_1 = 0$. By the definition (7) of the Voronoi region $\mathcal{V}_2$, $f_2 \leq f_1$, implying $\lim_{\mathcal{D}\to\infty} f_2 = 0$. Defining $f \triangleq f_1 - f_2$, (85) becomes

$$J_Z = \varphi(\hat{\boldsymbol{\theta}}^{(1)}) - f \tag{91}$$

where $f \geq 0$ and $\lim_{\mathcal{D}\to\infty} f = 0$. Since, due to the fundamental theorem of MMSE estimation, $J_{\text{MS}} \leq \varphi(\hat{\boldsymbol{\theta}}^{(1)})$, we have

$$J_Z \geq J_{\text{MS}} - f \tag{92}$$

which yields the proposition. □

Now set

$$a(\epsilon) \triangleq 1 + a_0(\epsilon) + L(\epsilon). \tag{93}$$

If $\hat{\boldsymbol{\theta}} \notin \mathcal{S}_{\hat{\boldsymbol{\theta}}}^{a(\epsilon)}$ then either

$$\|\hat{\boldsymbol{\theta}}^{(2)}\| > a(\epsilon) \text{ and } \|\hat{\boldsymbol{\theta}}^{(1)}\| \leq a(\epsilon) \tag{94}$$

or

$$\|\hat{\boldsymbol{\theta}}^{(1)}\| > a(\epsilon) \text{ and } \|\hat{\boldsymbol{\theta}}^{(2)}\| > a(\epsilon). \tag{95}$$

Assume, first, that (94) holds. If $\|\hat{\boldsymbol{\theta}}^{(1)}\| > a_0(\epsilon)$ then $\|\hat{\boldsymbol{\theta}}^{(2)}\| > a_0(\epsilon)$ and the lemma is proved based on Proposition 2. Conversely, if $\|\hat{\boldsymbol{\theta}}^{(1)}\| \le a_0(\epsilon)$, then, since $\|\hat{\boldsymbol{\theta}}^{(2)}\| > a(\epsilon)$,

$$\mathcal{D} = \|\hat{\boldsymbol{\theta}}^{(2)}\| - \|\hat{\boldsymbol{\theta}}^{(1)}\| \ge a(\epsilon) - a_0(\epsilon) = 1 + L(\epsilon) > L(\epsilon) \tag{96}$$

and the lemma follows from Proposition 3.

If (95) holds, then (93) implies that both $\|\hat{\boldsymbol{\theta}}^{(1)}\| > a_0(\epsilon)$ and $\|\hat{\boldsymbol{\theta}}^{(2)}\| > a_0(\epsilon)$, and the lemma follows from Proposition 2. □

Lemma 2. *There exists a number $b_0 \in \mathbb{R}_{>0}$ such that for any $b \ge b_0$, J_Z attains at least one minimum in $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$, at a point where $J_Z < J_{\mathrm{MS}}$. All such points satisfy (27).*

Proof. Assuming that the first moment of the joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{Z}$ is finite, let $\hat{\boldsymbol{\zeta}}^{(1)} \triangleq \hat{\boldsymbol{\theta}}_{\mathrm{MS}}$ and $\hat{\boldsymbol{\zeta}}^{(2)} \triangleq c\boldsymbol{e}_{\boldsymbol{\theta}_1}$ (recalling that $\boldsymbol{e}_{\boldsymbol{\theta}_1} \in \mathbb{R}^n$ is the unit vector along the first standard basis vector of Θ), with $c \in (0,1)$ such that $\hat{\boldsymbol{\zeta}}^{(2)} \neq \hat{\boldsymbol{\theta}}_{\mathrm{MS}}$. To show that $J_Z(\hat{\boldsymbol{\zeta}}) < J_{\mathrm{MS}}$, where $\hat{\zeta}$ comprises the estimators $\hat{\boldsymbol{\zeta}}^{(1)}$ and $\hat{\boldsymbol{\zeta}}^{(2)}$, we use the law of total probability to rewrite (5) as

$$J_Z(\hat{\boldsymbol{\zeta}}) = \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_1(\hat{\zeta})}(\|\hat{\boldsymbol{\zeta}}^{(1)} - \boldsymbol{\theta}\|^2)\,\mathbb{P}_{\mathrm{z}}(\mathcal{V}_1(\hat{\boldsymbol{\zeta}})) + \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2(\hat{\zeta})}(\|\hat{\boldsymbol{\zeta}}^{(2)} - \boldsymbol{\theta}\|^2)\,\mathbb{P}_{\mathrm{z}}(\mathcal{V}_2(\hat{\boldsymbol{\zeta}})). \tag{97}$$

Because $\boldsymbol{\theta}$ is continuous in Θ and $\hat{\boldsymbol{\zeta}}^{(2)}$ has a bounded norm, $\mathbb{P}_{\mathrm{z}}(\mathcal{V}_2(\hat{\boldsymbol{\zeta}})) > 0$. Thus, using the definition (7) of $\mathcal{V}_2$, (97) yields

$$\begin{aligned} J_Z(\hat{\boldsymbol{\zeta}}) &< \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_1(\hat{\zeta})}(\|\hat{\boldsymbol{\zeta}}^{(1)} - \boldsymbol{\theta}\|^2)\,\mathbb{P}_{\mathrm{z}}(\mathcal{V}_1(\hat{\boldsymbol{\zeta}})) \\ &\quad + \mathbb{E}_{\mathrm{z}}^{\mathcal{V}_2(\hat{\zeta})}(\|\hat{\boldsymbol{\zeta}}^{(1)} - \boldsymbol{\theta}\|^2)\,\mathbb{P}_{\mathrm{z}}(\mathcal{V}_2(\hat{\boldsymbol{\zeta}})) = \varphi(\hat{\boldsymbol{\zeta}}^{(1)}) = J_{\mathrm{MS}}. \end{aligned} \tag{98}$$

We next choose b_0 based on $\hat{\boldsymbol{\zeta}}$. Let

$$\epsilon \triangleq J_{\mathrm{MS}} - J_Z(\hat{\boldsymbol{\zeta}}). \tag{99}$$

According to (98) $\epsilon > 0$, and Lemma 1 states that there exists a number $a(\epsilon) \in \mathbb{R}_{>0}$ such that if $\hat{\boldsymbol{\theta}} \notin \mathcal{S}^{a(\epsilon)}_{\hat{\boldsymbol{\theta}}}$ then $J_Z(\hat{\boldsymbol{\theta}}) > J_{\mathrm{MS}} - \epsilon = J_Z(\hat{\boldsymbol{\zeta}})$. Set

$$b_0 \triangleq 1 + a(\epsilon) \vee \|\hat{\boldsymbol{\theta}}_{\mathrm{MS}}\| \tag{100}$$

and consider the set $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$ for any number $b \ge b_0$. Because $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$ is closed and bounded, it is compact. Since J_Z is continuous everywhere, it is necessarily continuous in $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$. Thus, according to the Weierstrass extreme value theorem, J_Z attains a minimum in at least one point in $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$ (the cost at all such points is, of course, identical). Henceforth denoting any of these minimum points as $\hat{\boldsymbol{\psi}}$, the first part of the lemma states that $J_Z(\hat{\boldsymbol{\psi}}) < J_{\mathrm{MS}}$. To show this, we notice that

$$\|\hat{\boldsymbol{\zeta}}^{(1)}\| \vee \|\hat{\boldsymbol{\zeta}}^{(2)}\| = \|\hat{\boldsymbol{\theta}}_{\mathrm{MS}}\| \vee c < b_0 \tag{101}$$

so that $\hat{\boldsymbol{\zeta}} \in (\mathcal{S}^{b_0}_{\hat{\boldsymbol{\theta}}})^\circ$, where we use the notation A° for the interior of A, and, consequently, $\hat{\boldsymbol{\zeta}} \in (\mathcal{S}^b_{\hat{\boldsymbol{\theta}}})^\circ$. Since $J_Z(\hat{\boldsymbol{\psi}}) \le J_Z(\hat{\boldsymbol{\zeta}})$, the first part of the lemma follows upon invoking (98).

Continuing to the second part of the lemma, we now prove, by contradiction, that $\hat{\boldsymbol{\psi}}$ satisfies (27). If it does not, then either 1) it is an interior point where J_Z is non-differentiable, or 2) it is located along the boundary of $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$. We contradict each case separately.

We first show that none of the minimum points can be located at an interior point where J_Z is non-differentiable. Let

$$\mathcal{Y} \triangleq \left\{\hat{\boldsymbol{\theta}} \in \Theta \times \Theta \mid \hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}^{(2)}\right\}. \tag{102}$$

Each point in $\mathcal{Y}$ is a pair of estimates for which J_Z is non-differentiable. Assume that $\hat{\boldsymbol{\psi}} \in (\mathcal{S}^b_{\hat{\boldsymbol{\theta}}})^\circ \cap \mathcal{Y}$. Then $\hat{\boldsymbol{\psi}}^{(1)} = \hat{\boldsymbol{\psi}}^{(2)}$, giving

$$J_Z(\hat{\boldsymbol{\psi}}) = \varphi(\hat{\boldsymbol{\psi}}^{(1)}). \tag{103}$$

Now set $\hat{\boldsymbol{\xi}}^{(1)} \triangleq \hat{\boldsymbol{\psi}}^{(1)}$ and $\hat{\boldsymbol{\xi}}^{(2)} \triangleq cb\boldsymbol{e}_{\boldsymbol{\theta}_1}$ with $c \in (0,1)$ such that $\hat{\boldsymbol{\xi}}^{(2)} \neq \hat{\boldsymbol{\psi}}^{(1)}$. Obviously, $\hat{\boldsymbol{\xi}} \notin \mathcal{Y}$, and

$$\|\hat{\boldsymbol{\xi}}^{(1)}\| \vee \|\hat{\boldsymbol{\xi}}^{(2)}\| = \|\hat{\boldsymbol{\psi}}^{(1)}\| \vee cb < b \tag{104}$$

so that $\hat{\boldsymbol{\xi}} \in (\mathcal{S}^b_{\hat{\boldsymbol{\theta}}})^\circ$. Analogously to (98) and using (103) we have

$$J_Z(\hat{\boldsymbol{\xi}}) < \varphi(\hat{\boldsymbol{\xi}}^{(1)}) = \varphi(\hat{\boldsymbol{\psi}}^{(1)}) = J_Z(\hat{\boldsymbol{\psi}}), \tag{105}$$

contradicting the assumption that J_Z has a minimum at $\hat{\boldsymbol{\psi}} \in (\mathcal{S}^b_{\hat{\boldsymbol{\theta}}})^\circ$.

Now assume that $\hat{\boldsymbol{\psi}}$ is located along the boundary of $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$. Then $\hat{\boldsymbol{\psi}} \notin \mathcal{S}^{a(\epsilon)}_{\hat{\boldsymbol{\theta}}}$, and, according to Lemma 1, $J_Z(\hat{\boldsymbol{\psi}}) > J_Z(\hat{\boldsymbol{\zeta}})$. Recalling that $\hat{\boldsymbol{\zeta}} \in (\mathcal{S}^b_{\hat{\boldsymbol{\theta}}})^\circ$ yields a contradiction to the assumption that $\hat{\boldsymbol{\psi}}$ is a minimum point of J_Z in $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$. □

Returning to the proof of Theorem 1, let b_0 be a number satisfying Lemma 2, and consider $b \geq b_0$. Then, according to Lemma 2, there exists at least one minimizer of J_Z in $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$, at a point that satisfies (27). Denote the value of J_Z at any such minimum point as J^*_Z (recall that the costs at all minimum points are identical). Then, according to Lemma 2,

$$J^*_Z < J_{\text{MS}}. \tag{106}$$

Define

$$\epsilon \triangleq J_{\text{MS}} - J^*_Z. \tag{107}$$

According to Lemma 1, there exists a number $a(\epsilon) > 0$ such that

$$J_Z(\hat{\boldsymbol{\theta}}) > J_{\text{MS}} - \epsilon = J^*_Z \quad \forall \hat{\boldsymbol{\theta}} \notin \mathcal{S}^{a(\epsilon)}_{\hat{\boldsymbol{\theta}}}. \tag{108}$$

Now consider the set $\mathcal{S}^c_{\hat{\boldsymbol{\theta}}}$ with

$$c \triangleq a(\epsilon) \vee b. \tag{109}$$

According to Lemma 2 and based on (109), there is at least one minimum point in $\mathcal{S}^c_{\hat{\boldsymbol{\theta}}}$, where the cost (which is identical for all minimum points) does not exceed J^*_Z because $\mathcal{S}^{a(\epsilon)}_{\hat{\boldsymbol{\theta}}} \in \mathcal{S}^c_{\hat{\boldsymbol{\theta}}}$. At any such point, thus, the cost is strictly smaller than J_{MS} due to (106). Outside of $\mathcal{S}^c_{\hat{\boldsymbol{\theta}}}$, $J_Z > J^*_Z$ due to Lemma 1 and (109). □

APPENDIX B
PROOF OF THEOREM 2

Proof. Let $\mathcal{T} \subset \Theta \times \Theta$ be defined as $\mathcal{T} \triangleq \{\hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}_{\text{MS}}\} \times \Theta$. In the hierarchical problem (3) the objective is to minimize a restriction of the cost J_Z, originally defined in (5) on $\Theta \times \Theta$, to the subdomain $\mathcal{T}$, that is

$$J_Z\restriction_{\mathcal{T}}(\hat{\boldsymbol{\theta}}^{(2)}) \triangleq J_Z(\hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}_{\text{MS}}, \hat{\boldsymbol{\theta}}^{(2)}). \tag{110}$$

As a restriction of J_Z, $J_Z\restriction_{\mathcal{T}}$ is continuous everywhere and differentiable except at $\hat{\boldsymbol{\theta}}^{(2)} = \hat{\boldsymbol{\theta}}_{\text{MS}}$.

The proof of Theorem 2 follows the proof of Theorem 1 in a restricted form. We begin with the following lemma.

Lemma 3. *For any number $\epsilon \in \mathbb{R}_{>0}$ there exists a number $a(\epsilon) \in \mathbb{R}_{>0}$ such that if $\|\hat{\boldsymbol{\theta}}^{(2)}\| > a(\epsilon)$ then $J_Z\restriction_{\mathcal{T}}(\hat{\boldsymbol{\theta}}^{(2)}) > J_{MS} - \epsilon$.*

Proof. Because $J_Z\restriction_{\mathcal{T}}$ is a restriction of J_Z, Proposition 3 applies with $\hat{\boldsymbol{\theta}}^{(1)} = \hat{\boldsymbol{\theta}}_{\text{MS}}$; hence, there exists a number $L(\epsilon) \in \mathbb{R}_{>0}$ such that if $\mathcal{D} > L(\epsilon)$ then $J_Z\restriction_{\mathcal{T}} > J_{\text{MS}} - \epsilon$. Set

$$a(\epsilon) \triangleq 1 + L(\epsilon) + \|\hat{\boldsymbol{\theta}}_{\text{MS}}\| \tag{111}$$

and assume that $\|\hat{\boldsymbol{\theta}}^{(2)}\| > a(\epsilon)$. The lemma follows upon observing that

$$\mathcal{D} \geq 1 + L(\epsilon) > L(\epsilon). \tag{112}$$

□

Lemma 4. *There exists a number $b_0 \in \mathbb{R}_{>0}$ such that for any $b \geq b_0$, $J_Z\restriction_{\mathcal{T}}$ attains a minimum in $\{\hat{\boldsymbol{\theta}}^{(2)} \in \Theta \mid \|\hat{\boldsymbol{\theta}}^{(2)}\| \leq b\}$, at a point where $J_Z\restriction_{\mathcal{T}} < J_{MS}$. This point satisfies* (32).

Proof. The proof follows the proof of Lemma 2 in restricted form, where J_Z is replaced by $J_Z\restriction_{\mathcal{T}}$, $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}}$ by its subset $\mathcal{S}^b_{\hat{\boldsymbol{\theta}}} \cap \{\|\hat{\boldsymbol{\theta}}^{(1)}\| = \|\hat{\boldsymbol{\theta}}_{\text{MS}}\|\}$, Lemma 1 by Lemma 3, and (27) by (32). □

Having these two lemmas on hand, the proof of Theorem 2 follows the proof of Theorem 1 in a restricted form, where J_Z is replaced by $J_Z\restriction_{\mathcal{T}}$, the domain $\Theta \times \Theta$ by the subdomain $\mathcal{T}$, Lemma 1 by Lemma 3, Lemma 2 by Lemma 4, and (27) by (32). □

APPENDIX C
THE GAUSSIAN ALTRUISM EQUATIONS

This Appendix investigates equations (31) and (35), the heterarchical and hierarchical altruism equations, respectively, in the *Gaussian* case. Recall that both equations are scalar, algebraic equations, that depend on $\hat{u}_m$.

In the sequel we define x to be the realization of the random variable $\hat{u}^m$ corresponding to the realization z of the measurement vector Z. Furthermore, given the realization z, we define the conditional random vector Y as

$$Y \triangleq u_1 \mid Z = z \tag{113}$$

and assume that $Y \sim \mathcal{N}\left(\mu_Y, \sigma^2_Y\right)$.

A. Heterarchical Altruistic Estimation

For a given realization z of Z, and using the definition (113), the heterarchical altruism equation (31) reduces to

$$x = \frac{1}{2}[\mathbb{E}(Y \mid Y < x) + \mathbb{E}(Y \mid Y > x)]. \tag{114}$$

Let X be the standardized version of Y

$$X \triangleq \frac{Y - \mu_Y}{\sigma_Y} \tag{115}$$

with probability density function and cumulative density function ϕ and Φ, respectively. Then, letting χ be the realization of X corresponding to the realization x of Y, we have [10, Theorem 19.2]

$$\mathbb{E}(Y \mid Y > x) = \mu_Y + \sigma_Y \frac{\phi(\chi)}{1 - \Phi(\chi)} \tag{116a}$$

$$\mathbb{E}(Y \mid Y < x) = \mu_Y - \sigma_Y \frac{\phi(\chi)}{\Phi(\chi)}. \tag{116b}$$

Using equations (116) in (114) yields

$$\frac{\phi(\chi)}{2[1 - \Phi(\chi)]} - \frac{\phi(\chi)}{2\Phi(\chi)} - \chi = 0. \tag{117}$$

It is easy to see that $\chi = 0$ is a solution of (117). Noting (115), this solution of the realization-based (117) is equivalent to (39), the solution of the general equation (31), thus proving Proposition 1. In the following lemma we prove that this solution is unique.

Lemma 5. *$\chi = 0$ is the only solution of (117).*

Proof. To prove the lemma we define the function

$$f_{\text{HT}}(\chi) \triangleq \frac{\phi(\chi)}{2[1 - \Phi(\chi)]} - \frac{\phi(\chi)}{2\Phi(\chi)} - \chi \tag{118}$$

and show that $\chi = 0$ is its only zero. Since f_{HT} is an anti-symmetric function, it is sufficient to prove that it does not vanish in $(0, \infty)$. We do this by proving that $f_{\text{HT}}(\chi) < 0$ for all $\chi > 0$.

Defining

$$f_{\text{HT}}^{(1)}(\chi) \triangleq -\frac{\phi(\chi)}{2\Phi(\chi)} + \phi(0) - \frac{\chi}{2} \tag{119}$$

and

$$f_{\text{HT}}^{(2)}(\chi) \triangleq \frac{\phi(\chi)}{2[1 - \Phi(\chi)]} - \phi(0) - \frac{\chi}{2} \tag{120}$$

yields

$$f_{\text{HT}}(\chi) = f_{\text{HT}}^{(1)}(\chi) + f_{\text{HT}}^{(2)}(\chi). \tag{121}$$

We thus proceed to prove, separately, that both

$$f_{\text{HT}}^{(1)}(\chi) < 0 \quad \forall \chi > 0 \tag{122}$$

and

$$f_{\text{HT}}^{(2)}(\chi) < 0 \quad \forall \chi > 0. \tag{123}$$

To prove (122) we recast it as

$$g_{\text{HT}}^{(1)}(\chi) < 0 \quad \forall \chi > 0 \tag{124}$$

where

$$g_{\text{HT}}^{(1)}(\chi) \triangleq -\phi(\chi) + 2\phi(0)\Phi(\chi) - \chi\Phi(\chi). \tag{125}$$

The definition (125) gives

$$g_{\text{HT}}^{(1)}(0) = 0. \tag{126}$$

Since $g_{\text{HT}}^{(1)}(\chi)$ is a continuous function of its argument over $(0, \infty)$, it suffices to show that it is monotonically strictly decreasing in that interval. To this end, we calculate its derivative

$$g_{\text{HT}}'^{(1)}(\chi) = 2\phi(0)\phi(\chi) - \Phi(\chi) \tag{127}$$

and notice that, since $\phi(\chi) < \phi(0)$ and $\Phi(\chi) > \frac{1}{2}$ in $(0, \infty)$,

$$g'^{(1)}_{\text{HT}}(\chi) < 2\phi(0)^2 - \frac{1}{2} \approx -0.1817 \quad \forall \chi > 0. \tag{128}$$

For illustrative purposes, the function $g^{(1)}_{\text{HT}}(\chi)$ and its first derivative are depicted in Fig. 3.

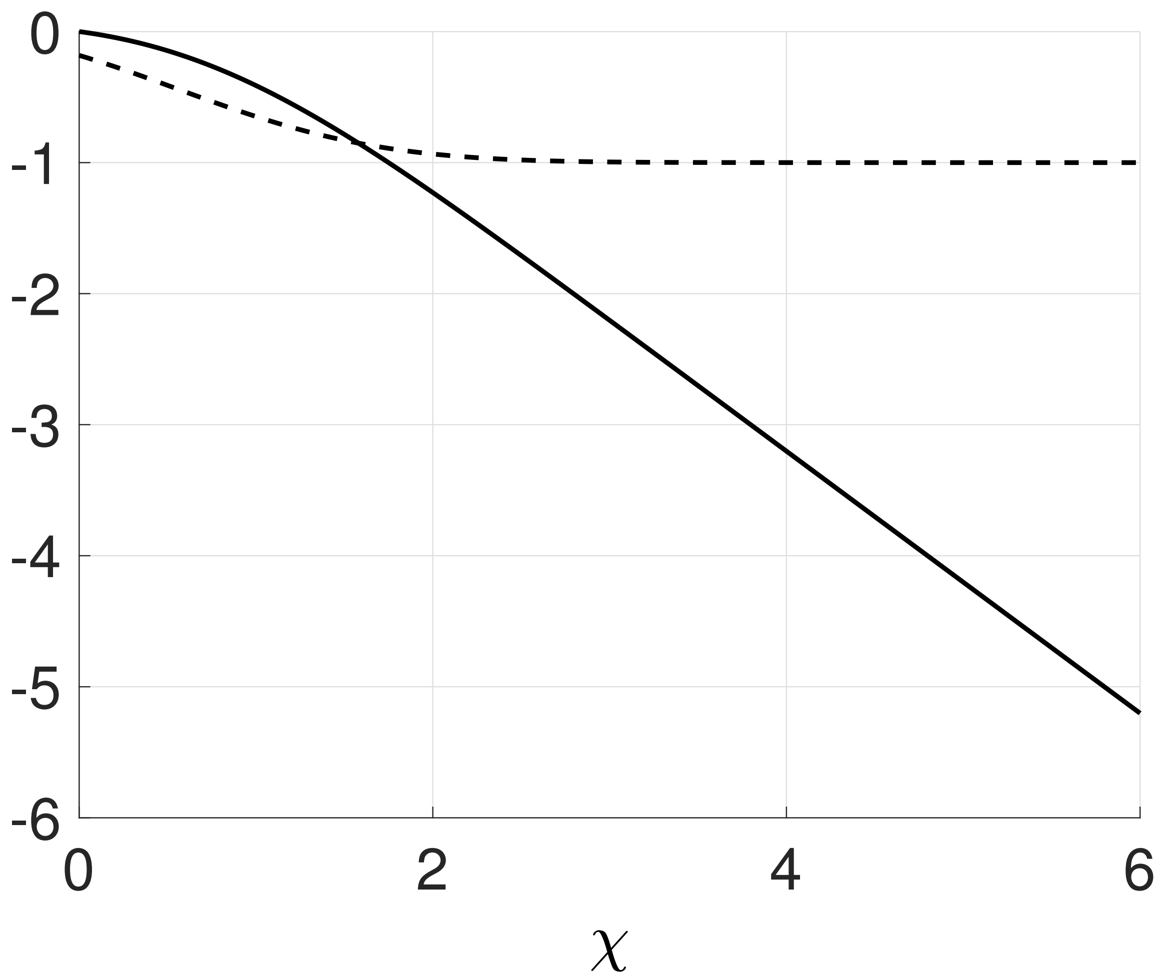

Fig. 3: The function $g^{(1)}_{\text{HT}}(\chi)$ (solid line) and its first derivative (dashed line).

To prove (123), we recast it as

$$g^{(2)}_{\text{HT}}(\chi) < 0 \quad \forall \chi > 0 \tag{129}$$

where

$$g^{(2)}_{\text{HT}}(\chi) \triangleq \phi(\chi) - 2\phi(0) + 2\phi(0)\Phi(\chi) - \chi[1 - \Phi(\chi)]. \tag{130}$$

The proof rests on the continuity of $g^{(2)}_{\text{HT}}(\chi)$ and its derivatives over $[0, \infty)$. Calculating the first three derivatives of $g^{(2)}_{\text{HT}}(\chi)$ yields

$$g'^{(2)}_{\text{HT}}(\chi) = \Phi(\chi) + 2\phi(0)\phi(\chi) - 1 \tag{131}$$

$$g''^{(2)}_{\text{HT}}(\chi) = \phi(\chi)[1 - 2\phi(0)\chi] \tag{132}$$

$$g'''^{(2)}_{\text{HT}}(\chi) = \phi(\chi)[2\phi(0)(\chi^2 - 1) - \chi]. \tag{133}$$

Clearly, $\chi_2 \triangleq \frac{1}{2\phi(0)}$ is the single zero of $g''^{(2)}_{\text{HT}}(\chi)$ in its entire domain. Since $g'''^{(2)}_{\text{HT}}(\chi_2) = -2\phi(0)\phi(\chi_2) < 0$, χ_2 is the single maximum point of $g'^{(2)}_{\text{HT}}(\chi)$, and we compute

$$g'^{(2)}_{\text{HT}}(\chi_2) \approx 0.04. \tag{134}$$

We now investigate the behavior of $g'^{(2)}_{\mathrm{HT}}(\chi)$ on both sides of its single extremal point. Clearly, $g'^{(2)}_{\mathrm{HT}}(\chi) \to 0$ as $\chi \to \infty$. Since χ_2 is the only extremal point of $g'^{(2)}_{\mathrm{HT}}(\chi)$, this yields that $g'^{(2)}_{\mathrm{HT}}(\chi) > 0$ for all $\chi \geq \chi_2$, from which we conclude that $g^{(2)}_{\mathrm{HT}}(\chi)$ does not have any extremal point in $[\chi_2, \infty)$.

Turning our attention to $\chi < \chi_2$, we first notice that $g'^{(2)}_{\mathrm{HT}}(0) \approx -0.1817$. Since $g'^{(2)}_{\mathrm{HT}}(\chi_2) > 0$, the mean value theorem yields that $g'^{(2)}_{\mathrm{HT}}(\chi)$ must have a zero in $(0, \chi_2)$. Moreover, since χ_2 is the only extremal point of $g'^{(2)}_{\mathrm{HT}}(\chi)$, we conclude that $g'^{(2)}_{\mathrm{HT}}(\chi)$ is monotonically strictly increasing in $(0, \chi_2)$, so that its zero in $(0, \chi_2)$ is unique. Denote that zero as χ_1. Clearly, χ_1 is a unique minimum point of $g^{(2)}_{\mathrm{HT}}(\chi)$ in $[0, \chi_2]$, since $g'^{(2)}_{\mathrm{HT}}(\chi) < 0$ for $0 \leq \chi < \chi_1$ and $g'^{(2)}_{\mathrm{HT}}(\chi) > 0$ for $\chi_1 < \chi \leq \chi_2$. Noting that $g^{(2)}_{\mathrm{HT}}(0) = 0$ and $g^{(2)}_{\mathrm{HT}}(\chi_2) \approx -0.0336$, we conclude that $g^{(2)}_{\mathrm{HT}}(\chi) < 0$ in $(0, \chi_2]$.

To prove that $g^{(2)}_{\mathrm{HT}}(\chi) < 0$ also for $\chi > \chi_2$, we observe that $g^{(2)}_{\mathrm{HT}}(\chi) \to 0$ as $\chi \to \infty$, which follows from both

$$\phi(\chi) - 2\phi(0) + 2\phi(0)\Phi(\chi) \to 0 \quad \text{as } \chi \to \infty \tag{135}$$

and

$$\chi[1 - \Phi(\chi)] \to 0 \quad \text{as } \chi \to \infty \tag{136}$$

where the latter limit results from using L'Hôpital's rule. Since $g^{(2)}_{\mathrm{HT}}(\chi_2) < 0$, the proof then follows from our previous conclusion that $g^{(2)}_{\mathrm{HT}}(\chi)$ does not have any extremal point in $[\chi_2, \infty)$. This also concludes the proof of the lemma.

For illustrative purposes, the function $g^{(2)}_{\mathrm{HT}}(\chi)$ and its first two derivatives are depicted in Fig. 4. □

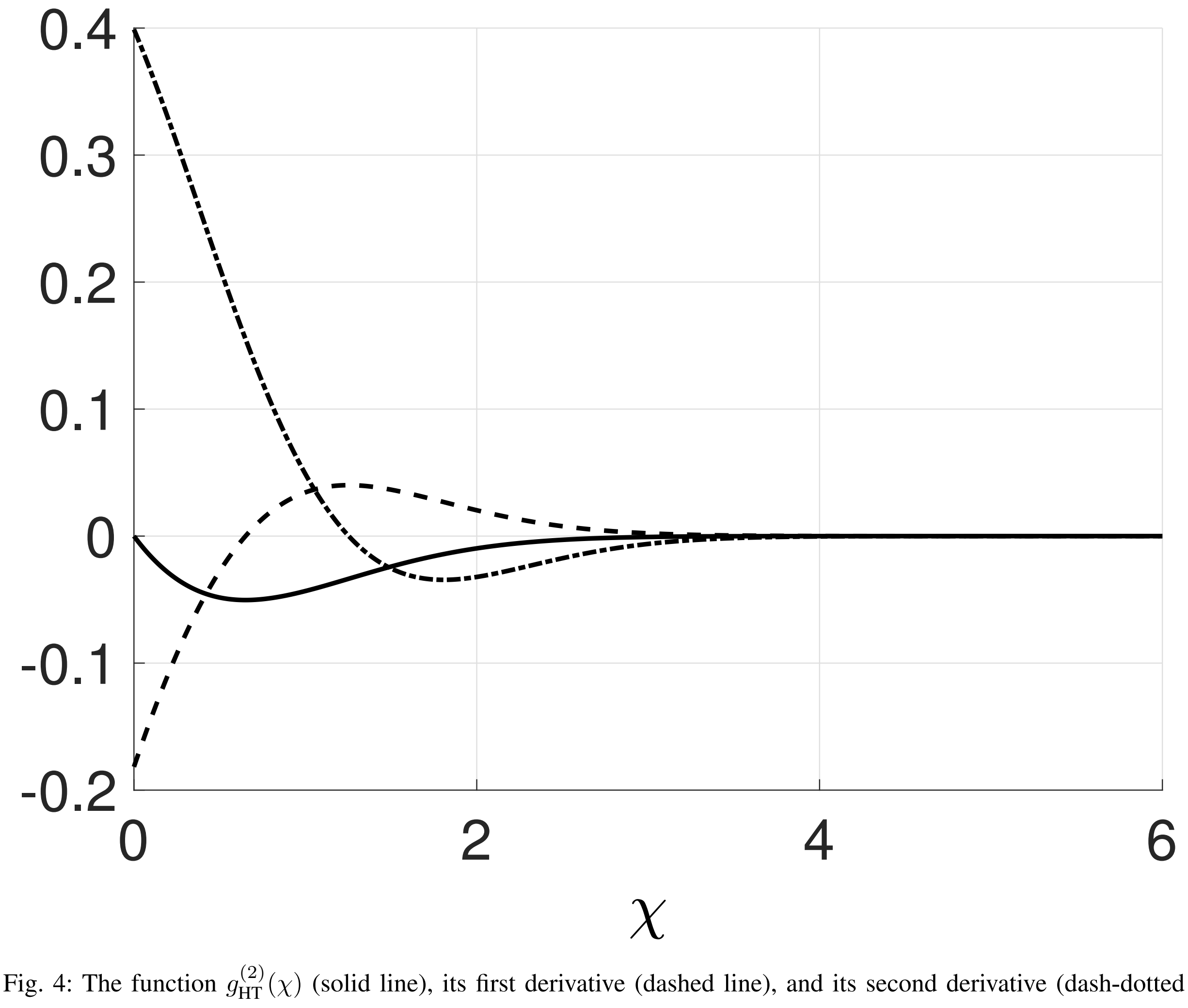

Fig. 4: The function $g^{(2)}_{\mathrm{HT}}(\chi)$ (solid line), its first derivative (dashed line), and its second derivative (dash-dotted line).

B. Hierarchical Altruistic Estimation

For a given realization z of Z the hierarchical altruism equation (35) reduces to

$$x = \frac{1}{2}[\mu_Y + \mathbb{E}(Y \mid Y > x)] \tag{137}$$

which, using (116a), yields

$$\frac{\phi(\chi)}{2[1-\Phi(\chi)]} - \chi = 0. \tag{138}$$

As opposed to its heterarchical counterpart (117), the hierarchical equation (138) does not lend itself to a closed form solution. Resorting to numerical methods we find its solution to be

$$\chi = \frac{1}{2} w_{\mathrm{HI}} \tag{139a}$$

with

$$w_{\mathrm{HI}} \approx 1.224. \tag{139b}$$

Furthermore, this solution is unique, as proved in the following lemma. Noting (115), this solution of the realization-based (138) is equivalent to (56), the solution of the general equation (35).

Lemma 6. (139) *is the unique solution to* (138).

Proof. Clearly, (138) cannot have a non-positive solution. To prove that it cannot have a positive solution other than (139), we recast (138) as

$$g_{\mathrm{HI}}(\chi) = 0 \tag{140}$$

where

$$g_{\mathrm{HI}}(\chi) \triangleq \phi(\chi) - 2\chi[1-\Phi(\chi)]. \tag{141}$$

The rest of the proof relies on the continuity of $g_{\mathrm{HI}}(\chi)$ and its derivatives, the first three of which are calculated to be

$$g'_{\mathrm{HI}}(\chi) = \chi\phi(\chi) - 2[1-\Phi(\chi)] \tag{142}$$

$$g''_{\mathrm{HI}}(\chi) = \phi(\chi)(3-\chi^2) \tag{143}$$

$$g'''_{\mathrm{HI}}(\chi) = \chi\phi(\chi^2-5). \tag{144}$$

To facilitate the ensuing development, we summarize in Table I the signs of $g_{\mathrm{HI}}(\chi)$ and its first three derivatives at $\chi = 0$ and at $\chi = \sqrt{3}$. The only root of $g''_{\mathrm{HI}}(\chi)$ in $[0,\infty)$ is $\chi = \sqrt{3}$. Since $g'''_{\mathrm{HI}}(\sqrt{3}) < 0$, it follows that this root is the only maximum

TABLE I: Signs of $g_{\mathrm{HI}}(\chi)$ and its first three derivatives at $\chi = 0$ and $\chi = \sqrt{3}$.

	$\chi = 0$	$\chi = \sqrt{3}$
$\operatorname{sgn} g_{\mathrm{HI}}(\chi)$	$+$	$-$
$\operatorname{sgn} g'_{\mathrm{HI}}(\chi)$	$-$	$+$
$\operatorname{sgn} g''_{\mathrm{HI}}(\chi)$	$+$	0
$\operatorname{sgn} g'''_{\mathrm{HI}}(\chi)$	0	$-$

point of $g'_{\mathrm{HI}}(\chi)$ in $[0,\infty)$, rendering $g'_{\mathrm{HI}}(\chi)$ monotonically non-increasing for $\chi > \sqrt{3}$. Furthermore, since $g'_{\mathrm{HI}}(\sqrt{3}) > 0$ and $g'_{\mathrm{HI}}(\chi) \to 0$ as $\chi \to \infty$, it follows that $g'_{\mathrm{HI}}(\chi) > 0$ in $[\sqrt{3},\infty)$, rendering $g_{\mathrm{HI}}(\chi)$ monotonically strictly increasing in that interval. Now, using (136), it is easy to see that $g_{\mathrm{HI}}(\chi) \to 0$ as $\chi \to \infty$. Since $g_{\mathrm{HI}}(\sqrt{3}) < 0$, we conclude that $g_{\mathrm{HI}}(\chi)$ does not possess any root in $[\sqrt{3},\infty)$.

To complete the proof, we need to show that $g_{\mathrm{HI}}(\chi)$ does not possess any root in $(0,\sqrt{3})$ other than (139) (as Table I shows that both 0 and $\sqrt{3}$ are not roots of $g_{\mathrm{HI}}(\chi)$). Since $g'_{\mathrm{HI}}(0) < 0$ and $g'_{\mathrm{HI}}(\sqrt{3}) > 0$, and as $g'_{\mathrm{HI}}(\chi)$ does not possess an extremum in $(0,\sqrt{3})$, it must be monotonically increasing in that interval, crossing zero at a single point in $(0,\sqrt{3})$. Thus, $g_{\mathrm{HI}}(\chi)$ can have only a single extremal point in that interval. This extremal point is a minimum point since $g''_{\mathrm{HI}}(\chi) > 0$ in $(0,\sqrt{3})$. Since $g_{\mathrm{HI}}(0) > 0$ and $g_{\mathrm{HI}}(\sqrt{3}) < 0$, we thus conclude that $g_{\mathrm{HI}}(\chi)$ can cross zero only once in $(0,\sqrt{3})$. This unique crossing is at (139).

For illustrative purposes, the function $g_{\mathrm{HI}}(\chi)$ is depicted in Fig. 5. □

ACKNOWLEDGMENT

The authors thank Vadim Indelman of the Technion's Department of Aerospace Engineering for his useful suggestions.

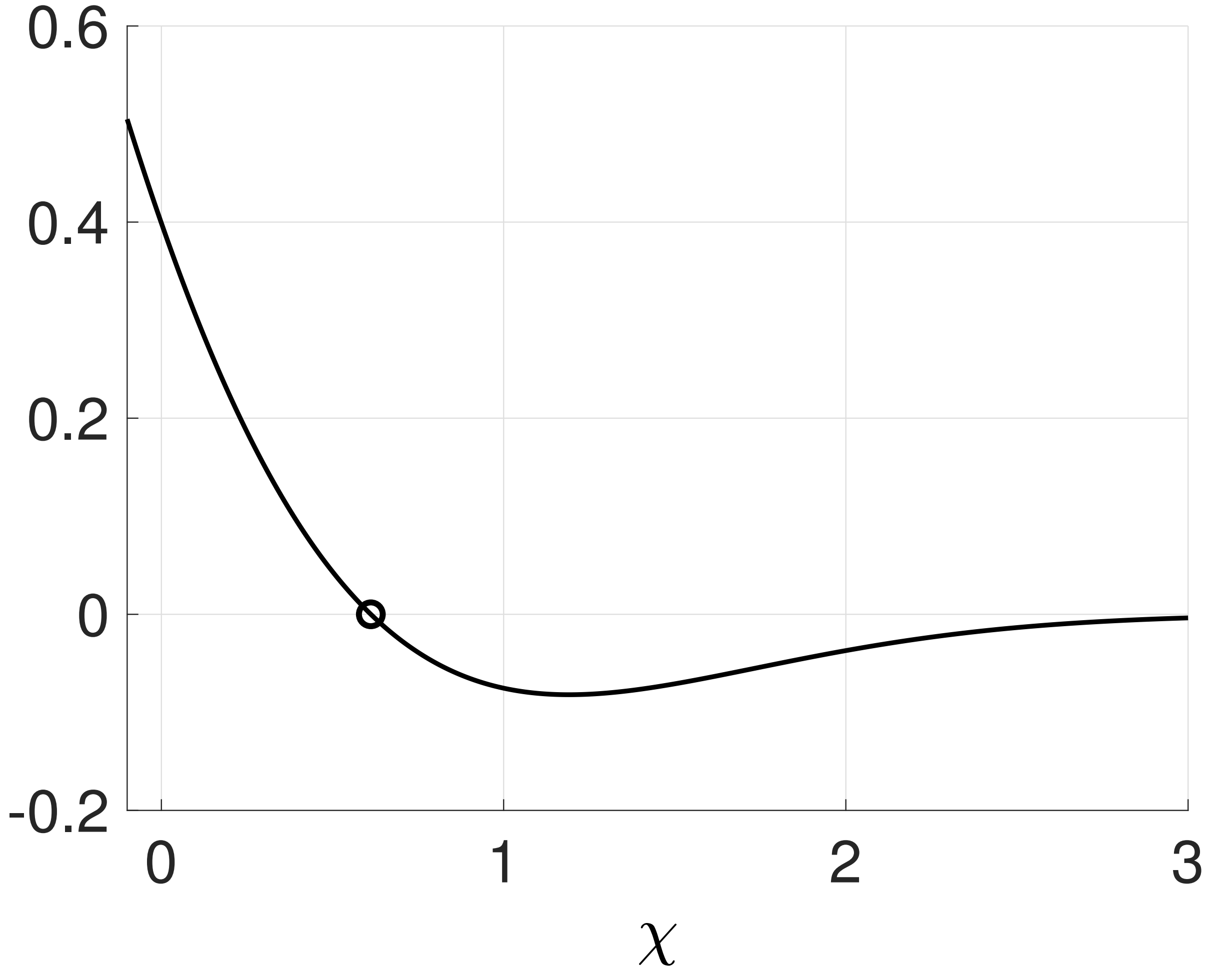

Fig. 5: The function $g_{\mathrm{HI}}(\chi)$. Superimposed is its numerically calculated root, $\chi \approx 0.612$ (circle).

References

[1] V. Shaferman and Y. Oshman, "Stochastic cooperative interception using information sharing based on engagement staggering," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 9, pp. 2127–2141, September 2016.

[2] S. Y. Hayoun and T. Shima, "Cooperative 2-on-1 bounded-control linear differential games," in *Advances in Aerospace Guidance, Control, and Dynamics, Selected Papers of the Third CEAS Specialist Conference on Guidance, Navigation and Control*, J. Bordeneuve-Guibe, A. Drouin, and C. Roos, Eds. Toulouse, France: Springer, April 2015.

[3] S. S. Kumkov, S. Le Ménec, and V. S. Patsko, "Solvability sets in pursuit problem with two pursuers and one evader," in *Proceedings of the 19th World Congress*. Cape Town, South Africa: The International Federation of Automatic Control, August 2014, pp. 1543–1549.

[4] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 3447–3457, Aug 2008.

[5] M. Guerriero, L. Svensson, D. Svensson, and P. Willett, "Shooting two birds with two bullets: How to find Minimum Mean OSPA estimates," in *Information Fusion (FUSION), 2010 13th Conference on*, July 2010, pp. 1–8.

[6] A. Okabe, B. Boots, K. Sugihara, S. N. Chiu, and D. G. Kendall, *Locational Optimization Through Voronoi Diagrams*. John Wiley & Sons, Inc., 2008, pp. 531–584.

[7] M. Iri, K. Murota, and T. Ohya, "A fast Voronoi-diagram algorithm with applications to geographical optimization problems," in *Lecture Notes in Control and Information Sciences*, vol. 59, 1984, pp. 273–288.

[8] Q. Du, V. Faber, and M. Gunzburger, "Centroidal Voronoi tessellations: Applications and algorithms," *SIAM Review*, vol. 41, no. 4, pp. 637–676, 1999.

[9] S. Martínez, J. Cortés, and F. Bullo, "Motion coordination with distributed information," *IEEE Control Systems Magazine*, vol. 27, no. 4, pp. 75–88, 2007.

[10] W. H. Greene, *Econometric Analysis*, 7th ed. New Jersey: Prentice Hall, 2012.

[11] F. C. Leone, L. S. Nelson, and R. B. Nottingham, "The folded normal distribution," *Technometrics*, vol. 3, no. 4, pp. 543–550, 1961.

[12] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985, pp. 176–177.