

On the Set of Possible Minimizers of a Sum of Known and Unknown Functions

Kananart Kuwaranancharoen and Shreyas Sundaram

Abstract— The problem of finding the minimizer of a sum of convex functions is central to the field of optimization. Thus, it is of interest to understand how that minimizer is related to the properties of the individual functions in the sum. In this paper, we consider the scenario where one of the individual functions in the sum is not known completely. Instead, only a region containing the minimizer of the unknown function is known, along with some general characteristics (such as strong convexity parameters). Given this limited information about a portion of the overall function, we provide a necessary condition which can be used to construct an upper bound on the region containing the minimizer of the sum of known and unknown functions. We provide this necessary condition in both the general case where the uncertainty region of the minimizer of the unknown function is arbitrary, and in the specific case where the uncertainty region is a ball.

I. INTRODUCTION

Optimization is an important tool in various fields, including machine learning [1], signal processing [2], control theory, [3]–[5], and robotics [6]–[8]. Given an objective function to be optimized, there are several standard algorithms that can be applied to find the optimal variables [9]–[12].

However, in many applications, it may be the case that the objective function is only partially known. For example, such scenarios are central to the field of robust optimization, where the objective function contains some parametric uncertainty, and the goal is to choose the optimization variable to be robust to the possible realizations of the uncertainty [13]–[15]. The problem that we consider in this paper also has a similar flavor, in that we assume that the optimization objective is not fully known. However, rather than seeking to find a single solution that is simultaneously robust to all possible realizations of the uncertain parameter (or learning that parameter [15]), we instead seek to characterize the region where the minimizer could lie for *each* possible realization of the uncertainty. This approach has the potential to yield insights regarding the nature of the possible solutions to the given uncertain optimization problem.

In our recent paper, [16] we determined a region containing the possible minimizers of a sum of two strongly convex functions, given only the minimizers of the local functions, their strong convexity parameters, and a bound on their gradients. In contrast, in this paper, we shall consider the case of optimizing a sum of known and unknown functions where only limited information about the unknown function is available. In this case, we are given some general

characteristics of the unknown function, namely a region containing the minimizer, and the strong convexity parameter of the function. Our goal is to determine necessary conditions for a point to be a minimizer of the sum. In particular, we will determine a region where the potential minimizer of the sum can lie. Thus, if a point from within this region is chosen as an estimate of the true minimizer of the sum, the size of the region can be used to quantify how far the estimate can be from the true minimizer. Below, we describe an example scenario to illustrate this problem.

An Example Scenario

In supervised machine learning problems, one uses labeled training data in order to construct a model that can be used to perform regression or classification tasks. The training data consists of pairs x_i and y_i which are the feature vector and label of the i -th example, respectively. For simplicity, assume that we have 2 training sets denoted by $\mathcal{D}_j = \{x_i^{(j)}, y_i^{(j)}\}_{i=1}^{N_j}$ for $j \in \{1, 2\}$. We can write the aggregate loss function of the whole dataset $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ as

$$L(w; \mathcal{D}) = \underbrace{\sum_{i=1}^{N_1} l(w; x_i^{(1)}, y_i^{(1)})}_{L_1(w; \mathcal{D}_1)} + \underbrace{\sum_{i=1}^{N_2} l(w; x_i^{(2)}, y_i^{(2)})}_{L_2(w; \mathcal{D}_2)}$$

where w is a model parameter that we need to optimize and $l(w; x_i^{(j)}, y_i^{(j)})$ is a loss function for each sample. Assume that $L(w; \mathcal{D})$ is a strongly convex function (which will be the case when we consider linear regression problems or functions incorporating l_2 regularization [17]). Suppose w^* and w_2^* are the minimizer of $L(w; \mathcal{D})$ and $L_2(w; \mathcal{D}_2)$, respectively.

Now suppose that the entity trying to find the optimal parameter w for $L(w; \mathcal{D})$ can only access the data set \mathcal{D}_1 , but not \mathcal{D}_2 (or alternatively, can only access a corrupted or *poisoned* version of \mathcal{D}_2 [18], [19]). In this case, the entity may only know certain properties of the function $L_2(w; \mathcal{D}_2)$ (such as its general form, convexity parameters, etc.), and a region containing the minimizer of $L_2(w; \mathcal{D}_2)$ (e.g., based on the statistical properties of the underlying data). Given this limited information about $L_2(w; \mathcal{D}_2)$, and with $L_1(w; \mathcal{D}_1)$ fully known, the entity could seek to find a region that is guaranteed to contain the minimizer of the true function $L(w; \mathcal{D})$. This is the problem tackled in this paper.

II. NOTATION AND PRELIMINARIES

A. Sets

We denote the closure, interior, and boundary of a set \mathcal{E} by $\bar{\mathcal{E}}$, \mathcal{E}° , and $\partial\mathcal{E} = \bar{\mathcal{E}} \setminus \mathcal{E}^\circ$, respectively.

B. Linear Algebra

We denote by \mathbb{R}^n the n -dimensional Euclidean space. For simplicity, we often use (x_1, \dots, x_n) to represent the column vector $[x_1 \ x_2 \ \dots \ x_n]^T$. We use e_i to denote the i -th basis vector (the vector of all zeros except for a one in the i -th position). We denote by $\langle u, v \rangle$ the Euclidean inner product of u and v i.e., $\langle u, v \rangle = u^T v$, by $\|\cdot\|$ the Euclidean norm $\|x\| := (\sum_i x_i^2)^{1/2}$ and by $\angle(u, v)$ the angle between vectors u and v . Note that

$$\angle(u, v) = \arccos \left(\frac{\langle u, v \rangle}{\|u\| \|v\|} \right).$$

We use $\mathcal{B}(x_0, r) = \{x \in \mathbb{R}^n : \|x - x_0\| < r\}$ and $\bar{\mathcal{B}}(x_0, r)$ to denote the open and closed ball, respectively, centered at x_0 of radius r . Moreover, the function $u(x_1, x_2) : (\mathbb{R}^n \times \mathbb{R}^n) \setminus \{(z_1, z_2) \in \mathbb{R}^n \times \mathbb{R}^n : z_1 = z_2\} \rightarrow \mathbb{R}^n$ denotes the unit vector in the direction of $x_1 - x_2$, i.e.,

$$u(x_1, x_2) = \frac{x_1 - x_2}{\|x_1 - x_2\|} \quad \text{with } x_1 \neq x_2. \quad (1)$$

C. Convex Sets and Convex Functions

A set \mathcal{C} in \mathbb{R}^n is said to be convex if, for all x_1 and x_2 in \mathcal{C} and all θ in the interval $(0, 1)$, the point $(1 - \theta)x_1 + \theta x_2 \in \mathcal{C}$.

We say a vector $g \in \mathbb{R}^n$ is a subgradient of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ at $x \in \text{dom } f$ if for all $z \in \text{dom } f$, $f(z) \geq f(x) + \langle g, z - x \rangle$.

If f is convex and differentiable, then its gradient at x is a subgradient; however, a subgradient can exist even when f is not differentiable at x . A function f is called subdifferentiable at x if there exists at least one subgradient at x . The set of subgradients of f at the point x is called the subdifferential of f at x , and is denoted $\partial f(x)$. The subdifferential $\partial f(x)$ is always a closed convex set, even if f is not convex. In addition, if f is continuous at x , then the subdifferential $\partial f(x)$ is bounded.

A function f is called strongly convex with parameter $\sigma > 0$ (or σ -strongly convex) if for all points $x, y \in \text{dom } f$, $\langle g_x - g_y, x - y \rangle \geq \sigma \|x - y\|^2$ for all $g_x \in \partial f(x)$ and $g_y \in \partial f(y)$. We denote the set of all convex functions by \mathcal{F} , and the set of all σ -strongly convex functions with minimizer x_u^* in the set $\mathcal{A} \subseteq \mathbb{R}^n$ and $\text{dom}(\cdot) = \mathbb{R}^n$ by $\mathcal{S}(\mathcal{A}, \sigma)$.

III. PROBLEM STATEMENT

We consider a function of the form

$$f(x) = f^k(x) + f^u(x) \quad (2)$$

where f^k and f^u are convex functions. We assume that we know f^k exactly, but do not know f^u , other than some general properties described below.

We assume that $f^k \in \mathcal{F}$ and $f^u \in \mathcal{S}(\mathcal{A}, \sigma)$ where \mathcal{A} is a compact set (i.e., we only know that f^u is σ -strongly convex and that its minimizer lies in some set \mathcal{A}). Our goal is to find the set of points $x \in \mathbb{R}^n$ that could potentially be the minimizer of $f(x)$ in (2). To this end, we will seek to characterize the region

$$\begin{aligned} \mathcal{M}(f^k, \mathcal{A}, \sigma) &\triangleq \{x \in \mathbb{R}^n : \exists f^u \in \mathcal{S}(\mathcal{A}, \sigma), \\ &\quad \mathbf{0} \in \partial f^k(x) + \partial f^u(x)\}. \end{aligned} \quad (3)$$

For simplicity of notation, we will omit the argument of the set $\mathcal{M}(f^k, \mathcal{A}, \sigma)$ and write it as \mathcal{M} . Note that \mathcal{M} contains all points $x \in \mathbb{R}^n$ that can potentially be a minimizer of f , given f^k , and the quantity σ and the set \mathcal{A} pertaining to f^u .

Remark 1: Returning to the regression scenario involving data that is not directly available to the optimizing entity (described in the Introduction), the unknown function would be of the form $f^u(x) = \|Ax - y\|^2$ where A is a matrix containing (unknown) training data and y is the (unknown) vector of corresponding labels. When A has full rank, the loss function is strongly convex. In addition, if some general underlying statistical properties of the data are known to the optimizing entity, it could estimate a lower bound on the strong convexity parameter σ , and a region containing the possible minimizer of $f^u(x)$. Thus, using this information, the central entity seeks to find the set of possible minimizers of the sum of this unknown function and its own loss function (corresponding to data that it has access to directly).

IV. ANALYSIS FOR GENERAL \mathcal{A}

In this section, we provide a necessary condition for a point $x^* \in \text{dom } f^k$ to be the minimizer of f in the general case where the uncertainty region \mathcal{A} of the minimizer of the unknown function is compact, but of arbitrary shape.

For any given point $x^* \in \mathbb{R}^n \setminus \mathcal{A}$, define the set

$$\tilde{\mathcal{A}}(\mathcal{A}, x^*) \triangleq \{x \in \partial\mathcal{A} : (1 - \theta)x + \theta x^* \notin \mathcal{A}, \forall \theta \in (0, 1)\}. \quad (4)$$

In words, $\tilde{\mathcal{A}}(\mathcal{A}, x^*)$ is the set of points x on the boundary of \mathcal{A} such that the line joining x to x^* does not intersect \mathcal{A} (except at x). We now present the main result of this section (in the interest of space, we provide the proof in [20]).

Theorem 1: Suppose $f^k \in \mathcal{F}$ and $\mathcal{A} \subseteq \text{dom } f^k$ is a compact set. A necessary condition for a point $x^* \in \mathbb{R}^n$ to be in $\mathcal{M}(f^k, \mathcal{A}, \sigma) \setminus \mathcal{A}$ is

$$\min_{x_u^* \in \mathcal{A}, g_{x^*}^k \in \partial f^k(x^*)} \frac{\langle g_{x^*}^k, u(x^*, x_u^*) \rangle}{\|x^* - x_u^*\|} \leq -\sigma. \quad (5)$$

Furthermore, the above inequality (5) can be reduced to

$$\min_{(x_u^*, g^k) \in \mathcal{X}(f^k, \mathcal{A}, x^*)} \frac{\langle g^k, u(x^*, x_u^*) \rangle}{\|x^* - x_u^*\|} \leq -\sigma \quad (6)$$

where

$$\mathcal{X}(f^k, \mathcal{A}, x^*) \triangleq \{(x, g) \in \tilde{\mathcal{A}}(\mathcal{A}, x^*) \times \partial f^k(x^*) : \langle g, u(x^*, x) \rangle < 0\}. \quad (7)$$

We can interpret the necessary condition in Theorem 1 as follows. To check whether $x^* \in \mathbb{R}^n$ can be a minimizer of $f(x)$, we can follow the inequality (5) and search for a pair $(x_u^*, g_{x^*}^k)$ with $x_u^* \in \mathcal{A}$ and $g_{x^*}^k \in \partial f^k(x^*)$ such that the pair satisfies the inequality

$$\frac{\langle g_{x^*}^k, u(x^*, x_u^*) \rangle}{\|x^* - x_u^*\|} \leq -\sigma. \quad (8)$$

However, the inequality (6) with $\mathcal{X}(f^k, \mathcal{A}, x^*)$ defined in (7) suggests that we do not have to search throughout the space $\mathcal{A} \times \partial f^k(x^*)$. Instead, we can restrict our attention to be in

the set \mathcal{X} . Now we have the variables x_u^* and $g_{x^*}^k$ that are coupled through the inequality $\langle g_{x^*}^k, u(x^*, x_u^*) \rangle < 0$. That is, if we first choose $g_{x^*}^k \in \partial f^k(x^*)$, then we can consider x_u^* that is in the set $\{x \in \partial \mathcal{A} : \langle g_{x^*}^k, u(x^*, x) \rangle < 0, (1-\theta)x + \theta x^* \notin \mathcal{A}, \forall \theta \in (0, 1)\}$. Similarly, if we first choose $x_u^* \in \{x \in \partial \mathcal{A} : (1-\theta)x + \theta x^* \notin \mathcal{A}, \forall \theta \in (0, 1)\}$, then we can consider $g_{x^*}^k$ that is in the set $\{g \in \partial f^k(x^*) : \langle g, u(x^*, x) \rangle < 0\}$.

If the function f^k is differentiable at x^* , we have a single element in the set $\partial f^k(x^*)$, namely $\nabla f^k(x^*)$, and we can search for $x_u^* \in \partial \mathcal{A}$ such that $\langle \nabla f^k(x^*), u(x^*, x_u^*) \rangle < 0$. However, if the set \mathcal{A} is arbitrary, this search may be computationally expensive. In the next section, we consider additional structure on the set \mathcal{A} to simplify the search.

Remark 2: Note that the set $\mathcal{A}^\circ \subseteq \mathcal{M}(f^k, \mathcal{A}, \sigma)$. To see this, note that for all $x^* \in \mathcal{A}^\circ$, there exists $\epsilon > 0$ such that $\mathcal{B}(x^*, \epsilon) \subset \mathcal{A}^\circ$. Suppose $g \in \partial f^k(x^*)$. We can choose $f^u(x) = \frac{\sigma_u}{2} \|x - (x^* + \frac{g}{\sigma_u})\|^2$ where $\sigma_u = \frac{2k\|g\|}{\epsilon}$ and $k = \max\{1, \frac{\sigma_u}{2\|g\|}\}$. One can verify that $x_u^* \in \mathcal{B}(x^*, \epsilon)$, $\nabla f^u(x^*) = -g$, and $\sigma_u \geq \sigma$.

V. ANALYSIS FOR THE CASE WHERE \mathcal{A} IS A BALL

Here, we consider additional structure on the uncertainty set \mathcal{A} in order to provide a more specific characterization of the region \mathcal{M} . In particular, we consider $\mathcal{A} = \bar{\mathcal{B}}(\bar{x}, \epsilon_0)$, where \bar{x} is the best guess of what the true parameter x_u^* is, and ϵ_0 is the maximum possible deviation of the true minimizer from our best guess.

We begin by investigating a property of the necessary condition (6) under a coordinate transformation. Suppose $x = (x_{(1)}, x_{(2)}, \dots, x_{(n)}) \in \mathbb{R}^n$ and $x^* \notin \bar{\mathcal{B}}(\bar{x}, \epsilon_0)$. Let \mathbf{T} and \mathbf{R} be the translation and rotation operators such that $\mathbf{R}(\mathbf{T}(\bar{x})) = \mathbf{0}$, $\mathbf{R}(\mathbf{T}(x^*)) = (\tilde{x}_{(1)}^*, 0, \dots, 0)$ with $\tilde{x}_{(1)}^* > 0$, and $\mathbf{R}(g^k) = (\tilde{g}_{(1)}, \tilde{g}_{(2)}, 0, \dots, 0)$ with $\tilde{g}_{(2)} \geq 0$ while preserving the distance between any two points. In other words, given the ball $\bar{\mathcal{B}}(\bar{x}, \epsilon_0)$, a point x^* and a vector g^k , we transform the coordinates so that the ball is centered at the origin, the point x^* lies on the x -axis, and the vector g^k lies on the x - y plane.

Next, consider the expression $\frac{\langle g, u(x^*, x_u^*) \rangle}{\|x^* - x_u^*\|}$. Notice that both numerator and denominator can be written as inner products. Since \mathbf{R} is a unitary operator, we have

$$\frac{\langle \mathbf{R}(g), u(\mathbf{R}(\mathbf{T}(x^*)), \mathbf{R}(\mathbf{T}(x_u^*))) \rangle}{\|\mathbf{R}(\mathbf{T}(x^*)) - \mathbf{R}(\mathbf{T}(x_u^*))\|} = \frac{\langle g, u(x^*, x_u^*) \rangle}{\|x^* - x_u^*\|}.$$

This means that even though we use the coordinate transformation $\mathbf{R}(\mathbf{T}(\cdot))$, we can still apply Theorem 1. Therefore, for the purpose of deriving our main result, without loss of generality, we can consider $\bar{x} = \mathbf{0}$, $x^* = (x_{(1)}^*, 0, \dots, 0)$ where $x_{(1)}^* > \epsilon_0$, and $g (= g^k) = (g_{(1)}, g_{(2)}, 0, \dots, 0)$, where $g_{(2)} \geq 0$.

Before going into the result, we introduce some definitions that will appear in the theorem. For any given $x^* \in \mathbb{R}^n$, define $z_1(x^*) \in \mathbb{R}^n$ as

$$z_1(x^*) \triangleq \arg \min_{x \in \bar{\mathcal{B}}(\bar{x}, \epsilon_0)} \|x - x^*\|. \quad (9)$$

By our assumption that $x^* = (x_{(1)}^*, 0, \dots, 0)$, we have $z_1(x^*) = (\epsilon_0, 0, \dots, 0)$. Since $x^* \notin \bar{\mathcal{B}}(\bar{x}, \epsilon_0)$, the point z_1 is unique and is on $\partial \bar{\mathcal{B}}(\bar{x}, \epsilon_0)$. If $g \neq \alpha(x^* - \bar{x}) = (\alpha x_{(1)}^*, 0, \dots, 0)$ for all $\alpha \geq 0$ (i.e., $\angle(g, x^* - \bar{x}) \neq 0$), we define the set \mathcal{P} to be such that

$$\mathcal{P}(g, x^*) \triangleq \arg \min_{x \in \partial \bar{\mathcal{B}}(\bar{x}, \epsilon_0)} \angle(g, x - x^*),$$

the point $z_2 \in \mathbb{R}^n$ to be such that

$$z_2(g, x^*) \triangleq \arg \min_{x \in \mathcal{P}(g, x^*)} \|x - x^*\|, \quad (10)$$

and the curve $\mathcal{C}_0(\bar{x}, \epsilon_0, g, x^*)$ to be the shortest path on the surface $\partial \bar{\mathcal{B}}(\bar{x}, \epsilon_0)$ that connects z_1 and z_2 together, i.e., \mathcal{C}_0 is the geodesic path between z_1 and z_2 on $\partial \bar{\mathcal{B}}(\bar{x}, \epsilon_0)$.

To clarify these definitions, we introduce two more objects. Let L be the ray that starts from the point x^* and runs parallel to the vector g i.e.,

$$L(g, x^*) = \{x \in \mathbb{R}^n : \exists t \in [0, \infty), x = x^* + tg\}.$$

If $g \neq \alpha(x^* - \bar{x}) = (\alpha x_{(1)}^*, 0, \dots, 0)$ for all $\alpha \in \mathbb{R}$, let P_2 be the 2-dimensional plane that contains the vectors g and $x^* - \bar{x}$ as its bases, and contains the point x^* , i.e.,

$$P_2(\bar{x}, g, x^*) \triangleq \{x \in \mathbb{R}^n : \exists s, t \in \mathbb{R} \text{ such that}$$

$$x = x^* + sg + t(x^* - \bar{x})\} \\ = \{x \in \mathbb{R}^n : x_{(3)} = x_{(4)} = \dots = x_{(n)} = 0\},$$

where the second equality follows from the fact that $x^* = (x_{(1)}^*, 0, 0, \dots, 0)$ and $g = (g_{(1)}, g_{(2)}, 0, \dots, 0)$.

There are two possible cases: (i) the ray L passes through the ball $\bar{\mathcal{B}}(\bar{x}, \epsilon_0)$ and (ii) the ray L does not pass through the ball $\bar{\mathcal{B}}(\bar{x}, \epsilon_0)$.

In the first case, we have

$$\min_{x \in \partial \bar{\mathcal{B}}(\bar{x}, \epsilon_0)} \angle(g, x - x^*) = 0,$$

and there are either one or two elements in the set \mathcal{P} . The point z_2 is the one that closer to the point x^* . Note that $z_2 \in P_2$. The illustration of the first case is shown in Fig. 1.

In the second case, we have

$$\min_{x \in \partial \bar{\mathcal{B}}(\bar{x}, \epsilon_0)} \angle(g, x - x^*) > 0.$$

The vector $z_2 - x^*$ is a tangent vector at the point z_2 on the ball $\bar{\mathcal{B}}$ and has angle $\angle(z_2 - x^*, \bar{x} - x^*) = \arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right)$. Furthermore, the point z_2 is on the plane P_2 since $z_2 - x^*$ and $\bar{x} - x^*$ must be on the same 2D-plane in order to minimize the angle between them. The illustration of the second case is shown in Fig. 2.

Since P_2 passes through the center \bar{x} of the ball $\bar{\mathcal{B}}(\bar{x}, \epsilon_0)$, we can define the great circle $\mathcal{G} \subset P_2$ which is the intersection of $\partial \bar{\mathcal{B}}$ with P_2 . Since z_1 and z_2 are in \mathcal{G} (and also in P_2), the geodesic path \mathcal{C}_0 is in P_2 . The geodesic path in both cases is also shown in Fig. 1 and Fig. 2.

Before stating the theorem, define the open half-space

$$\mathcal{H}(g, x^*) \triangleq \{x \in \mathbb{R}^n : \langle g, u(x^*, x) \rangle < 0\},$$

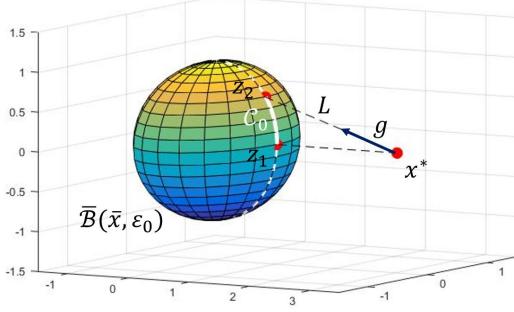


Fig. 1. The points z_1 and z_2 , and the curve C_0 on the surface $\partial\bar{B}(\bar{x}, \epsilon_0)$ in the case that the ray L passes through the ball $\bar{B}(\bar{x}, \epsilon_0)$.

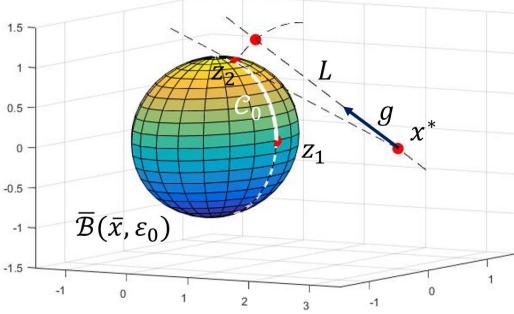


Fig. 2. The points z_1 and z_2 , and the curve C_0 on the surface $\partial\bar{B}(\bar{x}, \epsilon_0)$ in the case that the ray L does not pass through the ball $\bar{B}(\bar{x}, \epsilon_0)$.

Note that $C_0(\bar{x}, \epsilon_0, g, x^*) \cap \mathcal{H}(g, x^*) \neq \emptyset$ as long as $\angle(g, z_2 - x^*) < \frac{\pi}{2}$ or equivalently, $\angle(g, \bar{x} - x^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right)$ as shown in Fig. 3 and 4.

We now come to the main result of this section.

Theorem 2: Suppose $f^k \in \mathcal{F}$ and $\epsilon_0 > 0$. A necessary condition for a point $x^* \in \mathbb{R}^n$ to be in $\mathcal{M}(f^k, \bar{B}(\bar{x}, \epsilon_0), \sigma) \setminus \bar{B}(\bar{x}, \epsilon_0)$ is

$$\min_{(x_u^*, g^k) \in \tilde{\mathcal{X}}(f^k, \bar{x}, \epsilon_0, x^*)} \frac{\langle g^k, u(x^*, x_u^*) \rangle}{\|x^* - x_u^*\|} \leq -\sigma \quad (11)$$

where

$$\tilde{\mathcal{X}}(f^k, \bar{x}, \epsilon_0, x^*) \triangleq \{(x, g) \in \mathcal{C}_0(\bar{x}, \epsilon_0, g, x^*) \times \partial f^k(x^*)\}. \quad (12)$$

Proof: For a given $g_{x^*}^k \in \partial f^k(x^*)$ with $g_{x^*}^k \neq \mathbf{0}$, we consider the angle $\angle(g_{x^*}^k, x^* - \bar{x})$ in two disjoint cases:

(a) Suppose the gradient $g_{x^*}^k$ is colinear with the vector $x^* - \bar{x}$.

(i) If $g_{x^*}^k = \alpha(x^* - \bar{x})$ for some $\alpha > 0$ (i.e., $g_{x^*}^k$ is pointing directly away from $\bar{B}(\bar{x}, \epsilon_0)$ on the $x_{(1)}$ -axis), then $\langle g_{x^*}^k, u(x^*, x) \rangle > 0$ for all $x \in \bar{B}(\bar{x}, \epsilon_0)$. Thus, no points in $\bar{B}(\bar{x}, \epsilon_0)$ can satisfy the inequality (8).

(ii) If $g_{x^*}^k = \alpha(x^* - \bar{x})$ for some $\alpha < 0$ (i.e., $g_{x^*}^k$ is pointing directly toward $\bar{B}(\bar{x}, \epsilon_0)$ on the $x_{(1)}$ -axis), then the ray L passes through the ball $\bar{B}(\bar{x}, \epsilon_0)$ at z_1 , and thus $\{z_1(x^*)\} = \{z_2(g_{x^*}^k, x^*)\} = C_0$. Furthermore, $\bar{B} \subset \mathcal{H}$. For simplicity of notation,

we will omit the arguments and write $z_1(x^*)$ and $z_2(g_{x^*}^k, x^*)$ as z_1 and z_2 , respectively. From (10), for all $x \in \partial\bar{B}$, we have

$$\begin{aligned} \angle(g_{x^*}^k, u(z_2, x^*)) &\leq \angle(g_{x^*}^k, u(x, x^*)) \\ \Rightarrow \angle(g_{x^*}^k, u(x^*, z_2)) &\geq \angle(g_{x^*}^k, u(x^*, x)) \\ \Rightarrow \cos \angle(g_{x^*}^k, u(x^*, z_2)) &\leq \cos \angle(g_{x^*}^k, u(x^*, x)) \\ \Rightarrow \langle g_{x^*}^k, u(x^*, z_2) \rangle &\leq \langle g_{x^*}^k, u(x^*, x) \rangle. \end{aligned} \quad (13)$$

Since $z_2, x \in \mathcal{H}$, we have $\langle g_{x^*}^k, u(x^*, z_2) \rangle \leq \langle g_{x^*}^k, u(x^*, x) \rangle < 0$. In addition, from (9), for all $x \in \partial\bar{B}$, we have $0 < \|x^* - z_1\| \leq \|x^* - x\|$. Since $z_1 = z_2$ in this case, we obtain

$$\frac{-\|g_{x^*}^k\|}{\|x^* - z_1\|} = \frac{\langle g_{x^*}^k, u(x^*, z_1) \rangle}{\|x^* - z_1\|} \leq \frac{\langle g_{x^*}^k, u(x^*, x) \rangle}{\|x^* - x\|}$$

for all $x \in \partial\bar{B}$. Thus, it suffices to only check $z_1 \in C_0$ to see if (8) is satisfied.

(b) Suppose the gradient $g_{x^*}^k$ is not colinear with the vector $x^* - \bar{x}$. Then we can define the points z_1 and z_2 as described earlier. If

$$\angle(g_{x^*}^k, \bar{x} - x^*) \geq \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right),$$

then $\bar{B}(\bar{x}, \epsilon_0) \cap \mathcal{H}(g_{x^*}^k, x^*) = \emptyset$ as shown in Fig. 3, and no points in $\bar{B}(\bar{x}, \epsilon_0)$ can satisfy the inequality (8). If

$$\angle(g_{x^*}^k, \bar{x} - x^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right),$$

then $\bar{B}(\bar{x}, \epsilon_0) \cap \mathcal{H}(g_{x^*}^k, x^*) \neq \emptyset$ and $z_2 \in \mathcal{H}(g_{x^*}^k, x^*)$ as shown in Fig. 4. In this case, consider a point $x \in \partial\bar{B}(\bar{x}, \epsilon_0) \cap \mathcal{H}(g_{x^*}^k, x^*)$ and $x \notin C_0$.

(i) Suppose $\|x - x^*\| > \|z_2 - x^*\|$. By the definition of z_2 in (10), we have $\angle(g_{x^*}^k, u(z_2, x^*)) \leq \angle(g_{x^*}^k, u(x, x^*))$. Since $z_2, x \in \mathcal{H}$, using the same argument as (13), we get $\langle g_{x^*}^k, u(x^*, z_2) \rangle \leq \langle g_{x^*}^k, u(x^*, x) \rangle < 0$. Therefore,

$$\frac{\langle g_{x^*}^k, u(x^*, z_2) \rangle}{\|z_2 - x^*\|} < \frac{\langle g_{x^*}^k, u(x^*, x) \rangle}{\|x - x^*\|},$$

i.e., if x satisfies (8), then so does z_2 . This means that we can consider $z_2 \in C_0$ instead of any point in $\partial\bar{B}(\bar{x}, \epsilon_0) \cap \mathcal{H}(g_{x^*}^k, x^*)$ with greater distance from x^* .

(ii) Suppose $\|x - x^*\| \leq \|z_2 - x^*\|$. Since C_0 is connected and $h(y) = \|y - x^*\|$ is a continuous function, $\{h(y) : y \in C_0\}$ is connected. Then, we have

$$\left[\|z_1 - x^*\|, \|z_2 - x^*\| \right] \subseteq \{\|y - x^*\| : y \in C_0\}.$$

Thus, there exists a $z \in C_0 \cap \mathcal{H}$ such that $\|z - x^*\| = \|x - x^*\|$. However, since $C_0 \subset P_2$, we get that

$$\langle g_{x^*}^k, u(z, x^*) \rangle \leq \langle g_{x^*}^k, u(x, x^*) \rangle.$$

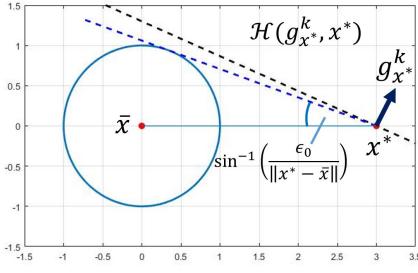


Fig. 3. The area above the black dotted line is $\mathcal{H}(g_{x^*}^k, x^*)$ and the blue dotted line shows the angle $\arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right)$. In this case, the angle $\angle(g_{x^*}^k, \bar{x} - x^*) \geq \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right)$, so $\mathcal{B}(\bar{x}, \epsilon_0) \cap \mathcal{H}(g_{x^*}^k, x^*) = \emptyset$.

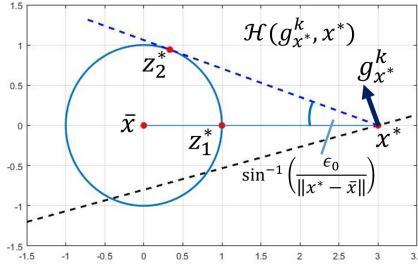


Fig. 4. The area above the black dotted line is $\mathcal{H}(g_{x^*}^k, x^*)$ and the blue dotted line shows the angle $\arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right)$. In this case, the angle $\angle(g_{x^*}^k, \bar{x} - x^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right)$, so $\mathcal{B}(\bar{x}, \epsilon_0) \cap \mathcal{H}(g_{x^*}^k, x^*) \neq \emptyset$.

Furthermore, since $z, x \in \mathcal{H}$, using the same argument as (13), we get $\langle g_{x^*}^k, u(x^*, z) \rangle < \langle g_{x^*}^k, u(x^*, x) \rangle < 0$. In this case, we also have

$$\frac{\langle g_{x^*}^k, u(x^*, z) \rangle}{\|z - x^*\|} \leq \frac{\langle g_{x^*}^k, u(x^*, x) \rangle}{\|x - x^*\|},$$

i.e., if x satisfies (8), then so does z .

Thus, we conclude that for each point $x \in \partial\mathcal{B} \cap \mathcal{H}$, there is a point $z \in \mathcal{C}_0$ such that $\frac{\langle g_{x^*}^k, u(x^*, z) \rangle}{\|x^* - z\|} \leq \frac{\langle g_{x^*}^k, u(x^*, x) \rangle}{\|x^* - x\|}$. Therefore, to check if there is a point $x \in \partial\mathcal{B} \cap \mathcal{H}$ satisfying (8), we only need to check points in \mathcal{C}_0 , yielding (11). ■

In fact, we can replace $\mathcal{C}_0(\bar{x}, \epsilon_0, g, x^*)$ in Theorem 2 by $\mathcal{C}_0(\bar{x}, \epsilon_0, g, x^*) \cap \mathcal{H}(g, x^*)$. However, for simplicity of exposition, we forego the discussion of this further reduction in search space.

The set $\tilde{\mathcal{X}}(f^k, \bar{x}, \epsilon_0, x^*)$ defined in (12) suggests that we do not have to search for a pair (x^*, g^k) that satisfies the inequality

$$\frac{\langle g^k, u(x^*, x_u^*) \rangle}{\|x^* - x_u^*\|} \leq -\sigma$$

throughout the set $\mathcal{X}(f^k, \mathcal{A}, x^*)$ defined in (7) but can instead restrict our attention to be in the set $\tilde{\mathcal{X}}(f^k, \bar{x}, \epsilon_0, x^*)$ in (12). Since the curve \mathcal{C}_0 depends on the vector g^k that we choose from $\partial f^k(x^*)$, we have to first select $g^k \in \partial f^k(x^*)$ and then we can consider the points on the curve \mathcal{C}_0 to see if they satisfy (8). We will use this in the algorithm for computing the region \mathcal{M} in the next section.

VI. ALGORITHM

Consider the case from the previous section where the uncertainty set is a ball, i.e., $\mathcal{A} = \mathcal{B}(\bar{x}, \epsilon_0)$. In this subsection, we will give an algorithm (Algorithm 1) to identify the region that satisfies the necessary condition (11). We provide a discussion of each of the steps below.

Algorithm 1 Region \mathcal{M} Identification (Ball Case)

Let $X \subseteq \text{dom } f^k$ be a set of points in the space

Input $X, f^k \in \mathcal{F}, \bar{x} \in \mathbb{R}^n, \epsilon_0 > 0$, and $\sigma > 0$

Output $\text{minimizer}(X)$

```

1: for  $x^* \in X$  do                                ▷ Loop through the space
2:    $\text{minimizer}(x^*) \leftarrow \text{false}$ 
3:    $d \leftarrow \|\bar{x} - x^*\|$ 
4:    $g \leftarrow \nabla f^k(x^*)$ 
5:    $\alpha \leftarrow \angle(g, \bar{x} - x^*)$ 
6:   if  $\alpha < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{d}\right)$  then
7:     for  $\theta \in [0, \arccos(\frac{\epsilon_0}{d})]$  do
8:        $\|x^* - x_u^*\| \leftarrow \sqrt{d^2 + \epsilon_0^2 - 2\epsilon_0 d \cos \theta}$ 
9:        $\angle(g, x^* - x_u^*) \leftarrow \alpha + \left[\pi - \arcsin\left(\frac{\epsilon_0 \sin \theta}{\|x^* - x_u^*\|}\right)\right]$ 
10:       $\langle g, u(x_u^*, x_u^*) \rangle \leftarrow \|g\| \cos \angle(g, x^* - x_u^*)$ 
11:      if  $\frac{\langle g, u(x_u^*, x_u^*) \rangle}{\|x^* - x_u^*\|} \leq -\sigma$  then
12:         $\text{minimizer}(x^*) \leftarrow \text{true}$ 
13: return  $\text{minimizer}(X)$ 

```

Let X be a set of points; we wish to check whether each point in X is a potential minimizer of $f^k + f^u$. For simplicity, we assume that the function f^k is differentiable, i.e., $\partial f^k(x^*) = \{\nabla f^k(x^*)\}$ and the set of points $X \subseteq \text{dom } f^k$. For example, we can use `linspace` in MATLAB to form a range for each axis, followed by using `meshgrid` to construct X . The object minimizer is an array that keeps a Boolean value for each point in X to indicate whether it is a potential minimizer. First, we loop through each point x^* in the set X and assign Boolean ‘false’ to that x^* . In order to change the Boolean to be ‘true’, the point x^* has to satisfy the inequality (11). Before checking that inequality, we need to compute several intermediate variables. In the algorithm, we compute the distance between the center of the ball \bar{x} and the point x^* ($d \leftarrow \|\bar{x} - x^*\|$), the gradient of f^k at x^* ($g \leftarrow \nabla f^k(x^*)$), and the angle between the gradient and reference ($\alpha \leftarrow \angle(g, \bar{x} - x^*)$). Note that we can compute α explicitly by

$$\alpha \leftarrow \angle(g, \bar{x} - x^*) = \arccos\left(\frac{\langle g, \bar{x} - x^* \rangle}{\|g\| \|\bar{x} - x^*\|}\right).$$

We then verify the condition

$$\angle(g, \bar{x} - x^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|x^* - \bar{x}\|}\right)$$

(line 6); if this is not satisfied, no points in $\mathcal{B}(\bar{x}, \epsilon_0)$ can satisfy the inequality (8) as argued in the proof of Theorem 2 and illustrated in Fig. 3. The next step is to compute the path \mathcal{C}_0 , which we parametrize by using the variable θ . The

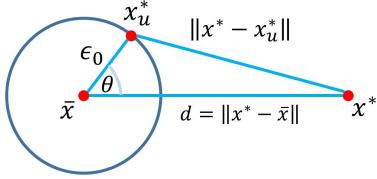


Fig. 5. Given ϵ_0 , d , and θ , we can compute $\|x^* - x_u^*\|$.

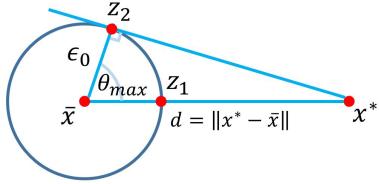


Fig. 6. Given ϵ_0 and d , we can compute $\angle(z_2 - \bar{x}, x^* - \bar{x})$.

variable θ in the algorithm corresponds to

$$\theta = \angle(x_u^* - \bar{x}, x^* - \bar{x}) \quad \text{where } x_u^* \in \mathcal{C}_0$$

as shown in Fig. 5. So, we need to know the range of θ that characterizes the path \mathcal{C}_0 . This range can be computed by considering the points z_1 and z_2 , at which the angle θ equals 0 and $\arccos(\frac{\epsilon_0}{\|\bar{x} - x^*\|})$, respectively, as shown in Fig. 6. Consider Fig. 5. For each θ in the range (discretized to a sufficiently fine resolution), we can compute the distance $\|x^* - x_u^*\|$ (line 8) by using the cosine law. Consider Fig. 7. We can compute the angle $\angle(g, x^* - x_u^*)$ (line 9) by using

$$\angle(g, x^* - x_u^*) = \angle(g, x^* - \bar{x}) + \arcsin\left(\frac{\epsilon_0 \sin \theta}{\|x^* - x_u^*\|}\right)$$

$$\text{and } \angle(g, x^* - \bar{x}) = (\pi - \angle(g, \bar{x} - x^*)).$$

After that we compute the inner product $\langle g, u(x^*, x_u^*) \rangle$ (line 10). Finally, we can compute the LHS of (11) and compare it to $-\sigma$. If the inequality (11) is satisfied by the current values x^* and θ , we set the Boolean associated to this x^* to be ‘true’. We provide an example illustrating this approach in [20].

VII. CONCLUSIONS

In this paper, we studied the properties of the minimizer of the sum of convex functions in which one of the functions is unknown but the others are known. However, we assumed that the unknown function is strongly convex with known convexity parameter, and that we have a region \mathcal{A} where the minimizer of this function lies. We established a necessary

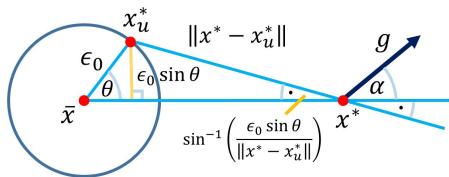


Fig. 7. Given ϵ_0 , θ , $\|x^* - x_u^*\|$, and α , we can compute $\angle(g, x^* - x_u^*)$.

condition for a given point to be a minimizer of the sum of known and unknown functions for general compact \mathcal{A} . We then considered a special case where the region of the unknown function’s minimizer is a ball. In this case, we simplified the necessary condition and provided an algorithm to determine the region that satisfies the necessary condition.

Future work could focus on providing sufficient conditions for a given point to be a minimizer (to complement our necessary condition). Alternatively, one could analyze properties of the set of solutions that satisfy the necessary condition.

REFERENCES

- [1] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1, no. 10.
- [2] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*. Prentice Hall Upper Saddle River, NJ, 2000, vol. 1.
- [3] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, “Constrained model predictive control: Stability and optimality,” *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.
- [4] A. E. Bryson, *Applied optimal control: optimization, estimation and control*. Routledge, 2018.
- [5] G. C. Calafiori and L. Fagiano, “Robust model predictive control via scenario optimization,” *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 219–224, 2013.
- [6] M. Zhu and S. Martínez, *Distributed optimization-based control of multi-agent networks in complex environments*. Springer, 2015.
- [7] E. Montijano and A. Mosteo, “Efficient multi-robot formations using distributed optimization,” in *53rd IEEE Conference on Decision and Control*, 2014, pp. 6167–6172.
- [8] K. Shin and N. McKay, “Minimum-time control of robotic manipulators with geometric path constraints,” *IEEE Transactions on Automatic Control*, vol. 30, no. 6, pp. 531–541, 1985.
- [9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [10] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [11] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *International Conference on Learning Representations (ICLR)*, 2015.
- [12] S. Hosseini, A. Chapman, and M. Mesbahi, “Online distributed ADMM via dual averaging,” in *53rd IEEE Conference on Decision and Control*, 2014, pp. 904–909.
- [13] A. Ben-Tal and A. Nemirovski, “Robust convex optimization,” *Mathematics of Operations Research*, vol. 23, no. 4, pp. 769–805, 1998.
- [14] D. Bertsimas, D. B. Brown, and C. Caramanis, “Theory and applications of robust optimization,” *SIAM Review*, vol. 53, no. 3, pp. 464–501, 2011.
- [15] H. Jiang and U. V. Shanbhag, “On the solution of stochastic optimization and variational problems in imperfect information regimes,” *SIAM Journal on Optimization*, vol. 26, no. 4, pp. 2394–2429, 2016.
- [16] K. Kuwaranancharoen and S. Sundaram, “On the location of the minimizer of the sum of two strongly convex functions,” in *IEEE Conference on Decision and Control (CDC)*, 2018, pp. 1769–1774.
- [17] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
- [18] B. Biggio, B. Nelson, and P. Laskov, “Poisoning attacks against support vector machines,” in *Proceedings of the 29th International Conference on Machine Learning*. Omnipress, 2012, pp. 1467–1474.
- [19] M. Mozaffari-Kermani, S. Sur-Kolay, A. Raghunathan, and N. K. Jha, “Systematic poisoning attacks on and defenses for machine learning in healthcare,” *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1893–1905, 2015.
- [20] K. Kuwaranancharoen and S. Sundaram, “On the set of possible minimizers of a sum of known and unknown functions,” *arXiv preprint*, 2020.