Chapter 5

Extended Karush-Kuhn-Tucker Condition for Constrained Interval Optimization Problems and its Application in Support Vector Machines

5.1 Introduction

Optimization theory has applications in various fields with wide engineering applications. Data acquisition and its quantification play a significant role to model an optimization problem. The data behind an optimization model is generally taken from measurements or observations. Often the sets of acquired data are reported with a given error percentage or with an imprecision. Such data is appropriately represented by fuzzy numbers or intervals. Hence, the different parameters/coefficients in the expression of the objective and constraint functions, which are modeled through the acquired data, become intervals or fuzzy numbers [89, 90]. This makes the objective and/or constraint functions of the optimization problem fuzzy-valued or interval-valued. In this chapter, we deal with interval-valued functions. The optimization problems with interval-valued objective and/or constraint functions are called *Interval Optimization Problems* (IOPs). In finding solutions to such optimization problems, the conventional optimization techniques are not directly applicable since they deal with real-valued functions.

5.2 Motivation

There have been numerous studies on IOPs. In many of the existing approaches, depending on best or worst case scenarios, the lower or the upper function (or their average) of the objective function is optimized [91–96]. Thereby the resulting problem becomes a conventional optimization problem, which has been solved by traditional optimization techniques. This strategy of transforming the IOP to a conventional optimization problem provides a single solution to the problem. It ignores to analyze the complete set of solutions. The Karush-Kuhn-Tucker (KKT) optimality conditions have also been applied to IOPs. It has been extensively studied by Wu in [97], [98] and [99]. Chalco-Cano *et al.* [100] have proposed KKT optimality conditions for IOPs using generalized derivative. Singh *et al.* [101, 102] utilized the partial ordering of intervals from [97] and the generalized derivatives from [100] to formulated KKT conditions for IOPs using the sum of lower and upper functions. It is to observe that existing literature on finding optimality conditions for IOPs attempted to generalize the conventional optimality conditions by some algebraic manipulations instead of the geometrical analysis of an optimal point. Apart from this, there are several other issues with the existing KKT theory for IOPs (for details, see the subsection 5.5.3). Therefore, we extend KKT conditions for constrained IOPs.

5.3 Contributions

In this chapter, we aim to find the KKT-optimality results by the geometrical analysis of the solutions of constrained and unconstrained IOPs. As an application of the derived results, we attempt to apply them on the binary classification problem with interval-valued data using *support vector machines* [103, 104].

5.4 Fundamentals of intervals and interval-valued functions

5.4.1 Interval arithmetic

Consider two intervals $\mathbf{A} = [\underline{a}, \bar{a}]$ and $\mathbf{B} = [\underline{b}, \bar{b}]$. Parametrically, $\mathbf{A}$ can be written as set of $a(t)$'s where $a(t) = \underline{a} + t(\overline{a} - \underline{a})$, $t \in [0, 1]$. The addition and the scalar multiplication defined, respectively,

$$\mathbf{A} \oplus \mathbf{B} = \left\{a(t_1) + b(t_2) :\ t_1, t_2 \in [0, 1]\right\} = [\underline{a} + \underline{b},\ \overline{a} + \overline{b}], \text{ and}$$

$$\lambda \odot \mathbf{A} = \left\{\lambda a(t) :\ t \in [0, 1]\right\} = \begin{cases} [\lambda\underline{a},\ \lambda\overline{a}] & \text{if } \lambda \geq 0 \\ [\lambda\overline{a},\ \lambda\underline{a}] & \text{if } \lambda < 0, \end{cases}$$

where λ is a real constant.

The difference between two intervals needs slightly more attention. This is mainly because of the following two reasons for the definition $\mathbf{A} \ominus \mathbf{B} = \left\{a(t_1) - b(t_2) :\ t_1, t_2 \in [0, 1]\right\}$:

(*i*) $\mathbf{A} \ominus \mathbf{A} \neq \{0\}$, and

(*ii*) given $\mathbf{C} = \mathbf{A} \ominus \mathbf{B}$, the relation $\mathbf{A} = \mathbf{B} \oplus \mathbf{C}$ does not necessarily hold.

A much refined definition was proposed in [105] and [106], which resolves these two drawbacks and gives the difference between two intervals as the interval $\mathbf{C}$ such that $\mathbf{A} = \mathbf{B} \oplus \mathbf{C}$. However, there is still a problem as the difference may not always exist for any two arbitrary compact intervals. This presents difficulties in defining differentiability of interval-valued functions (see [106] for details). Therefore, we employ the generalized Hukuhara difference (gH-difference) in order to appropriately define the difference between two intervals.

Definition 5.1 (*gH-difference of intervals* [105]). The gH-difference between two intervals $\mathbf{A} = [\underline{a}, \overline{a}]$ and $\mathbf{B} = [\underline{b}, \overline{b}]$ is denoted by $\mathbf{A} \ominus_{gH} \mathbf{B}$, defined by

$$\mathbf{A} \ominus_{gH} \mathbf{B} = [\min\{\underline{a} - \underline{b}, \overline{a} - \overline{b}\}, \max\{\underline{a} - \underline{b}, \overline{a} - \overline{b}\}]$$

.

It is easy to see that the gH-difference between any two intervals exists and also the relation $\mathbf{A} \ominus_{gH} \mathbf{A} = \{0\}$ holds.

Definition 5.2 (*Dominance relation of intervals* [107]). Let $\mathbf{A}$ and $\mathbf{B}$ be two elements of $I(\mathbb{R})$.

(i) $\mathbf{B}$ is said to be dominated by $\mathbf{A}$ if $a(t) \leq b(t)$ for all $t \in [0,1]$, and then we write $\mathbf{A} \preceq \mathbf{B}$;

(ii) we say $\mathbf{A} \neq \mathbf{B}$ if there exists a $t_0 \in [0,1]$ such that $a(t_0) \neq b(t_0)$;

(iii) $\mathbf{B}$ is said to be strictly dominated by $\mathbf{A}$ if $\mathbf{A} \preceq \mathbf{B}$ and $\mathbf{A} \neq \mathbf{B}$, and then we write $\mathbf{A} \prec \mathbf{B}$.

For two interval vectors $\mathbf{A}_v^k = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k)^\top$ and $\mathbf{B}_v^k = (\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_k)^\top$, we write $\mathbf{A}_v^k \preceq \mathbf{B}_v^k$ if $\mathbf{A}_i \preceq \mathbf{B}_i$ for each $i = 1, 2, \ldots, k$. Similarly, we can define $\mathbf{A}_v^k \prec \mathbf{B}_v^k$ also.

Lemma 5.1 *For any* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *in* $I(\mathbb{R})$, $\mathbf{A} \preceq \mathbf{B}$ *if and only if* $\mathbf{A} \ominus_{gH} \mathbf{B} \preceq \mathbf{0}$.

Proof: Let

$$\boldsymbol{A} = [\underline{a}, \bar{a}] = \{a(t) \ : \ a(t) = \underline{a} + t(\bar{a} - \underline{a}), 0 \leq t \leq 1\}$$

and

$$\boldsymbol{B} = [\underline{b}, \bar{b}] = \{b(t) \ : \ b(t) = \underline{b} + t(\bar{b} - \underline{b}), 0 \leq t \leq 1\}.$$

Then (see [105]), $\mathbf{A} \ominus_{gH} \mathbf{B} = \left[\min\left\{\underline{a} - \underline{b}, \bar{a} - \bar{b}\right\}, \max\left\{\underline{a} - \underline{b}, \bar{a} - \bar{b}\right\}\right]$.

Let $\boldsymbol{A} \preceq \boldsymbol{B}$. Then, by Definition 5.2, we note that

$$\begin{aligned}
&\boldsymbol{A} \preceq \boldsymbol{B} \\
&\implies \underline{a} + t(\bar{a} - \underline{a}) = a(t) \leq b(t) = \underline{b} + t(\bar{b} - \underline{b}) \text{ for all } t \in [0,1] \\
&\implies a(0) \leq b(0) \text{ and } \; \bar{a}(1) \leq \bar{b}(1) \\
&\implies \underline{a} \leq \underline{b} \text{ and } \; \bar{a} \leq \bar{b} \\
&\implies \underline{a} - \underline{b} \leq 0 \text{ and } \bar{a} - \bar{b} \leq 0 \\
&\implies \mathbf{A} \ominus_{gH} \mathbf{B} \preceq \mathbf{0}.
\end{aligned}$$

Conversely, let $\mathbf{A} \ominus_{gH} \mathbf{B} \preceq \mathbf{0}$. Therefore, we have $\underline{a} - \underline{b} \leq 0$ and $\bar{a} - \bar{b} \leq 0$, i.e., $\underline{a} \leq \underline{b}$ and $\bar{a} \leq \bar{b}$.

Depending on $\underline{b} < \bar{a}$ or $\bar{a} \leq \underline{b}$, we consider the following two cases.

• Case 1. Let $\underline{b} < \bar{a}$.

Then, $\underline{a} \leq \underline{b} < \bar{a} \leq \bar{b}$. In order to show that $\boldsymbol{A} \preceq \boldsymbol{B}$, we need to show that $a(t) \leq b(t)$ for all $t \in [0,1]$.

Let us assume that there exists $t_0 \in [0,1]$, such that $a(t_0) > b(t_0)$.

As $\underline{a} \leq \underline{b}$ and $\bar{a} \leq \bar{b}$, therefore $t_0 \neq 0$ and $t_0 \neq 1$. Thus, $\frac{1}{t_0} > 1$.

Note that from $a(t_0) = \underline{a} + t_0(\bar{a} - \underline{a})$, we have $\bar{a} = \frac{1}{t_0}\, a(t_0) - \left(\frac{1}{t_0} - 1\right)\underline{a}$. Similarly $\bar{b} = \frac{1}{t_0}\, b(t_0) - \left(\frac{1}{t_0} - 1\right)\underline{b}$. Since $a(t_0) > b(t_0)$, $\frac{1}{t_0} > 1$ and $\underline{a} \leq \underline{b}$, we see that

$$\bar{a} \;=\; \frac{1}{t_0}\, a(t_0) - \left(\frac{1}{t_0} - 1\right)\underline{a} \;>\; \frac{1}{t_0}\, b(t_0) - \left(\frac{1}{t_0} - 1\right)\underline{b} = \bar{b}.$$

This contradicts $\bar{a} \leq \bar{b}$. Therefore, for any $t \in [0,1]$, $a(t) \leq b(t)$. Hence, $\boldsymbol{A} \preceq \boldsymbol{B}$.

• Case 2. Let $\bar{a} \leq \underline{b}$.

Note that $a(t)$ and $b(t)$ are increasing functions. Therefore, for any $t \in [0,1]$ we have

$$a(t) \leq a(1) = \bar{a} \leq \underline{b} = b(0) \leq b(t).$$

Hence, $\boldsymbol{A} \preceq \boldsymbol{B}$ and the proof is complete. □

5.4.1.1 Interval-valued functions

There have been numerous works in interval-valued functions, each developing the theory further. We utilize the definitions proposed by Moore [108], Hansen [109], Wu [98], Bhurjee and Panda [107] and Ghosh [106]. A parametric definition of interval-valued functions as in [106] is given below.

Consider an interval-valued function $\mathbf{F}_{\mathbf{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$, where $\mathbf{C}_v^k = (\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_k)^\top$ denotes the vector of k interval coefficients. Taking $\mathbf{C}_j = \left[\underline{c_j}, \overline{c_j}\right]$, $j = 1, 2, \ldots, k$, the interval vector $\mathbf{C}_v^k$ can be presented by

$$\Big\{c(t) \ : \ c(t) \ = \big(c_1(t_1), c_2(t_2), \ldots, c_k(t_k)\big)^\top, \ \ t = (t_1, t_2, \ldots, t_k)^\top,$$
$$c_j(t_j) = \underline{c_j} + t_j(\overline{c_j} - \underline{c_j}), \ 0 \leq t_j \leq 1, \ j = 1, 2, \ldots, k\Big\}.$$

Therefore, the interval-valued function $\mathbf{F}_{\mathbf{C}_v^k}$ can be represented as a bunch of functions $f_{c(t)}$'s, where $c(t)$ is a vector in $\mathbf{C}_v^k$ in the parametric form. In other words, for all x in $\mathbb{R}^n$ we have

$$\mathbf{F}_{\mathbf{C}_v^k}(x) = \Big\{f_{c(t)}(x) : f_{c(t)} : \mathbb{R}^n \to \mathbb{R}, \ c(t) \in \mathbf{C}_v^k, \ t \in [0,1]^k\Big\}.$$

Definition 5.3 (*Interval-valued convex function* [107]). Let $X \subseteq \mathbb{R}^n$ be a convex set. An interval-valued function $\mathbf{F}_{\mathbf{C}_v^k} : X \to I(\mathbb{R})$ is said to be a convex on X if for any x_1

and x_2 in X,

$$\mathbf{F}_{\mathbf{C}_v^k}\left(\lambda x_1 + (1-\lambda)x_2\right) \preceq \lambda \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1) \oplus (1-\lambda) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_2) \text{ for all } \lambda \in [0,1].$$

5.4.1.2 gH-differentiability and gH-partial derivative

In all the definitions in this subsection, we consider that $\mathbf{F}_{\mathbf{C}_v^k}$ is an interval-valued function defined on $X \subseteq \mathbb{R}^n$.

Definition 5.4 (*gH-partial derivative* [106]). Let $x_0 = (x_1^0, x_2^0, \ldots, x_n^0)$ be an interior point of X and $h = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$ be such that $x_0 + h \in X$. Define a function

$$\mathbf{\Phi}_i(x_i) = \mathbf{F}_{\mathbf{C}_v^k}(x_1^0, x_2^0, \ldots, x_{i-1}^0, x_i, x_{i+1}^0, \ldots, x_n^0).$$

If the generalized Hukuhara derivative (gH-derivative) of $\mathbf{\Phi}_i$ exists at x_i^0, i.e.,

$$\lim_{h_i \to 0} \frac{\mathbf{\Phi}_i(x_i^0 + h_i) \ominus_{gH} \mathbf{\Phi}_i(x_i^0)}{h_i}$$

exists, then we say that $\mathbf{F}_{\mathbf{C}_v^k}$ has the i-th gH-partial derivative at x_0 and it is denoted by $D_i\mathbf{F}_{\mathbf{C}_v^k}(x^0)$, $i = 1, 2, \ldots, n$.

Note 5.1 (See [106]). It is evident that if $D_i\mathbf{F}_{\mathbf{C}_v^k}(x_0)$ exists, then $\mathbf{\Phi}_i(x_i^0 + h_i) \ominus_{gH} \mathbf{\Phi}_i(x_i^0)$ can be written as $h_i \odot \left(D_i\mathbf{F}_{\mathbf{C}_v^k}(x_0) \oplus \mathbf{E}_i(x_0; h)\right)$, where $\lim_{\|h\| \to 0} \mathbf{E}_i(x_0; h) = \mathbf{0}$.

Definition 5.5 (*gH-gradient* [106]). The gH-gradient of an interval-valued function $\mathbf{F}_{\mathbf{C}_v^k}$ at a point $x_0 \in X$ is denoted by $\nabla\mathbf{F}_{\mathbf{C}_v^k}(x_0)$ and defined by the vector

$$\nabla\mathbf{F}_{\mathbf{C}_v^k}(x_0) = \left(D_1\mathbf{F}_{\mathbf{C}_v^k}(x_0), D_2\mathbf{F}_{\mathbf{C}_v^k}(x_0), \ldots, D_n\mathbf{F}_{\mathbf{C}_v^k}(x_0)\right)^\top,$$

where $D_i\mathbf{F}_{\mathbf{C}_v^k}(x_0)$ is i-th gH-partial derivative of $\mathbf{F}_{\mathbf{C}_v^k}$ at $x_0 \in X$ for $i = 1, 2, \ldots, n$.

Definition 5.6 (*gH-differentiability* [106]). A function $\mathbf{F}_{\mathbf{C}_v^k} : X \to I(\mathbb{R})$ is said to be gH-differentiable at x_0 in X if there exist two interval-valued functions $\mathbf{E}(x_0; h)$ and $\mathbf{L}_{x_0} : \mathbb{R}^n \to I(\mathbb{R})$ such that

$$\mathbf{F}_{\mathbf{C}_v^k}(x_0 + h) \ominus_{gH} \mathbf{F}_{\mathbf{C}_v^k}(x_0) = \mathbf{L}_{x_0}(h) \oplus \|h\| \odot \mathbf{E}(x_0; h)$$

for $\|h\| < \delta$ for some $\delta > 0$, where $\lim_{\|h\| \to 0} \mathbf{E}(x_0; h) = \mathbf{0}$ and $\mathbf{L}_{x_0}$ is such a function that

(i) $\mathbf{L}_{x_0}(x + y) = \mathbf{L}_{x_0}(x) \oplus \mathbf{L}_{x_0}(y)$ for all $x, y \in X$, and

(ii) $\mathbf{L}_{x_0}(cx) = c \odot \mathbf{L}_{x_0}(x)$ for all $c \in \mathbb{R}$ and $x \in X$.

Theorem 5.1 (See [106]). Let $\mathbf{F}_{\mathbf{C}_v^k}$ be gH-differentiable at x_0. Then $\mathbf{L}_{x_0}$ exists for every h in $\mathbb{R}^n$ and $\mathbf{L}_{x_0}(h) = h^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0)$.

Theorem 5.2 *Let $\boldsymbol{F}_{\boldsymbol{C}_v^k} : X \to I(\mathbb{R})$ be a gH-differentiable at any $x \in X$, where X is a nonempty open convex subset of $\mathbb{R}^n$. Then, $\boldsymbol{F}_{\boldsymbol{C}_v^k}$ is convex on X if and only if*

$$(x_2 - x_1)^\top \odot \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_1) \preceq \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_2) \oplus (-1) \odot \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_1) \textit{ for all } x_1, x_2 \in X.$$

Proof: First, we assume that $\mathbf{F}_{\mathbf{C}_v^k}$ is convex on X, and x_1 and x_2 are any two elements of X. Then, for $h = x_2 - x_1$ and $0 < \tau_0 < 1$,

$$\begin{aligned}\mathbf{F}_{\mathbf{C}_v^k}(x_1 + \tau_0 h) &= \mathbf{F}_{\mathbf{C}_v^k}\left((1 - \tau_0)x_1 + \tau_0(x_1 + h)\right)\\ &\preceq (1 - \tau_0) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1) \oplus \tau_0 \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1 + h).\end{aligned}$$

Therefore, by Lemma 5.1,

$$\begin{aligned}
&\mathbf{F}_{\mathbf{C}_v^k}(x_1 + \tau_0 h) \ominus_{gH} \mathbf{F}_{\mathbf{C}_v^k}(x_1) \preceq \tau_0 \odot \left(\mathbf{F}_{\mathbf{C}_v^k}(x_1 + h) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1)\right)\\
\text{or,}\quad &\frac{1}{\tau_0} \odot \left(\mathbf{F}_{\mathbf{C}_v^k}(x_1 + \tau_0 h) \ominus_{gH} \mathbf{F}_{\mathbf{C}_v^k}(x_1)\right) \preceq \mathbf{F}_{\mathbf{C}_v^k}(x_1 + h) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1)\\
\text{or,}\quad &\frac{1}{\tau_0} \odot \left(\mathbf{F}_{\mathbf{C}_v^k}(x_1 + \tau_0 h) \ominus_{gH} \mathbf{F}_{\mathbf{C}_v^k}(x_1)\right) \preceq \mathbf{F}_{\mathbf{C}_v^k}(x_2) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1).
\end{aligned}$$

Hence, as $\tau_0 \to 0+$, with the help of Definition 5.6 and Theorem 5.1, we get

$$(x_2 - x_1)^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_1) \preceq \mathbf{F}_{\mathbf{C}_v^k}(x_2) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1).$$

Conversely, let

$$(x_2 - x_1)^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_1) \preceq \mathbf{F}_{\mathbf{C}_v^k}(x_2) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1)$$

be true for any x_1 and x_2 in X. Thus, for $0 \leq \lambda \leq 1$, denoting $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$, the following two inequalities hold true

$$(1 - \lambda) \odot \left((x_1 - x_2)^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_1)\right) \preceq \mathbf{F}_{\mathbf{C}_v^k}(x_1) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_\lambda) \tag{5.1}$$

$$\text{and}\quad \lambda \odot \left((x_2 - x_1)^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_1)\right) \preceq \mathbf{F}_{\mathbf{C}_v^k}(x_2) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_\lambda). \tag{5.2}$$

Multiplying (5.1) by λ and (5.2) by $(1 - \lambda)$, and then adding, we obtain

$$\mathbf{0} \preceq \left(\lambda \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1) \oplus (1 - \lambda) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_2)\right) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_\lambda).$$

Hence,

$$\mathbf{F}_{\mathbf{C}_v^k}(\lambda x_1 + (1 - \lambda)x_2) \preceq \lambda \odot \mathbf{F}_{\mathbf{C}_v^k}(x_1) \oplus (1 - \lambda) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_2).$$

Arbitrariness of $\lambda \in [0, 1]$ proves that $\mathbf{F}_{\mathbf{C}_v^k}$ is convex on X. $\square$

5.5 Fritz John and Karush-Kuhn-Tucker optimality conditions

Theorem 5.3 *Let $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$ be an interval-valued function which is gH-differentiable at x_0. Let there exists a vector $d \in \mathbb{R}^n$ such that $d^\top \odot \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) \prec \mathbf{0}$. Then, there exists $\delta > 0$ such that for each $\alpha \in (0, \delta)$, $\boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0 + \alpha d) \prec \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0)$.*

Proof: As $\mathbf{F}_{\mathbf{C}_v^k}$ is gH-differentiable at x_0, by Definition 5.6 and Theorem 5.1 we get

$$\mathbf{F}_{\mathbf{C}_v^k}(x_0 + h) \ominus_{gH} \mathbf{F}_{\mathbf{C}_v^k}(x_0) = h^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \ \oplus \ \|h\| \odot \mathbf{E}(x_0; h),$$

for some $\mathbf{E}(x_0; h)$ which tends to $\mathbf{0}$ as $\|h\| \to 0$. On replacing $h = \alpha d$, for $\alpha > 0$, we get

$$\mathbf{F}_{\mathbf{C}_v^k}(x_0 + \alpha d) = \mathbf{F}_{\mathbf{C}_v^k}(x_0) \ \oplus \ \alpha d^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \ \oplus \ |\alpha| \ \|d\| \odot \mathbf{E}(x_0; h).$$

Since $d^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \prec \mathbf{0}$ and $\mathbf{E}(x_0; \alpha d) \to \mathbf{0}$ as $\alpha \to 0+$, we have $\mathbf{F}_{\mathbf{C}_v^k}(x_0 + \alpha d) \prec \mathbf{F}_{\mathbf{C}_v^k}(x_0)$, for each $\alpha \in (0, \delta)$, for some $\delta > 0$. □

Note 5.2 *Theorem 5.3 shows that the vector d is a descent direction of $\boldsymbol{F}_{\boldsymbol{C}_v^k}$ at x_0.*

Definition 5.7 *(Cone of descent directions). For an interval-valued function $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$ which is gH-differentiable at x_0, the set of descent directions at x_0 is given by the set*

$$\hat{F}(x_0) = \{d \in \mathbb{R}^n : d^\top \odot \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) \prec \mathbf{0}\}.$$

As for any d in $\hat{F}(x_0)$, $\lambda d \in \hat{F}(x_0)$ for all $\lambda > 0$, the set $\hat{F}(x_0)$ is called the cone of descent directions.

Definition 5.8 (*Cone of feasible directions* [110]). Given a nonempty set $X \subseteq R^n$ and $x_0 \in X$. At x_0, the cone of feasible directions of X is defined by

$$\hat{S}(x_0) = \{d \in \mathbb{R}^n : d \neq 0, \ x_0 + \alpha d \in X \ \forall \alpha \in (0, \delta) \text{ for some } \delta > 0\}.$$

Analogous to the efficient solution concept in multi-objective optimization problems, we use the following efficient solution concept for IOPs.

Definition 5.9 (*Efficient solution* [107]). A feasible solution $\bar{x} \in X$ is called a (local) efficient solution of the IOP

$$\min_{x \in X \subseteq \mathbb{R}^n} \mathbf{F}_{\mathbf{C}_v^k}(x)$$

if there does not exist any $x \in X$ ($\in N_\delta(\bar{x})$) such that $\mathbf{F}_{\mathbf{C}_v^k}(x) \prec \mathbf{F}_{\mathbf{C}_v^k}(\bar{x})$ ($N_\delta(\bar{x})$ is a δ-neighborhood of $\bar{x}$). If a solution $\bar{x}$ is (local) efficient, then we call $\mathbf{F}_{\mathbf{C}_v^k}(\bar{x})$ as a (local) *non-dominated solution* to the IOP.

Theorem 5.4 *Given a nonempty open set $X \subseteq \mathbb{R}^n$, consider the interval optimization problem,*

$$\min_{x \in X \subseteq \mathbb{R}^n} \boldsymbol{F}_{\boldsymbol{C}_v^k}(x),$$

where $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$. If $\boldsymbol{F}_{\boldsymbol{C}_v^k}$ is gH-differentiable at a point $x_0 \in X$ and x_0 is a local efficient solution, then $\hat{F}(x_0) \cap \hat{S}(x_0) = \emptyset$.

Proof: We shall prove the theorem by contradiction. Let $\hat{F}(x_0) \cap \hat{S}(x_0) \neq \emptyset$ and d be an element in $\hat{F}(x_0) \cap \hat{S}(x_0)$. Then, in view of Theorem 5.3, there exists $\delta_1 > 0$ such that

$$\mathbf{F}_{\mathbf{C}_v^k}(x_0 + \alpha d) \prec \mathbf{F}_{\mathbf{C}_v^k}(x_0) \text{ for each } \alpha \in (0, \delta_1).$$

Also, by Definition 5.8, there exists $\delta_2 > 0$ such that $x_0 + \alpha d \in X$ for each $\alpha \in (0, \delta_2)$. Defining $\delta = \min\{\delta_1, \delta_2\} > 0$, we see that for all $\alpha \in (0, \delta)$,

$$x_0 + \alpha d \in X \text{ and } \mathbf{F}_{\mathbf{C}_v^k}(x_0 + \alpha d) \prec \mathbf{F}_{\mathbf{C}_v^k}(x_0).$$

This is contradictory to x_0 a local efficient solution. Hence, $\hat{F}(x_0) \cap \hat{S}(x_0) = \emptyset$. □

The following corollary is immediately followed from the Theorem 5.4.

Corollary 1 *Let $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$ be an interval-valued function. If at a point $x_0 \in X$, $\hat{F}(x_0) \cap \hat{S}(x_0) \neq \emptyset$, then x_0 is not an efficient point for the problem $\min_{x \in X} \boldsymbol{F}_{\boldsymbol{C}_v^k}(x)$.*

Example 5.1 (Example to support the Theorem 5.4 and Corollary 1).

Let $X \subset \mathbb{R}^2$ be the set $\{(x_1, x_2) | 1 \leq x_1 \leq 2, 1 \leq x_2 \leq 2\}$. Consider the interval optimization problem

$$\min_{x \in X} \mathbf{F}(x_1, x_2),$$

where $\mathbf{F}(x_1, x_2) = \left[\underline{F}(x_1, x_2), \overline{F}(x_1, x_2)\right]$ and

$$\underline{F}(x_1, x_2) = 1 + 2(x_1 - 1)^2 + 2(x_2 - 2)^2$$
$$\text{and } \underline{F}(x_1, x_2) = 5 + 5(x_1 - 1)^2 + 5(x_2 - 1)^2.$$

The objective function $\mathbf{F}(x_1, x_2)$ is depicted in the Figure 5.1. The red dots in the surfaces of $\underline{F}(x_1, x_2)$ and $\overline{F}(x_1, x_2)$ are the locations of the minima of the functions $\underline{F}$ and $\overline{F}$, respectively. From the figure, it is evident that $x_0 = (1, 1.5) \in X$ is an efficient point. At x_0, the cone of feasible directions is given by

$$\hat{S}(x_0) = \{(d_1, d_2) \neq (0, 0) : (1 + \alpha d_1, 1.5 + \alpha d_2) \in X \ \forall \alpha \in (0, \delta) \text{ for some } \delta > 0\}$$
$$= \{(d_1, d_2) \neq (0, 0) : d_1 \geq 0\}.$$

At x_0, the partial derivatives of $\mathbf{F}$ are

$$D_1\mathbf{F}(x_0) = \left[\min\left\{\frac{\partial \underline{F}}{\partial x_1}(x_0), \frac{\partial \overline{F}}{\partial x_1}(x_0)\right\}, \max\left\{\frac{\partial \underline{F}}{\partial x_1}(x_0), \frac{\partial \overline{F}}{\partial x_1}(x_0)\right\}\right] = [0, 0]$$

$$\text{and } D_2\mathbf{F}(x_0) = \left[\min\left\{\frac{\partial \underline{F}}{\partial x_2}(x_0), \frac{\partial \overline{F}}{\partial x_2}(x_0)\right\}, \max\left\{\frac{\partial \underline{F}}{\partial x_2}(x_0), \frac{\partial \overline{F}}{\partial x_2}(x_0)\right\}\right] = [-2, 5]$$

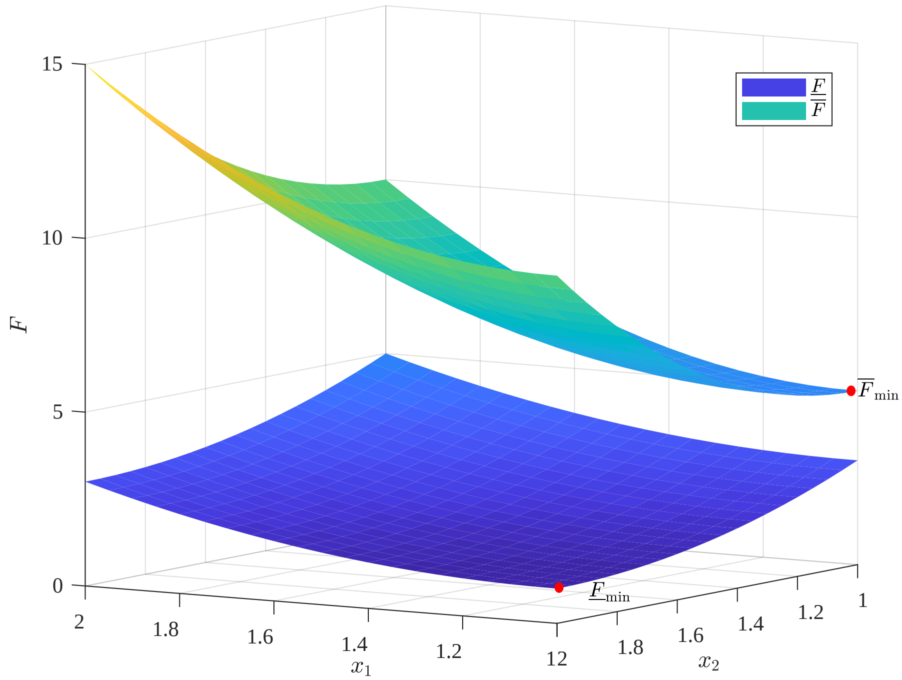

Figure 5.1: The objective function $\mathbf{F}(x_1, x_2)$ of the Example 5.1

Hence, at x_0, the cone of descent directions is given by

$$\begin{aligned}\hat{F}(x_0) &= \{(d_1, d_2) \in \mathbb{R}^2 : (d_1, d_2) \odot \nabla \mathbf{F}(x_0)^\top \prec \mathbf{0}\} \\ &= \{(d_1, d_2) \in \mathbb{R}^2 : d_1 \odot D_1\mathbf{F}(x_0) \oplus d_2 \odot D_1\mathbf{F}(x_0) \prec \mathbf{0}\} \\ &= \{(d_1, d_2) \in \mathbb{R}^2 : d_2 \odot [-2, 5] \prec 0\} \\ &= \emptyset.\end{aligned}$$

Thus, at the efficient solution $x_0 = (1, 1.5)$, we see that $\hat{S}(x_0) \cap \hat{F}(x_0) = \emptyset$.

Let us take another point $x_{00} = (2, 1.5)$. In a similar way as that for the point x_0, we can check that at the point x_{00}:

$$\hat{S}(x_{00}) = \{(d_1, d_2) \neq (0, 0) : d_1 \leq 0\}$$

$$\text{and } \hat{F}(x_{00}) = \{(d_1, d_2) : 2d_1 < d_2 < -2d_1\}.$$

The cone $\hat{F}(x_{00})$ is depicted in the Figure 5.2.

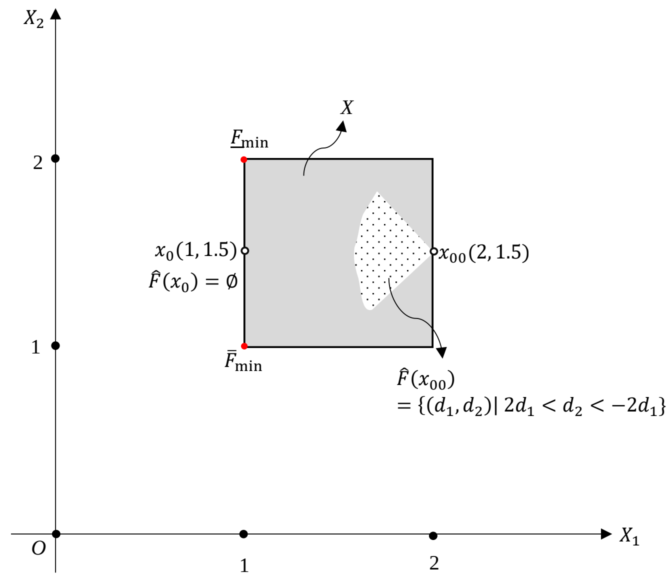

Figure 5.2: The cones of descent directions $\hat{F}(x_0)$ and $\hat{F}(x_{00})$ for the Example 5.1

We note that

$$\hat{S}(x_{00}) \cap \hat{F}(x_{00}) = \{(d_1, d_2) : 2d_1 < d_2 < -2d_1\} \neq \emptyset.$$

Thus, due to the Corollary 1 of the Theorem 5.4, the point x_{00} must not be an efficient point.

For any δ in $(0, 1)$, we observe that the point $(2 - \delta, 1.5)$ of X lies in the (circular) δ-neighborhood of x_{00} and

$$\mathbf{F}(2 - \delta, 1.5) = [1.5 + 2(1 - \delta)^2,\ 6.25 + 5(1 - \delta)^2] \prec [3.5, 11.25] = \mathbf{F}(2, 1.5).$$

Thus, indeed, the point $x_{00} = (2, 1.5)$ is not an efficient point for $\min_{x \in X} \mathbf{F}(x_1, x_2)$.

Theorem 5.5 *For the interval-valued functions $\boldsymbol{G}^i_{\boldsymbol{D}^p_v} : \mathbb{R}^n \to I(\mathbb{R})$, $i = 1, 2, \ldots, m$, consider the set $S = \{x \in X : \boldsymbol{G}^i_{\boldsymbol{D}^p_v}(x) \preceq \mathbf{0}$ for $i = 1, 2, \ldots, m\}$, where X is a nonempty open set in $\mathbb{R}^n$. Let $x_0 \in S$ and $I(x_0) = \{i : \boldsymbol{G}^i_{\boldsymbol{D}^p_v}(x_0) = \mathbf{0}\}$. Assuming $\boldsymbol{G}^i_{\boldsymbol{D}^p_v}$ to be gH-*

differentiable at x_0 for all $i \in I(x_0)$ and gH-continuous for $i \notin I(x_0)$, define

$$\hat{G}(x_0) = \{d : d^\top \odot \nabla \boldsymbol{G}^i_{\mathbf{D}^p_v}(x_0) \prec \mathbf{0} \text{ for all } i \in I(x_0)\}.$$

Then, $\hat{G}(x_0) \subseteq \hat{S}(x_0)$, where $\hat{S}(x_0) = \{d \in \mathbb{R}^n : d \neq 0,\ x_0 + \alpha d \in S\ \forall \alpha \in (0, \delta)$ for some $\delta > 0\}$.

Proof: Let d be an element in $\hat{G}(x_0)$.

As $x_0 \in X$ and X is an open set, there exists $\delta_0 > 0$ such that

$$x_0 + \alpha d \in X \text{ for } \alpha \in (0, \delta_0). \tag{5.3}$$

For each $i \notin I(x_0)$, as $\mathbf{G}^i_{\mathbf{D}^p_v}$ is gH-continuous at x_0,

$$\mathbf{G}^i_{\mathbf{D}^p_v}(x_0 + \alpha d) = \mathbf{G}^i_{\mathbf{D}^p_v}(x_0) \oplus \mathbf{E}^i(x_0; \alpha d),$$

where $\mathbf{E}^i(x_0; \alpha d) \to \mathbf{0}$ as $\|d\| \to 0$.

Since $\mathbf{G}^i_{\mathbf{D}^p_v}(x_0) \prec \mathbf{0}$, for $i \notin I(x_0)$, there exists $\delta_i > 0$ such that

$$\mathbf{G}^i_{\mathbf{D}^p_v}(x_0 + \alpha d) \prec \mathbf{0} \text{ for } \alpha \in (0, \delta_i) \text{ and } i \notin I(x_0). \tag{5.4}$$

Also, as $d \in \hat{G}(x_0)$, for each $i \in I(x_0)$ there exists $\delta_i > 0$ such that (see Theorem 5.3)

$$\mathbf{G}^i_{\mathbf{D}^p_v}(x_0 + \alpha d) \prec \mathbf{G}^i_{\mathbf{D}^p_v}(x_0) = \mathbf{0} \text{ for all } \alpha \in (0, \delta_i). \tag{5.5}$$

Let $\delta = \min\{\delta_0, \delta_1, \delta_2, \ldots, \delta_m\}$. Evidently, $\delta > 0$. From (5.3), (5.4) and (5.5), we see that the points of the form $x_0 + \alpha d$ belong to S for each $\alpha \in (0, \delta)$. Therefore, $d \in \hat{S}(x_0)$. Hence, $\hat{G}(x_0) \subseteq \hat{S}(x_0)$. □

Theorem 5.6 *Let X be a nonempty open set in $\mathbb{R}^n$. Consider an interval optimization* *problem*

$$\left.\begin{aligned} &\min\ \boldsymbol{F}_{\boldsymbol{C}_v^k}(x) \\ &\text{s.t.}\quad \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x) \preceq \boldsymbol{0}\ \text{ for } i = 1, 2, \ldots, m \\ &\qquad x \in X, \end{aligned}\right\} \tag{5.6}$$

where $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$ *and* $\boldsymbol{G}^i_{\boldsymbol{D}_v^p} : \mathbb{R}^n \to I(\mathbb{R})$ *for* $i = 1, 2, \ldots, m$. *For a feasible point* x_0, *define* $I(x_0) = \{i : \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) = 0\}$. *Let at* x_0, $\boldsymbol{F}_{\boldsymbol{C}_v^k}$ *and* $\boldsymbol{G}^i_{\boldsymbol{D}_v^p}$, $i \in I(x_0)$, *be gH-differentiable, and for* $i \notin I(x_0)$, $\boldsymbol{G}^i_{\boldsymbol{D}_v^p}$ *be gH-continuous. If* x_0 *is a local efficient solution of (5.6), then*

$$\hat{F}(x_0) \cap \hat{G}(x_0) = \emptyset,$$

where $\hat{F}(x_0) = \{d : d^\top \odot \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) \prec \boldsymbol{0}\}$ and $\hat{G}(x_0) = \{d : d^\top \odot \nabla \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) \prec \boldsymbol{0}$ *for each* $i \in I(x_0)\}$.

Proof: We can infer the following using Theorem 5.4 and Theorem 5.5:

$$x_0 \text{ is a local efficient solution } \implies \hat{F}(x_0) \cap \hat{S}(x_0) = \emptyset \implies \hat{F}(x_0) \cap \hat{G}(x_0) = \emptyset.$$

□

5.5.1 Unconstrained interval optimization problems

Theorem 5.7 *(Extended First Gordan's Theorem). Consider a vector* $\boldsymbol{A}_v^n = (\boldsymbol{a}_i)_{n\times 1}$ *in* $I(\mathbb{R})^n$. *Then, exactly one of the following systems has a solution:*

(i) $y^\top \odot \boldsymbol{A}_v^n \prec \boldsymbol{0}$ *for some* $y = (y_i)_{n\times 1} \in \mathbb{R}^n$,

(ii) $\boldsymbol{0}_v^n \in x \odot \boldsymbol{A}_v^n$ *for some* $x \in \mathbb{R}$, $x > 0$.

Proof: Let (i) be true. We prove that (ii) cannot be true. On contrary, if possible let (ii) be also true.

As (i) is true, we have

$$\begin{aligned}
& y_0^\top \odot \mathbf{A}_v^n \prec \mathbf{0} \text{ for some } y_0 \in \mathbb{R}^n \\
\text{or, } & x \odot (y_0^\top \odot \mathbf{A}_v^n) \prec \mathbf{0} \text{ for all } x \in \mathbb{R}, x > 0 \\
\text{or, } & y_0^\top \odot (x \odot \mathbf{A}_v^n) \prec \mathbf{0} \text{ for all } x \in \mathbb{R}, x > 0.
\end{aligned} \tag{5.7}$$

As (ii) is also true, we have

$$\begin{aligned}
& \mathbf{0}_v^n \in x_0 \odot \mathbf{A}_v^n \text{ for some } x_0 \in \mathbb{R}, x_0 > 0 \\
\text{or, } & \mathbf{0} \in y^\top \odot (x_0 \odot \mathbf{A}_v^n) \text{ for all } y \in \mathbb{R}^n.
\end{aligned} \tag{5.8}$$

As (5.7) and (5.8) cannot hold together, we have a contradiction. Thus, if (i) is true, (ii) cannot be true.

In order to prove the other case, let us assume that (i) is false. We prove that (ii) is true. On contrary, let us assume that (ii) is false. Therefore,

$$\begin{aligned}
& \mathbf{0}_v^n \notin x \odot \mathbf{A}_v^n \text{ for all } x \in \mathbb{R},\ x > 0 \\
\text{or, } & \mathbf{0}_v^n \notin \mathbf{A}_v^n \\
\text{or, } & \exists\, i \in \{1, 2, \ldots, n\} \text{ such that } 0 \notin \boldsymbol{a}_i \\
\text{or, } & \exists\, i \in \{1, 2, \ldots, n\} \text{ such that } \boldsymbol{a}_i \prec \mathbf{0} \text{ or } \mathbf{0} \prec \boldsymbol{a}_i.
\end{aligned} \tag{5.9}$$

Let us consider the sets $J = \big\{j : 0 \in \boldsymbol{a}_j,\ j \in \{1, 2, \ldots, n\}\big\}$ and $K = \{k : 0 \notin \boldsymbol{a}_k,\ k \in \{1, 2, \ldots, n\}\}$.

Evidently, by (5.9), $K \neq \emptyset$. Also, $J \cup K = \{1, 2, \ldots, n\}$ and $J \cap K = \emptyset$.

We now construct a vector $y_0 = (y_1^0, y_2^0, \ldots, y_n^0)^\top \in \mathbb{R}^n$ by

$$y_i^0 = \begin{cases} 0 & \text{if } i \in J \\ 1 & \text{if } i \in K \text{ and } \boldsymbol{a}_i \prec \mathbf{0} \\ -1 & \text{if } i \in K \text{ and } \mathbf{0} \prec \boldsymbol{a}_i. \end{cases}$$

With this $y_0 \in \mathbb{R}^n$, we note that

$$\sum_{k \in K} y_k^0 \odot \boldsymbol{a}_k \oplus \sum_{j \in J} y_j^0 \odot \boldsymbol{a}_j \prec \mathbf{0}$$
$$\text{or, } y_0^\top \odot \mathbf{A}_v^n \prec \mathbf{0}. \tag{5.10}$$

However, as (i) is false $y^\top \odot \mathbf{A}_v^n \prec \mathbf{0}$ for no $y \in \mathbb{R}^n$, which is contradictory to (5.10). Thus, (ii) must be true. Hence, the result is followed. □

Theorem 5.8 *If x_0 is a local efficient solution of the IOP:*

$$\min_{x \in \mathbb{R}^n} \quad \boldsymbol{F}_{\boldsymbol{C}_v^k}(x),$$

where $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$ is gH-differentiable at x_0. Then, $\mathbf{0}_v^n \in \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0)$.

Proof: By Definition 5.7 and Theorem 5.3, if x_0 is a local efficient solution, then $\hat{F}(x_0) = \emptyset$. Therefore, $d^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \prec \mathbf{0}$ for no $d \in \mathbb{R}^n$.

By Theorem 5.7 with $\mathbf{A}_v^n = \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0)$, there exists $x_0 \in \mathbb{R}$, $x_0 > 0$, such that

$$\mathbf{0}_v^n \in x_0 \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0)$$
$$\text{or, } \mathbf{0}_v^n \in \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0).$$

□

Note 5.3 *It is very worthy to note that the optimality condition on Theorem 5.8 is an* *inclusion relation* $\mathbf{0}_v^n \in \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0)$ *instead of the perfect equality* $\nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) = \mathbf{0}_v^n$*. Note that the optimality condition* $\nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) = \mathbf{0}_v^n$ *is not only very restrictive, but also not correct. For instance, consider the problem in Example 5.1. We observe that at the efficient point* $x_0 = (1, 1.5)$*,* $\nabla \boldsymbol{F}(x_0) = ([0,0], [-2,5]) \neq \mathbf{0}_v^2$*, but* $\mathbf{0}_v^2 \in \nabla \boldsymbol{F}(x_0)$*.*

5.5.2 Interval optimization problem with inequality constraints

Theorem 5.9 *(Extended Second Gordan's Theorem). For a matrix with interval entries* $\boldsymbol{\mathcal{A}} = \left(\boldsymbol{a}_{ij}\right)_{m\times n}$*, where* $\boldsymbol{a}_{ij} \in I(\mathbb{R})$*, exactly one of the following systems has a solution:*

(i) $\boldsymbol{\mathcal{A}}^\top \odot y \prec \mathbf{0}_v^n$ *for some* $y = (y_i)_{m\times 1} \in \mathbb{R}^m$*,*

(ii) $\mathbf{0}_v^m \in \boldsymbol{\mathcal{A}} \odot x$ *for some nonzero* $x = (x_i)_{n\times 1} \in \mathbb{R}^n$ *with all* $x_i \geq 0$*.*

Proof: Let (i) be true. Then, we show that (ii) cannot be true. On the contrary, let (ii) be true.

As (i) is true, we have

$$\begin{aligned} &\boldsymbol{\mathcal{A}}^\top \odot y_0 \prec \mathbf{0}_v^n \text{ for some } y_0 = (y_1^0, y_2^0, \ldots, y_m^0)^\top \in \mathbb{R}^m \\ \text{or, } &x^\top \odot (\boldsymbol{\mathcal{A}}^\top \odot y_0) \prec \mathbf{0} \text{ for all nonzero } x = (x_i)_{n\times 1} \in \mathbb{R}^n, x_i \geq 0 \quad (5.11) \\ \text{or, } &(\boldsymbol{\mathcal{A}} \odot x)^\top \odot y_0 \prec \mathbf{0} \text{ for all nonzero } x = (x_i)_{n\times 1} \in \mathbb{R}^n, x_i \geq 0. \quad (5.12) \end{aligned}$$

If (ii) is also true, then for some nonzero $x_0 = (x_i^0)_{n\times 1} \in \mathbb{R}^n$ with $x_i^0 \geq 0$ we have

$$\mathbf{0}_v^m \in \boldsymbol{\mathcal{A}} \odot x_0. \quad (5.13)$$

Let $\boldsymbol{w} = \boldsymbol{\mathcal{A}} \odot x_0 = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)^\top$. Then, $\boldsymbol{w} \in I(\mathbb{R})^m$ and $(\boldsymbol{\mathcal{A}} \odot x_0)^\top \odot y_0 = \sum_{i=1}^m y_i^0 \odot \boldsymbol{w}_i$.

From (5.13), we now have

$$\begin{aligned} &\mathbf{0} \in \boldsymbol{w}_i \text{ for all } i = 1, 2, \ldots, m \\ \text{or, } &\mathbf{0} \in y_i^0 \odot \boldsymbol{w}_i \text{ for all } i = 1, 2, \ldots, m \\ \text{or, } &\mathbf{0} \in (\boldsymbol{\mathcal{A}} \odot x_0)^\top \odot y_0. \end{aligned} \tag{5.14}$$

As (5.12) and (5.14) cannot hold together, we have a contradiction. Thus, if (i) is true, (ii) cannot be true. In order to prove the other case, let us assume that (i) is false. Then, we prove that (ii) must be true. As (i) is false,

$$\boldsymbol{\mathcal{A}}^\top \odot y \prec \mathbf{0}_v^n \text{ for no } y \in \mathbb{R}^m. \tag{5.15}$$

Let us assume, on contrary, that (ii) is also false. Then,

$$\begin{aligned} &\mathbf{0}_v^m \notin \boldsymbol{\mathcal{A}} \odot x \text{ for all nonzero } x = (x_i)_{n\times 1} \in \mathbb{R}^n \text{ with all } \; x_i \geq 0 \\ \text{or, } &\exists\, i \in \{1, 2, \ldots, m\} \text{ such that } 0 \notin \boldsymbol{w}_i & (5.16) \\ \text{or, } &\exists\, i \in \{1, 2, \ldots, m\} \text{ such that } \boldsymbol{w}_i \prec \mathbf{0} \text{ or } \mathbf{0} \prec \mathbf{0} \preceq \boldsymbol{w}_i, & (5.17) \end{aligned}$$

where $\boldsymbol{\mathcal{A}} \odot x = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)^\top$.

Let us consider the sets $J = \big\{j : 0 \in \boldsymbol{w}_j,\ j \in \{1, 2, \ldots, m\}\big\}$ and $K = \{k : 0 \notin \boldsymbol{w}_k,\ k \in \{1, 2, \ldots, m\}\}$.

Evidently, by (5.16), $K \neq \emptyset$. Also, $J \cup K = \{1, 2, \ldots, n\}$ and $J \cap K = \emptyset$.

We now construct a vector $y_0 = (y_1^0, y_2^0, \ldots, y_m^0)^\top \in \mathbb{R}^n$ by

$$y_i^0 = \begin{cases} 0 & \text{if } i \in J \\ 1 & \text{if } i \in K \text{ and } \boldsymbol{w}_i \prec \mathbf{0} \\ -1 & \text{if } i \in K \text{ and } \mathbf{0} \prec \boldsymbol{w}_i. \end{cases}$$

With this $y_0 \in \mathbb{R}^m$, we note that

$$\sum_{k\in K} y_k^0 \odot \boldsymbol{w}_k \oplus \sum_{j\in J} y_j^0 \odot \boldsymbol{w}_j \prec \mathbf{0}$$

$$\text{or, } y_0^\top \odot (\boldsymbol{\mathcal{A}} \odot x) \prec \mathbf{0} \text{ for all nonzero } x = (x_i)_{n\times 1} \in \mathbb{R}^n \text{ with all } x_i \geq 0$$

$$\text{or, } x^\top \odot (\boldsymbol{\mathcal{A}}^\top \odot y_0) \prec \mathbf{0} \text{ for all nonzero } x = (x_i)_{n\times 1} \in \mathbb{R}^n \text{ with all } x_i \geq 0. \tag{5.18}$$

The inequality (5.18) can be true only when $\boldsymbol{\mathcal{A}}^\top \odot y_0 \prec \mathbf{0}$. As (5.15) and (5.18) are contradictory, our assumption was wrong and (ii) must be true. Hence, the result is followed. □

Theorem 5.10 *(Extended Fritz John condition). Let X be a nonempty open set in $\mathbb{R}^n$; $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$ and $\boldsymbol{G}^i_{\boldsymbol{D}_v^p} : \mathbb{R}^n \to I(\mathbb{R})$ for $i = 1, 2, \ldots, m$ be interval-valued functions. Consider the IOP:*

$$\left.\begin{aligned} \min \quad & \boldsymbol{F}_{\boldsymbol{C}_v^k}(x), \\ \text{s.t.} \quad & \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x) \preceq \mathbf{0}, i = 1, 2, \ldots, m \\ & x \in X. \end{aligned}\right\} \tag{5.19}$$

For a feasible point x_0, define $I(x_0) = \{i : \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) = \mathbf{0}\}$. Let $\boldsymbol{F}_{\boldsymbol{C}_v^k}$ and $\boldsymbol{G}^i_{\boldsymbol{D}_v^p}$ be gH-differentiable at x_0 for $i \in I(x_0)$ and gH-continuous for $i \notin I(x_0)$. If x_0 is a local efficient point of (5.19), then there exist constants u_0 and u_i for $i \in I(x_0)$ such that

$$\begin{cases} \mathbf{0}_v^n \in \left(u_0 \odot \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) \oplus \displaystyle\sum_{i\in I(x_0)} u_i \odot \nabla \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) \right), \\ u_0 \geq 0, u_i \geq 0 \text{ for } i \in I(x_0), \\ (u_0, u_I) \neq \left(0, 0_v^{|I(x_0)|}\right), \end{cases}$$

where u_I is the vector whose components are u_i for $i \in I(x_0)$.

Further, if $\boldsymbol{G}^i_{\boldsymbol{D}_v^p}$ for all $i \notin I(x_0)$ are also gH-differentiable at x_0, then there exist *constants* $u_0, u_1, u_2, \ldots, u_m$ *such that*

$$\begin{cases} \mathbf{0}_v^n \in \left(u_0 \odot \nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) \oplus \sum_{i=1}^{m} u_i \odot \nabla \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) \right), \\ u_i \odot \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) = \mathbf{0},\ i = 1, 2, \ldots, m, \\ u_0 \geq 0, u_i \geq 0,\ i = 1, 2, \ldots, m, \\ (u_0, u) \neq (0, 0_v^m), \end{cases}$$

where u *is the vector* $(u_1, u_2, \ldots, u_m)$.

Proof: Since x_0 is a local efficient point of (5.19), by Theorem 5.6, we get

$$\hat{F}(x_0) \cap \hat{G}(x_0) = \emptyset$$

$$\text{or, } \nexists\ d \in \mathbb{R}^n \text{ s.t. } d^\top \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \prec \mathbf{0} \text{ and } d^\top \odot \nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) \prec \mathbf{0}\ \forall i \in I(x_0). \quad (5.20)$$

Let $\boldsymbol{\mathcal{A}}$ be the matrix whose columns are $\nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0)$ and $\nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0)$, $i \in I(x_0)$, i.e.,

$$\boldsymbol{\mathcal{A}} = \left[\nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0), \quad \left[\nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) \right]_{i \in I(x_0)} \right]_{n \times (1 + |I(x_0)|)}$$

By (5.20), we see that

$$\boldsymbol{\mathcal{A}}^\top \odot d \prec \mathbf{0}_v^{1+|I(x_0)|} \text{ for no } d \in \mathbb{R}^n. \quad (5.21)$$

Therefore, by Theorem 5.9, there exists a nonzero $p = (p_i)_{|I(x_0)+1| \times 1} \in \mathbb{R}^{|I(x_0)+1|}, p_i \geq 0$ such that $\mathbf{0}_v^n \in \boldsymbol{\mathcal{A}} \odot p$. Let the vector p be represented by

$$p = \begin{bmatrix} u_0 \\ u_i \end{bmatrix}_{i \in I(x_0)}. \quad (5.22)$$

Substituting (5.22) in $\mathbf{0}_v^n \in \boldsymbol{\mathcal{A}} \odot p$, we get

$$\begin{cases} \mathbf{0}_v^n \in \left(u_0 \odot \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \oplus \sum_{i \in I} u_i \odot \nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) \right), \\ u_0, u_i \geq 0 \text{ for } i \in I(x_0), \\ (u_0, u_I) \neq (0, 0, \ldots, 0). \end{cases}$$

This proves the first part of the theorem.

For $i \in I(x_0)$, $\mathbf{G}^i_{\mathbf{D}_v^p}(x_0) = \mathbf{0}$. Therefore, $u_i \odot \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) = \mathbf{0}$. If $\mathbf{G}^i_{\mathbf{D}_v^p}$ for all $i \notin I(x_0)$ are also gH-differentiable at x_0, by setting $u_i = 0$ for $i \notin I(x_0)$ the second part of the theorem is followed. □

Definition 5.10 (*Linearly independent interval vectors*). The set of m interval vectors $\{(\boldsymbol{X}_v^k)_1, (\boldsymbol{X}_v^k)_2, \ldots, (\boldsymbol{X}_v^k)_m\}$ is said to be linearly independent if for m real numbers $c_1, c_2, \ldots, c_m$:

$$\mathbf{0}_v^k \in c_1 \odot (\boldsymbol{X}_v^k)_1 \oplus c_2 \odot (\boldsymbol{X}_v^k)_2 \oplus \cdots \oplus c_m \odot (\boldsymbol{X}_v^k)_m \text{ if and only if } c_1 = 0, c_2 = 0, \ldots, c_m = 0.$$

Theorem 5.11 *(Extended Karush-Kuhn-Tucker necessary optimality condition). Let X be a nonempty open set in $\mathbb{R}^n$ and $\boldsymbol{F}_{\boldsymbol{C}_v^k} : \mathbb{R}^n \to I(\mathbb{R})$ and $\boldsymbol{G}^i_{\boldsymbol{D}_v^p} : \mathbb{R}^n \to I(\mathbb{R})$, $i = 1, 2, \ldots, m$, be interval-valued functions. Suppose x_0 be a feasible point of the IOP:*

$$\begin{cases} \min & \boldsymbol{F}_{\boldsymbol{C}_v^k}(x) \\ \text{s.t.} & \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x) \preceq \mathbf{0} \;\; i = 1, 2, \ldots, m \\ & x \in X. \end{cases}$$

Define $I(x_0) = \{i : \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) = 0\}$. Let

(i) $\boldsymbol{F}_{\boldsymbol{C}_v^k}$ and $\boldsymbol{G}^i_{\boldsymbol{D}_v^p}$ be gH-differentiable at x_0 for all $i \in I(x_0)$,

(ii) $\boldsymbol{G}^i_{\boldsymbol{D}_v^p}$ be gH-continuous for all $i \notin I(x_0)$, and

(iii) the collection of interval vectors $\{\nabla \boldsymbol{G}^i_{\boldsymbol{D}^p_v}(x_0) \ : \ i \in I(x_0)\}$ *is linearly independent.*

If x_0 *is a local efficient solution, then there exist constants* u_i *for all* $i \in I(x_0)$ *such that*

$$\begin{cases} \mathbf{0}^n_v \in \left(\nabla \boldsymbol{F}_{\boldsymbol{C}^k_v}(x_0) \oplus \sum_{i \in I} u_i \odot \nabla \boldsymbol{G}^i_{\boldsymbol{D}^p_v}(x_0)\right), \\ u_i \geq 0 \textit{ for all } i \in I(x_0). \end{cases}$$

If $\boldsymbol{G}^i_{\boldsymbol{D}^p_v}$ *for* $i \notin I(x_0)$ *are also gH-differentiable at* x_0*, then there exist constants* u_1*,* u_2*,* $\ldots$*,* u_m *such that*

$$\begin{cases} \mathbf{0}^n_v \in \left(\nabla \boldsymbol{F}_{\boldsymbol{C}^k_v}(x_0) \oplus \sum_{i=1}^{m} u_i \odot \nabla \boldsymbol{G}^i_{\boldsymbol{D}^p_v}(x_0)\right), \\ u_i \odot \boldsymbol{G}^i_{\boldsymbol{D}^p_v}(x_0) = \mathbf{0}, \ i = 1, 2, \ldots, m, \\ u_i \geq 0, \ i = 1, 2, \ldots, m. \end{cases}$$

Proof: By Theorem 5.10, there exist real constants u_0 and u'_i for all $i \in I(x_0)$, not all zeros, such that

$$\left.\begin{array}{l} \mathbf{0}^n_v \in \left(u_0 \odot \nabla \mathbf{F}_{\mathbf{C}^k_v}(x_0) \oplus \sum_{i \in I} u'_i \odot \nabla \mathbf{G}^i_{\mathbf{D}^p_v}(x_0)\right) \\ u_0 \geq 0, u'_i \geq 0 \text{ for all } i \in I(x_0). \end{array}\right\}$$

Then, we must have $u_0 > 0$. Since otherwise, the set $\{\nabla \mathbf{G}^i_{\mathbf{D}^p_v}(x_0) \ : \ i \in I(x_0)\}$ will become not linearly independent.

Define $u_i = u'_i/u_0$. Then, $u_i \geq 0$ for all $i \in I(x_0)$ and $\mathbf{0}^n_v \in \left(\nabla \mathbf{F}_{\mathbf{C}^k_v}(x_0) \oplus \sum_{i \in I} u_i \odot \nabla \mathbf{G}^i_{\mathbf{D}^p_v}(x_0)\right)$. For $i \in I(x_0)$, $\mathbf{G}^i_{\mathbf{D}^p_v}(x_0) = \mathbf{0}$. Therefore, $0 \in u_i \mathbf{G}^i_{\mathbf{D}^p_v}(x_0)$. If the functions $\mathbf{G}^i_{\mathbf{D}^p_v}$ for $i \notin I(x_0)$ are also gH-differentiable at x_0, then by setting $u_i = 0$ for $i \notin I(x_0)$, the latter part of the theorem is followed. □

Example 5.2 *In this example, we verify the* extended Fritz John condition *(Theorem*

5.10) *and the* extended Karush-Kuhn-Tucker necessary optimality condition *(Theorem 5.11) on the following IOP at the feasible point* $x_0 = (0, 2) \in \mathbb{R}^2$*:*

$$\begin{aligned} \min \quad & \boldsymbol{F}_{\boldsymbol{C}_v^4}(x_1, x_2) = [-3, 0] \odot x_1^2 \oplus [0, 1] \odot x_2^3 \oplus [-2, -1] \odot x_2^2 \oplus [1, 2] \odot (x_1^2 x_2) \\ s.t. \quad & \boldsymbol{G}^1_{\boldsymbol{D}_v^3}(x_1, x_2) = [-2, 3] \odot x_1 \oplus [-2, -1] \odot x_2 \ominus_{gH} [-4, -2] \preceq \boldsymbol{0} \\ & \boldsymbol{G}^2_{\boldsymbol{D}_v^3}(x_1, x_2) = [1, 2] \odot x_1^2 \oplus [-5, -3] \odot x_2 \ominus_{gH} [-1, 0] \preceq \boldsymbol{0}. \end{aligned}$$

Here, the functions $\boldsymbol{F}_{\boldsymbol{C}_v^4}$, $\boldsymbol{G}^1_{\boldsymbol{C}_v^3}$ *and* $\boldsymbol{G}^2_{\boldsymbol{C}_v^3}$ *are* gH*-differentiable on* $\mathbb{R}^2$.
At x_0*, we note that* $\boldsymbol{G}^1_{\boldsymbol{D}_v^3}(x_0) = \boldsymbol{0}$ *and* $\boldsymbol{G}^2_{\boldsymbol{D}_v^3}(x_0) = [-11, -6]$*. Hence,* $I(x_0) = \{1\}$*.*
We observe that

$$\begin{aligned} \nabla \boldsymbol{F}_{\boldsymbol{C}_v^4}(x_0) &= \left(D_1 \boldsymbol{F}_{\boldsymbol{C}_v^5}(0, 2), D_2 \boldsymbol{F}_{\boldsymbol{C}_v^5}(0, 2)\right)^\top = \left(\boldsymbol{0}, [-8, 8]\right)^\top \text{ and} \\ \nabla \boldsymbol{G}^1_{\boldsymbol{D}_v^3}(x_0) &= \left(D_1 \boldsymbol{G}^1_{\boldsymbol{D}_v^3}(0, 2), D_2 \boldsymbol{G}^1_{\boldsymbol{D}_v^3}(0, 2)\right)^\top = \left([-2, 3], [-2, -1]\right)^\top. \end{aligned}$$

Taking $u_0 = 2$*,* $u_1 = 1$ *and* $u_2 = 0$ *we see that conclusions of the Theorem 5.10 is true. Taking* $u_0 = 1$*,* $u_1 = 1$ *and* $u_2 = 0$ *we see that conclusions of the Theorem 5.11 holds true.*

Theorem 5.12 *(Extended Karush-Kuhn-Tucker sufficient optimality condition). Let* X *be a nonempty open convex set in* $\mathbb{R}^n$*;* $\boldsymbol{F}_{\boldsymbol{C}_v^k} : X \to I(\mathbb{R})$ *and* $\boldsymbol{G}^i_{\boldsymbol{D}_v^p} : X \to I(\mathbb{R})$*,* $i = 1, 2, \ldots, m$*, be interval-valued* gH*-differentiable convex functions on* X*. Suppose* x_0 *be a feasible point of the IOP:*

$$\begin{cases} \min & \boldsymbol{F}_{\boldsymbol{C}_v^k}(x) \\ \text{s.t.} & \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x) \preceq \boldsymbol{0}, \ i = 1, 2, \ldots, m \\ & x \in X. \end{cases}$$

If there exist real constants $u_1, u_2, \ldots, u_m$ for which

$$\begin{cases} \mathbf{0}_v^n \in \left(\nabla \boldsymbol{F}_{\boldsymbol{C}_v^k}(x_0) \oplus \sum_{i=1}^{m} u_i \odot \nabla \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) \right), \\ u_i \odot \boldsymbol{G}^i_{\boldsymbol{D}_v^p}(x_0) = \mathbf{0},\ i = 1, 2, \ldots, m. \\ u_i \geq 0,\ i = 1, 2, \ldots, m. \end{cases}$$

then x_0 is an efficient point of the IOP.

Proof: By the hypothesis, for every $x \in X$ satisfying $\mathbf{G}^i_{\mathbf{D}_v^p}(x) \preceq \mathbf{0}$ for all $i = 1, 2, \ldots, m$, we have

$$\begin{aligned} \mathbf{0}_v^n \quad &\in \left(\nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \oplus \sum_{i=1}^{n} u_i \odot \nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) \right)^{\top} (x - x_0) \\ &= \nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0)^{\top}(x - x_0) \oplus \sum_{i=1}^{m} u_i \odot \nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0)^{\top}(x - x_0) \\ &\preceq \left(\mathbf{F}_{\mathbf{C}_v^k}(x) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_0) \right) \oplus \sum_{i=1}^{m} u_i \odot \left(\mathbf{G}^i_{\mathbf{D}_v^p}(x) \oplus (-1) \odot \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) \right), \\ &\qquad \text{by Theorem 5.2} \\ &\preceq \mathbf{F}_{\mathbf{C}_v^k}(x) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_0). \end{aligned}$$

Thus, for every $x \in X$,

$$\text{either} \quad \mathbf{0}_v^n \in \mathbf{F}_{\mathbf{C}_v^k}(x) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_0) \quad \text{or} \quad \mathbf{0}_v^n \preceq \mathbf{F}_{\mathbf{C}_v^k}(x) \oplus (-1) \odot \mathbf{F}_{\mathbf{C}_v^k}(x_0).$$

In either case, x_0 is an efficient point of the considered IOP. Hence, the result is followed. □

5.5.3 Comparison with existing KKT conditions for IOPs

In this section, we make a comparison of the proposed KKT optimality condition with the existing ones. The comparison is based on

(i) the use of the lower function $\underline{F}$ and the upper function $\overline{F}$,

(ii) the use of gH-derivative, and

(iii) appearance (inclusion or equation) of the optimality condition.

To the best of the knowledge of the authors the existing articles on KKT theory for IOPs are [97], [99], [111], [112] and [100].

In developing KKT theory for IOPs, the articles [97], [99], [111] and [112], used the idea of H-derivative for interval-valued functions. However, the notion of H-derivative is very restrictive for interval-valued functions since H-derivative may not exist for very simple interval-valued functions (for details, see Subsection 2.2 of [100]). For instance, the function $\boldsymbol{F}(x) = (1 - x^5) \odot [-3, 1]$ does not have H-derivative at $x = 0$.
Further, the KKT theory for IOPs in [97], [99], [111] and [112] depends on existence of the derivative of the lower and upper functions, $\underline{F}$ and $\overline{F}$, respectively, of a given interval-valued function $\boldsymbol{F}(x) = [\underline{F}(x), \overline{F}(x)]$. This assumptions is also very difficult to follow even for very simple functions. For instance, $\underline{F}(x_1, x_2)$ and $\overline{F}(x_1, x_2)$ for the function $\boldsymbol{F}_{C_v^2}(x_1, x_2) = [-1, 1] \odot x^2 \oplus [0, 2] \odot x_2$ are not differentiable (for details, see Example 1 of [113]).

Additionally, in [97], [99], [111], [101] and [112] the KKT condition

$$\mathbf{0}_v^n = \left(\nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \oplus \sum_{i=1}^{m} u_i \odot \nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) \right) \tag{5.23}$$

is very restrictive as there is hardly a few interval-valued functions which follow this condition.

Several other deficiencies of the approaches of [97] and [99] and [111] and [112] are reported by Chalco-Cano et al. [100]. However, in [100] the objective function $\boldsymbol{F}(x)$ is taken in the form of $[\underline{F}(x), \overline{F}(x)]$. This is also not a mild assumption and it is not always an easy task to find the expressions of $\underline{F}(x)$ and $\overline{F}(x)$ even for very simple interval-valued function. For instance, consider the function

$$\boldsymbol{F}(x_1, x_2) = \frac{[-1, 2] \odot \sin x_1 + [-2, 1] \odot x_2}{[-1, 2] \odot \cos x_2 + [-2, 1] \odot x_1}.$$

Besides, the KKT condition in [100] also appear alike to (5.23). This is also very restrictive and very rare functions have gH-derivative *exactly* the 'zero' vector.

However, note that the studied KKT condition in this chapter

$$\mathbf{0}_v^n \in \left(\nabla \mathbf{F}_{\mathbf{C}_v^k}(x_0) \oplus \sum_{i=1}^{m} u_i \odot \nabla \mathbf{G}^i_{\mathbf{D}_v^p}(x_0) \right) \tag{5.24}$$

is very flexible as it appears with the inclusion relation instead of *perfectly* equal the to zero vector. Also this chapter uses gH-derivative of interval-valued functions which is the most general (see [100,105,113]) definition of derivative for interval-valued functions.

5.6 Application to Support Vector Machines

SVMs are generally used in solving classification problems. Here we consider a binary classification problem. For a given data set $D = \{(x_i, y_i) : x_i \in \mathbb{R}^n,\ y_i \in \{-1, 1\}, i = 1, 2, \ldots, m\}$, the problem of classifying data using SVMs is equivalent to the following optimization problem:

$$\left.\begin{aligned} &\min_{w,b}\ F(w, b) = \tfrac{1}{2}\|w\|^2 \\ &\text{s.t.}\quad y_i(w^\top x_i + b) \geq 1,\ i = 1, 2, \ldots, m, \end{aligned}\right\} \tag{5.25}$$

where $w \in \mathbb{R}^n$ is the weight vector and $b \in \mathbb{R}$ is the bias. The constraints represent the condition that the data points lie on either side of the separating hyperplanes

$w^\top x + b = \pm 1$.

In many classification problems, the data set may not be precise and thus involves uncertainty. This may be due to errors in measurement, implementation, etc. For example, let us assume we want to predict whether it will rain tomorrow or not. The data we may require are the wind speed, humidity levels, temperature, etc. These variables usually have values in intervals like 10–13 km/hr wind speed, $40 - 50$ percent humidity, $30 - 35^o C$ temperature, etc. The standard SVM formulation is not applicable for such data as it is interval-valued, whereas the problem (5.25) requires real-valued data. Thus, we adjust the SVM problem for the interval-valued data set

$$\big\{(\boldsymbol{X}_i, y_i) \ : \ \boldsymbol{X}_i \in I(\mathbb{R})^n, \ y_i \in \{-1, 1\}, i = 1, 2, \ldots, m\big\}$$

by

$$\left.\begin{aligned} &\min_{w,b} \ \ F(w, b) = \tfrac{1}{2}\|w\|^2 \\ &\text{s.t.} \ \ \mathbf{G}^i_{\mathbf{D}^n_v}(w, b) = [1, 1] \ominus_{gH} y_i \odot \Big(w^\top \odot \boldsymbol{X}_i \oplus b\Big) \preceq \mathbf{0}, \quad i = 1, 2, \ldots, m. \end{aligned}\right\} \tag{5.26}$$

We notice that the functions $F(w)$ and $\mathbf{G}^i_{\mathbf{D}^n_v}$ are gH-differentiable and convex. The gH-gradients of these functions are

$$\begin{aligned} \nabla F(w, b) &= \big(D_1 F(w, b), D_2 F(w, b)\big)^\top = (w, 0)^\top \\ \text{and } \nabla \mathbf{G}^i_{\mathbf{D}^n_v}(w, b) &= \Big(D_1 \mathbf{G}^i_{\mathbf{D}^n_v}(w, b), D_2 \mathbf{G}^i_{\mathbf{D}^n_v}(w, b)\Big)^\top = (-y_i \odot \boldsymbol{X}_i, -y_i)^\top, \end{aligned}$$

where D_1 and D_2 are the gH-partial derivatives with respect to w and b, respectively. According to Theorem 5.11, for an efficient point (w^*, b^*) of (5.26) there exist nonnegative scalars $u_1, u_2, \ldots, u_m$ such that

$$\mathbf{0}_v^{n+1} \in \left((w^*, 0)^\top \oplus \sum_{i=1}^{m} u_i \odot (-y_i \odot \boldsymbol{X}_i, -y_i)^\top \right) \tag{5.27}$$

$$\text{and} \quad \mathbf{0} = u_i \odot \mathbf{G}^i_{\mathbf{D}_v^n}(w^*, b^*), \; i = 1, 2, \ldots, m. \tag{5.28}$$

The condition (5.27) can be simplified as

$$\begin{aligned} \mathbf{0}_v^n &\in \left([w^*, w^*] \oplus \sum_{i=1}^{m} (-u_i y_i) \odot \boldsymbol{X}_i \right) \\ \text{and} \quad &\sum_{i=1}^{m} u_i y_i = 0. \end{aligned}$$

The data points $\boldsymbol{X}_i$ for which $u_i \neq 0$ are called support vectors. By (5.28), we observe that corresponding to any $u_i > 0$, we have $\mathbf{G}^i_{\mathbf{D}_v^n}(w^*, b^*) = \mathbf{0}$. Thus, corresponding to w^*, the value of the bias b^* is such a quantity that $\mathbf{G}^i_{\mathbf{D}_v^n}(w^*, b^*) = \mathbf{0}$ for all of those $i \in \{1, 2, \ldots, m\}$ for which $u_i > 0$.

Hence, as the functions $F(w, b)$ and $\mathbf{G}^i_{\mathbf{D}_v^n}(w, b)$ are gH-differentiable and convex, by Theorems 5.11 and 5.12, the set of conditions solving which we obtain the efficient solutions of the SVM IOP (5.26) are

$$\left. \begin{aligned} \mathbf{0}_v^n &\in \left([w, w] \oplus \sum_{i=1}^{m} (-u_i y_i) \odot \boldsymbol{X}_i \right), \\ &\sum_{i=1}^{m} u_i y_i = 0 \\ \text{and} \quad &\mathbf{0} = u_i \odot \mathbf{G}^i_{\mathbf{D}_v^n}(w, b), \; i = 1, 2, \ldots, m. \end{aligned} \right\} \tag{5.29}$$

Corresponding to any of the value of w that satisfies (5.29), we define the set of possible values of the bias by

$$\bigcap_{i:\; u_i > 0} \left\{ b \; : \; \mathbf{G}^i_{\mathbf{D}_v^n}(w, b) = \mathbf{0} \right\}. \tag{5.30}$$

Using any solution $\bar{w}$ and $\bar{b}$ of (5.29) and (5.30), a classifying hyperplane and the SVM classifier function are given by:

$$\bar{w}^\top \boldsymbol{X} + \bar{b} = \mathbf{0} \text{ and } s^*(\boldsymbol{X}) = \operatorname{sign}\left(\bar{w}^\top \boldsymbol{X} + \bar{b}\right).$$

Example 5.3 *Consider the interval data set*

$$\begin{aligned}
\mathbf{X}_1 &= \Big[[3,4],[1,2]\Big], y_1 = 1, & \mathbf{X}_2 &= \Big[[4,5],[2,3]\Big], y_2 = 1,\\
\mathbf{X}_3 &= \Big[[5,6],[1,2]\Big], y_3 = 1, & \mathbf{X}_4 &= \Big[[0,1],[1,2]\Big], y_4 = -1,\\
\mathbf{X}_5 &= \Big[[1,2],[2,3]\Big], y_5 = -1, & \mathbf{X}_6 &= \Big[[0,2],[3,4]\Big], y_6 = -1.
\end{aligned}$$

For this data set we find a classifying hyperplane with the help of the IOP SVM (5.26).

In order to find a classifying hyperplane, we need to find a possible solution (w, b) of (5.29) along with the corresponding u_i's.

We observe that for $(u_1, u_2, u_3, u_4, u_5, u_6) = (1, 0, 0, 0, 1, 0)$ we have $\sum_{i=1}^{6} u_i y_i = 0$. For these values of u_i's, the first condition in (5.29) reduces to

$$\begin{aligned}
&\mathbf{0}_v^n \in \big([w,w] \oplus (-1) \odot \boldsymbol{X}_1 \oplus \boldsymbol{X}_5\big)\\
&\text{or, } [w,w] \in (-1) \odot ((-1) \odot \boldsymbol{X}_1 \oplus \boldsymbol{X}_5)\\
&\text{or, } w \in ([1,3],[-2,0]). \qquad (5.31)
\end{aligned}$$

Denoting $w = (w_1, w_2) \in \mathbb{R}^2$, the condition (5.31) reduces to

$$1 \le w_1 \le 3 \text{ and } -2 \le w_2 \le 0. \qquad (5.32)$$

Let us choose $w_1^* = 1$ and $w_2^* = -2$. Corresponding to this $w^* = (w_1^*, w_2^*) = (1, -2)$, from (5.30) and the third condition in (5.29), the set of possible values of the bias b is given by

$$\begin{aligned}
&\bigcap_{i=1,5}\left\{b\in\mathbb{R}\ :\ \mathbf{G}^i_{\mathbf{D}^2_v}(w^*,b)=\mathbf{0}\right\}\\
&=\left\{b\in\mathbb{R}\ :\ \mathbf{G}^1_{\mathbf{D}^2_v}(w^*,b)=\mathbf{0}\right\}\bigcap\left\{b\in\mathbb{R}\ :\ \mathbf{G}^5_{\mathbf{D}^2_v}(w^*,b)=\mathbf{0}\right\}\\
&=\{b\in\mathbb{R}\ :\ b\in[-2,1]\}\cap\{b\in\mathbb{R}\ :\ b\in[-6,-1]\}\\
&=\{b\in\mathbb{R}\ :\ -2\le b\le -1\}.
\end{aligned}$$

Thus corresponding to $w_1^*=1$ and $w_2^*=-2$ the set of classifying hyperplanes is given by

$$\begin{aligned}
&w_1^*x_1+w_2^*x_2+b=0,\ -2\le b\le -1\\
\text{i.e.,}\quad &x_1-2x_2+b=0,-2\le b\le -1.
\end{aligned}$$

For any choice of b in $[-2,-1]$, note that the value of the objective function F is identical (and it is $\frac{5}{2}$).

5.6.1 Comparison with existing solutions to interval uncertainty in SVM

5.6.1.1 Robust optimization

In robust optimization approach, the uncertain parameters are assumed to take their worst case values and the constraints are to be satisfied for all the possible values for the uncertain parameters. Ben-Tal *et al.* [114] have applied robust optimization to solve interval uncertainty in SVM. The input data having interval uncertainty is given by the uncertainty set

$$X_u=\left\{\boldsymbol{X}_i\in I(\mathbb{R})^n:\ \boldsymbol{X}_i=X_i^{\text{nom}}+\sigma_i,\ \|\sigma_i\|_\infty\le\kappa,\ i=1,2,\dots,m\right\},\tag{5.33}$$

where κ is a positive constant. This means that each input point $\boldsymbol{X}_i$ lies inside the $\|.\|_\infty$-ball of radius κ centred at the nominal data X_i^{nom}, where for $p = (p_1, p_2, \ldots, p_n) \in \mathbb{R}^n$, $\|p\|_\infty = \max\{|p_1|, |p_2|, \ldots, |p_n|\}$.

The data set used for binary classification is represented as $D = \{(\boldsymbol{X}_i, y_i) : \ \boldsymbol{X}_i \in X_u,\ y_i \in \{-1, 1\}, i = 1, 2, \ldots, m\}$. Using robust optimization, in [114] the constraints given in classical SVM formulation in the problem (5.25) is reduced to

$$y_i(w^\top X_i^{\text{nom}} + b) \geq \rho_i \|w\|_1,\ \ i = 1, 2, \ldots, m,$$

where $\|w\|_1 = \sum_{i=1}^n |w_i|$ and $\rho_i = \|\sigma_i\|_\infty$.

The robust maximum margin classifier is thus obtained by (see in [114]) solving the so-called robust counterpart

$$\left.\begin{aligned} &\min_{w,b} \quad \|w\|_1, \\ &\text{s.t.} \quad y_i(w^\top X_i^{\text{nom}} + b) \geq \rho_i, \quad i = 1, 2, \ldots, m. \end{aligned}\right\} \tag{5.34}$$

Applying (5.34) on the data set given in Example 5.3, we get the optimal classifying line as $2x_1 - \frac{4}{3}x_2 - \frac{5}{3} = 0$. We notice that this line lies inside the set of lines derived in Example 5.3 using interval optimization ($w_1^* = 2 \in [1, 3],\ \leq w_2^* = -\frac{4}{3} \in [-2, 0],\ b^* = -\frac{5}{3} \in [-2, -1]$). Thus, robust optimization technique provides a single solution by solving the worst case possibility for the uncertain parameters and ignores all other possibilities of the solution.

However, the proposed approach characterizes and obtain the complete solution set of the IOP SVM problem.

5.6.1.2 Optimization using boundary functions

In [115], the application of SVMs in regression is considered where the data set has interval uncertainty. The authors have used the classical SVM formulation and applied duality results directly to an interval optimization problem. For the interval vectors, they have taken the upper or lower bound whichever maximizes their objective function. This ignores the overall set of values and takes only the best case scenario. Here a natural question arises: if the data set is uncertain, how can the solution be certain? It is noteworthy that the proposed approach does not reduce the optimization problem for the interval-data to the best or worst case analysis. Rather, it carries the interval uncertainty till the end of the final decision.

5.7 Conclusion

In this chapter, we have considered the problem of interval optimization for constrained IOPs with the aim of characterizing the efficient solution from a geometrical viewpoint. We have proposed extensions to Gordon's Theorems of the alternatives for an interval-valued system of inequalities and used it to derive the Fritz John conditions for IOPs. We also derived an extension to KKT conditions for IOPs and thereby proposed the optimality conditions for both constrained and unconstrained IOPs. These proposed optimality conditions have been applied to binary classification problem using SVMs for interval-valued dataset and a comparison has been drawn with existing methods. We have identified a set of efficient solutions for the formulated SVM problem using proposed extended KKT conditions.

****