

---

# Extremality, Stationarity and Generalized Separation of Collections of Sets

Hoa T. Bui · Alexander Y. Kruger

Received: date / Accepted: date

**Abstract** The core arguments used in various proofs of the *extremal principle* and its extensions as well as in primal and dual characterizations of *approximate stationarity* and *transversality* of collections of sets are exposed, analyzed and refined, leading to a unifying theory, encompassing all existing approaches to obtaining ‘extremal’ statements. For that, we examine and clarify quantitative relationships between the parameters involved in the respective definitions and statements. Some new characterizations of extremality properties are obtained.

**Keywords** extremal principle · approximate stationarity · transversality · regularity · separation

**Mathematics Subject Classification (2000)** 49J52; 49J53; 49K40; 90C30

## 1 Introduction

Models involving pairs, or more generally finite (and even infinite) collections of sets, are pretty common in various fields of mathematics, especially optimization and neighbouring areas. For instance, the classical *separation theorem* [1–3] for convex sets, one of the key results of nonlinear analysis, is instrumental when establishing multiplier rules or sub-differential calculus rules; the ubiquitous *feasibility problems* [4, 5] cover solving systems of (generalized) equations, while appropriate *transversality* (regular intersection) conditions produce *constraint qualifications* and convergence estimates for *alternating projections* (*von Neuman method*) [6–9], or more generally *cyclic projections*, which in turn provide a convenient model for convergence analysis of computational algorithms. On the other hand, (local) *optimality* in optimization problems can naturally be interpreted as ‘irregular’ (extremal) intersection of certain sets, and sufficient transversality conditions immediately produce necessary optimality conditions.

In the framework of convex analysis, the property of two sets that one set does not meet the interior of the other set (while both are nonempty), which is the key assumption of the conventional separation theorem, provides an example of the absence of regular intersection. Thus, the separation theorem can be interpreted as a characterization of the absence of regular intersection of convex sets. On the other hand, the opposite property, i.e., when the intersection of one set with the interior of the other set is nonempty, is a typical qualification condition in various convex analysis statements [2].

There have been many successful applications of convex analysis, and particularly the separation theorem in the nonconvex settings by considering appropriate local convex approximations of sets and functions. The most prominent example in the optimization area is probably given by the *Dubovitskii–Milyutin formalism* [10]. The Clarke tangent cone (9) is an important example of a local convex approximation of a set.

However, in many situations it is impossible to construct satisfactory local convex approximations of nonconvex sets. For instance, if a set consists of two intersecting lines on the plain, it is easy to check that its Clarke tangent cone at the point of intersection is trivial (contains only the zero vector), and provides no meaningful information about the set. The powerful tools of convex analysis generally fail outside of the comfortable convex setting.

---

Hoa T. Bui · Alexander Y. Kruger (✉)

Centre for Informatics and Applied Optimization, School of Science, Engineering and Information Technology, Federation University Australia, POB 663, Ballarat, Vic, 3350, Australia

E-mail: h.bui@federation.edu.au, a.kruger@federation.edu.au

A new tool with the potential to substitute the conventional separation theorem in the general nonconvex environment — currently known as the *Extremal principle* — was suggested in 1979–1980 in [11–13]. It was observed that the separation theorem actually characterizes a kind of extremal arrangement of sets. Building on this observation, the key assumption of the conventional separation theorem, that one set does not meet the interior of the other set, was replaced by a more general geometric (local) *extremality* property that arbitrarily small translations of the sets make their intersection empty (in a neighbourhood of a given *extremal point*); cf. Definition 2.1(i) and (ii). This property assumes neither convexity of the sets nor that one of the sets has nonempty interior, while still embracing many conventional (and generalized) optimality notions. Moreover, it is applicable to any finite ( $n \geq 2$ ) collections of sets (cf. [11–13]) and, if the space is Banach and the sets are closed, the function measuring the distance between points in different translated sets satisfies the assumptions of the Ekeland variational principle. Employing an appropriate sum rule to the perturbation of the distance function arising from the application of the Ekeland variational principle, one can formulate dual characterizations of extremality in terms of appropriate normals to individual sets. Such dual conditions can be interpreted as a kind of generalized separation; cf. conditions (ii) and (iii) in Theorem 2.2.

All existing infinite dimensional versions of the extremal principle are based on the Ekeland variational principle and differ mainly in the kind of sum rule used in the proof and, accordingly, the kind of normals used in the statement. In the beginning, the statements were restricted to *Fréchet smooth* spaces (i.e. Banach spaces admitting an equivalent norm, Fréchet differentiable away from zero); cf. [12–15]. This allowed application of the *differentiable sum rule* (Lemma 2.1(ii)); the generalized separation conditions were formulated in terms of sets of  $\varepsilon$ -normals [16]. The Asplund space *approximate sum rule* for Fréchet subdifferentials (Lemma 2.1(iii)) established by Fabian [17] in the end of the 1980s allowed extending the extremal principle to Asplund spaces and replacing sets of  $\varepsilon$ -normals in the generalized separation conditions with the Fréchet normal cones. This Asplund space version of the extremal principle appeared in 1996 in [18]. It was also shown in [18] that with the Fréchet normal cones the result cannot be in general extended beyond Asplund spaces.

Similarly, plugging the *Clarke–Rockafellar subdifferential sum rule* (Lemma 2.1(iv)) into the proof, one can formulate a general Banach space version of the extremal principle in terms of the Clarke normal cones. Of course, since the Fréchet normal cone is a subset of the Clarke one and the latter is often unreasonably large, when the space is Asplund, the version in terms of Fréchet normals gives better estimates, unless the sets are convex. In the latter case, the estimates coincide. Several publications have gone a little further and considered abstract subdifferentials possessing appropriate sum rules (in the respective *trustworthy* [19] spaces) and the corresponding normal cones; see e.g. [19–22].

Since its inception almost 40 years ago, the extremal principle has indeed impacted strongly on the nonconvex analysis and optimization substituting the conventional separation theorem. Several examples of its application to proving necessary optimality conditions (multiplier rules) can be found already in the very first publications [11–13]. Numerous applications of the extremal principle in multiple publications on optimization and analysis are exposed and commented on in the monograph [21]. The original publications have been followed by several studies of the concept of extremality of collections of sets and the extremal principle, resulting in its additional characterizations and several extensions.

It was established in [23, 24] that, similar to the classical Lagrange multiplier rule, the conclusion of the extremal principle (the generalized separation), being a dual necessary condition of extremality, actually characterizes a property which is weaker than local extremality. This property can be interpreted as a kind of stationarity of a collection of sets. The explicit primal space definition of this property was introduced (cf. Definition 2.1(iv)), and it was shown (by refining slightly the original proof of the extremal principle) that for this property the conclusion of the extremal principle is not only necessary (in the Asplund space setting), but also sufficient, thus producing the ultimate conventional version of the extremal principle, known as the *extended extremal principle*; cf. Theorem 2.2. This stationarity property called *approximate stationarity* [25] happens to be strongly connected to (in fact the negation of) another important geometric property of a collection of sets called *transversality* (also known under other names; cf. Definition 3.2 and Theorem 3.1), having its roots in the classical differential geometry and instrumental for the convergence analysis of alternating projections; see a discussion of the role of the latter property in [5], while a table illustrating the evolution of the terminology can be found in [26, Section 2]. The connections of the approximate stationarity and (extended) extremal principle can be traced further to the fundamental for variational analysis *metric regularity* property of set-valued mappings and the corresponding coderivative criterion; see [25].

The powerful (extended) extremal principle, despite its recognized universality and wide applicability, has its limitations. The second author encountered a problem while attempting to extend the extremal principle to infinite collections

of sets in [22]. The initial idea to consider families of finite subcollections and apply the conventional extremal principle to each of them failed, as well as the belief of the authors in the unlimited universality of the conventional extremal principle. Uniform estimates were required, holding for all finite subcollections, and the conventional extremal principle was unable to provide such estimates. The solution was found within the conventional (extended) extremal principle, more precisely, in its proof, providing another piece of evidence for the well-known fact that important mathematical results are often deeper than their statements.

The conventional extended extremal principle asserts the equivalence of two properties of a collection of sets: one in the primal space (approximate stationarity) and another one in the dual space (generalized separability); see Theorem 2.2, which for completeness gives two (equivalent) versions of the latter property. Both (in the case of Theorem 2.2 all three) properties are formulated “for any  $\varepsilon$  there exist ...”, and as such, the  $\varepsilon$ 's and the other parameters in different properties seem completely independent. However, this cannot be true. To prove that one property of this kind implies another one, there must be a way, given any  $\varepsilon$  in the second property, to construct another one to be plugged into the first property to produce the needed estimates, i.e., the parameters must be related. The relationship between the parameters in the two parts of the conventional (extended) extremal principle is exactly what was required to establish the uniform estimates needed in [22], and it indeed could be found in the proof of the conventional result. This observation made the authors of [22] carve out the core part of (the proof of) the conventional (extended) extremal principle with  $\varepsilon$  and other parameters fixed (a kind of  $\varepsilon$ -extremality), and formulate it as a separate statement in [22, Theorem 3.1] (see Proposition 6.1(iii) below), exposing the relationship between the parameters hidden in the proof of the conventional statement. As a result, both the conventional statement and its extension to infinite collections of sets are corollaries of [22, Theorem 3.1].

There are many equivalent formulations of the primal space approximate stationarity property (as well as other extremality and stationarity properties) and the dual space generalized separability properties, all formulated in the form “for any  $\varepsilon$  there exist ...”; see Section 3. The relationships between the parameters involved in these formulations can also be established, and are important for identifying and analyzing the core arguments in the conventional proofs of metric and dual characterizations of the extremality and stationarity and their extensions. This analysis is performed in Sections 5 and 6.

Another successful ‘surgical operation’ on the proof of the conventional extremal principle was done earlier by Zheng and Ng in [27], producing an  $\varepsilon$ -separability characterization of another kind of  $\varepsilon$ -extremality property with the explicit relationship between the  $\varepsilon$ 's in both properties. Unlike the conventional extremal principle and its extension in [22, Theorem 3.1] which assume the sets to have a common point, [27, Lemmas 2.2 and 2.2'] assume, on the contrary, that the intersection of the sets is empty. At the same time, their proofs follow the original ideas from [12, 13], utilizing the Ekeland variational principle and either the Asplund space fuzzy or the general Banach space Clarke–Rockafellar subdifferential sum rule; cf. Lemma 2.1 below. The statements have been further polished and analyzed in a sequence of subsequent papers [28–31]. In particular, it was proved by Guoyin Li et al in [30, Theorem 3.1] that the conclusion of [27, Lemma 2.2] is actually equivalent to the Ekeland variational principle, and as such implies (the Banach space with Clarke normal cones version of) the extremal principle. This fact confirms the need to move from the conventional “for any  $\varepsilon$  ...” extremality statements to more subtle ones with  $\varepsilon$  fixed. The most advanced version of the Zheng and Ng lemma was given in [31, Theorems 3.1 and 3.4] (*unified separation theorems*; see Theorem 6.2 below), where, besides other improvements, an additional condition was added to the concluding part, relating the dual vectors involved in the  $\varepsilon$ -separability characterization with certain primal space vectors involved in the original  $\varepsilon$ -extremality property.

A partial comparison of the assumptions and conclusions in the two existing  $\varepsilon$ -extremality statements in [22, Theorem 3.1] and [31, Theorems 3.1 and 3.4] has been done recently in [32], where it was also noted that they are formulated using in a sense different languages and are in general incomparable. In the current paper, we formulate in Theorem 6.1 (and prove) a new general  $\varepsilon$ -extremality statement which exposes the core arguments and the role of the parameters in the conventional proofs of dual characterizations of the extremality and stationarity and, in particular, implies [22, Theorem 3.1] and [31, Theorems 3.1 and 3.4]. In view of the mentioned above result by Guoyin Li et al, the conclusion of Theorem 6.1 is also equivalent to the Ekeland variational principle. We also establish in Section 6 a series of other consequences of Theorem 6.1, covering primal and dual space conditions involved in (hopefully) all known formulations of extremality/stationarity and generalized separability properties.

The structure of the paper is as follows. The next Section 2 contains some preliminary definitions and facts used throughout the paper. Some of the facts are new. In Subsection 2.1 dedicated to the Ekeland variational principle, two new versions of this classical result are formulated: the *Geometric Ekeland Variational Principle (GEVP)* and the *Asymmetric*

*Geometric Ekeland Variational Principle (AGEVP)*, and the equivalence of these two statements to the conventional Ekeland variational principle is proved in Proposition 2.1. In Subsection 2.2, we recall the definitions of the Fréchet and Clarke normal cones and subdifferentials, several versions of subdifferential sum rules and a few other calculus facts needed in the sequel. Subsection 2.3 is dedicated to the extremality and stationarity properties, and the (extended) extremal principle, and collects in one place all the relevant definitions, primal space metric and dual space normal characterizations as well as some discussions and historical comments. In Subsection 2.4, we prove several assertions containing elementary arguments which are used in proving the equivalence of various generalized separation statements and, more importantly, provide quantitative estimates for parameters involved in such statements. Such arguments are usually hidden within proofs. We make them explicit here for use in the subsequent sections and also outside this paper.

Section 3 recalls several known modifications of the extremality and stationarity properties and their dual characterizations, which are going to be important for our study of the  $\varepsilon$ -versions of these properties in the subsequent sections, and introduces the new *approximate  $\alpha$ -stationarity* property. In Section 4, we discuss several ways of defining ‘distances’ between  $n$  sets, i.e. quantifying the closeness of a finite number of sets, i.e. how ‘far apart’ they are, or, at least, whether they have a common point. Such distances play an important role when studying extremality, stationarity and regularity properties of collections of sets. An example of a distance of this kind was considered in [31] under the name *nonintersect index*. Primal space characterizations of ‘ $\varepsilon$ -closest’ points of a finite collection of sets are established in the form common for conventional formulations of approximate stationarity, thus, building a bridge between the two seemingly different languages used in defining ‘extremal’ properties of collections of sets.

Sections 5 and 6 are dedicated to the comparison of, respectively, primal and dual space ‘fixed  $\varepsilon$ ’ conditions involved in definitions and characterizations of extremality, stationarity and generalized separation properties. In Section 5, we extend the Asymmetric Geometric Ekeland Variational Principle (AGEVP) from Section 2 to the case of  $n \geq 2$  sets and establish a metric counterpart of the ‘fixed  $\varepsilon$ ’ conditions involved in the definition of local extremality, thus, producing the core component of the equivalent metric characterizations of the approximate  $\alpha$ -stationarity, and consequently of the approximate stationarity and transversality. Section 6 presents a series of ‘generalized separation’ statements, providing dual characterizations of typical ‘extremal’ arrangements of collections of sets, discussed in the preceding sections, and examines the relationships between them. These statements refine core arguments, which can be found in various existing versions of the (extended) extremal principle, as well as some new extensions. Their conclusions combine assertions in terms of the Fréchet normal cones in Asplund spaces and in terms of the Clarke normal cones in general Banach spaces. All the separation statements in this section as well as the conventional extremal principle and its extensions in [22] and [31] are consequences of the general Theorem 6.1. Its proof uses the extension of the AGEVP from Section 5 as a replacement for the conventional Ekeland variational principle in generalized separation statements. The dual characterizations of the ‘extremal’ arrangements of collections of sets given in the statements in the first part of Section 6 are partially reversed in several statements at the end of the section in the setting of a general normed vector space. This leads, in particular, to a full dual characterization of the approximate  $\alpha$ -stationarity, and as a consequence of the approximate stationarity and transversality.

## 2 Preliminaries

Our basic notation is standard; see e.g., [3, 21, 33–36]. Throughout the paper,  $X$  is either a metric or (more often) a normed vector space. In the latter case, we often require it to be Banach or Asplund. The distance and the norm are denoted by  $d(\cdot, \cdot)$  and  $\|\cdot\|$ , respectively. We use the same symbols to denote distances and norms in all spaces (primal and dual). When considering products of spaces, we usually assume them equipped with the maximum distance or norm.  $B_\delta(x)$  and  $\overline{B}_\delta(x)$  denote, respectively, the open and closed balls with center  $x$  and radius  $\delta > 0$ . Given a point  $x$  and a set  $A$  in  $X$ ,  $d(x, A) := \inf_{a \in A} d(x, a)$  denotes the distance from  $x$  to  $A$ ; in particular  $d(x, \emptyset) := +\infty$ . Given two subsets  $A, B \subset X$ ,  $d(A, B) := \inf_{a \in A} d(a, B)$  denotes the distance between  $A$  and  $B$ . Given a set  $A$ , a point  $a \in A$  and a number  $\delta > 0$ , we call the set  $A \cap B_\delta(a)$  a *localization* of the set  $A$  near  $a$ . If  $X$  is a normed vector space, its topological dual is denoted by  $X^*$ , while  $\langle \cdot, \cdot \rangle$  denotes the bilinear form defining the pairing between the two spaces. The open unit balls in  $X$  and  $X^*$  are denoted by  $\mathbb{B}$  and  $\mathbb{B}^*$ , respectively.  $\mathbb{N}$  stands for the set of all positive integers. We also use the notation  $\mathbb{R}_\infty := \mathbb{R} \cup \{+\infty\}$ . Given a function  $f : X \rightarrow \mathbb{R}_\infty$ , its domain is the set  $\text{dom } f := \{x \in X : f(x) < \infty\}$ .

## 2.1 Ekeland Variational Principle

Not surprisingly, the following classical result due to Ekeland [37] (see also [3, 21, 34–36]) plays the key role in the subsequent studies.

**Ekeland Variational Principle (EVP)** *Suppose  $X$  is a complete metric space,  $f : X \rightarrow \mathbb{R}_\infty$  is lower semicontinuous,  $\bar{x} \in X$  and  $\varepsilon > 0$ . If*

$$f(\bar{x}) < \inf_X f + \varepsilon,$$

*then, for any  $\lambda > 0$ , there exists an  $\hat{x} \in X$  such that*

- (i)  $d(\hat{x}, \bar{x}) < \lambda$ ;
- (ii)  $f(\hat{x}) \leq f(\bar{x})$ ;
- (iii)  $f(x) + (\varepsilon/\lambda)d(x, \hat{x}) > f(\hat{x})$  for all  $x \in X \setminus \{\bar{x}\}$ .

In the current paper dedicated to collections of sets, we are going to use the next two geometric versions of the Ekeland variational principle. They characterise the mutual arrangement of a pair of sets in a complete metric space with respect to a pair of points, being almost (up to  $\varepsilon$ ) closest points of these sets and, similarly to the conventional EVP, establish the existence of another pair of points arbitrarily close (up to an additional parameter or a pair of parameters) to the given one and minimizing a certain perturbed function. The perturbed functions in both assertions involve the distance between localizations of the sets near these points, i.e. intersections of the sets with neighbourhoods of the points.

**Geometric Ekeland Variational Principle (GEVP)** *Suppose  $X$  is a complete metric space,  $A$  and  $B$  are closed subsets of  $X$ ,  $a \in A$ ,  $b \in B$ , and  $\varepsilon > 0$ . If*

$$d(a, b) < d(A, B) + \varepsilon, \quad (1)$$

*then, for any  $\lambda > 0$ , there exist  $\hat{a} \in A \cap B_\lambda(a)$  and  $\hat{b} \in B \cap B_\lambda(b)$  such that*

- (i)  $d(\hat{a}, \hat{b}) \leq d(a, b)$ ;
- (ii)  $d(A \cap B_\xi(\hat{a}), B \cap B_\xi(\hat{b})) + \frac{\xi\varepsilon}{\lambda} > d(\hat{a}, \hat{b})$  for all  $\xi > 0$ .

**Asymmetric Geometric Ekeland Variational Principle (AGEVP)** *Suppose  $X$  is a complete metric space,  $A$  and  $B$  are closed subsets of  $X$ ,  $a \in A$ ,  $b \in B$ , and  $\varepsilon > 0$ . If condition (1) is satisfied, then, for any  $\lambda, \rho > 0$ , there exist  $\hat{a} \in A \cap B_\lambda(a)$  and  $\hat{b} \in B \cap B_\rho(b)$  such that*

- (i)  $d(\hat{a}, \hat{b}) \leq d(a, b)$ ;
- (ii)  $d(A \cap B_{\xi\lambda}(\hat{a}), B \cap B_{\xi\rho}(\hat{b})) + \xi\varepsilon > d(\hat{a}, \hat{b})$  for all  $\xi > 0$ .

Unlike GEVP, its asymmetric version AGEVP allows for balls with different radii to be used in the localizations of the sets. This feature is going to play an important role in our subsequent analysis. It is easy to see that GEVP is a particular case of AGEVP with  $\rho = \lambda$ . We now show that the two geometric versions of the Ekeland variational principle formulated above are both equivalent to the conventional one.

**Proposition 2.1** *EVP  $\Leftrightarrow$  GEVP  $\Leftrightarrow$  AGEVP.*

**Proof.** We first show that EVP  $\Rightarrow$  AGEVP. Let the assumptions of AGEVP be satisfied. Given numbers  $\lambda, \rho > 0$ , we consider the space  $X \times X$  with a metric defined as follows:

$$d_{\lambda, \rho}((x, y), (u, v)) := \max \left\{ \frac{1}{\lambda} d(x, u), \frac{1}{\rho} d(y, v) \right\} \quad (x, y, u, v \in X). \quad (2)$$

Since  $X$  is complete,  $(A \times B, d_{\lambda, \rho})$  is a complete space. Choose an  $\varepsilon' \in ]0, \varepsilon[$  such that (1) is satisfied with  $\varepsilon'$  in place of  $\varepsilon$ . EVP applied to the function  $d$  on  $(A \times B, d_{\lambda, \rho})$  gives the existence of points  $\hat{a} \in A$  and  $\hat{b} \in B$  such that

$$d_{\lambda, \rho}((\hat{a}, \hat{b}), (a, b)) < 1, \quad d(\hat{a}, \hat{b}) \leq d(a, b), \quad (3)$$

$$d(x, y) - d(\hat{a}, \hat{b}) + \varepsilon' d_{\lambda, \rho}((x, y), (\hat{a}, \hat{b})) \geq 0 \quad \text{for all } (x, y) \in A \times B. \quad (4)$$

In view of the definition (2), the first inequality in (3) is equivalent to the following two:  $d(\hat{a}, a) < \lambda$  and  $d(\hat{b}, b) < \rho$ . Given any  $\xi > 0$ ,  $x \in A \cap B_{\xi\lambda}(\hat{a})$  and  $y \in B \cap B_{\xi\rho}(\hat{b})$ , by (2) and (4), we have, respectively,  $d_{\lambda, \rho}((x, y), (\hat{a}, \hat{b})) < \xi$  and

$$d(x, y) \geq d(\hat{a}, \hat{b}) - \varepsilon' d_{\lambda, \rho}((x, y), (\hat{a}, \hat{b})) > d(\hat{a}, \hat{b}) - \xi\varepsilon'.$$

It follows that  $d(A \cap B_{\xi\lambda}(\hat{a}), B \cap B_{\xi\rho}(\hat{b})) \geq d(\hat{a}, \hat{b}) - \xi\varepsilon' > d(\hat{a}, \hat{b}) - \xi\varepsilon$ . This proves AGEVP.

The implication AGEVP  $\Rightarrow$  GEVP is straightforward.

To complete the proof, we next show that GEVP  $\Rightarrow$  EVP. Let the assumptions of EVP be satisfied. Choose a positive number  $\varepsilon' < \varepsilon$  such that  $f(\bar{x}) < \inf_X f + \varepsilon'$  and another number  $\alpha$  such that  $0 < \alpha < \lambda(\varepsilon'^{-1} - \varepsilon^{-1})$ . We are going to consider the space  $X \times \mathbb{R}$  with the metric  $d := d_X + \alpha|\cdot|$ , which makes  $X \times \mathbb{R}$  a complete metric space, two closed subsets of  $X \times \mathbb{R}$ :  $A := \{(x, y) : x \in X, y \geq f(x)\}$  and  $B := X \times \{M\}$ , where  $M := \inf_X f$ , and two points  $a := (\bar{x}, f(\bar{x})) \in A$  and  $b := (\bar{x}, M) \in B$ . We have  $d(A, B) = 0$  and

$$d(a, b) = \alpha(f(\bar{x}) - M) < \alpha\varepsilon' = d(A, B) + \alpha\varepsilon'.$$

GEVP gives the existence of points  $\hat{a} = (\hat{x}, \hat{y}) \in A \cap B_\lambda(a)$  and  $\hat{b} = (\hat{x}', M) \in B \cap B_\lambda(b)$  satisfying

$$d(\hat{a}, \hat{b}) \leq d(a, b), \quad (5)$$

$$d(A \cap B_\xi(\hat{a}), B \cap B_\xi(\hat{b})) + \frac{\xi\alpha\varepsilon'}{\lambda} > d(\hat{a}, \hat{b}) \quad \text{for all } \xi > 0. \quad (6)$$

Observe that  $d(\hat{x}, \bar{x}) < \lambda$ , and consequently,  $(\hat{x}, M) \in B \cap B_\lambda(b)$ . Moreover,  $\hat{x}' = \hat{x}$ . Indeed, if  $\hat{x} \neq \hat{x}'$ , then we can take  $\xi := d(\hat{x}, \hat{x}')$ . Then  $(\hat{x}, \hat{y}) \in A \cap B_\xi(\hat{a})$ ,  $(\hat{x}, M) \in B \cap B_\xi(\hat{b})$  and

$$d((\hat{x}, \hat{y}), (\hat{x}, M)) + \frac{\xi\alpha\varepsilon'}{\lambda} < \alpha(\hat{y} - M) + \left(1 - \frac{\varepsilon'}{\varepsilon}\right) d(\hat{x}, \hat{x}') < \alpha(\hat{y} - M) + d(\hat{x}, \hat{x}') = d(\hat{a}, \hat{b}),$$

which contradicts (6); hence  $\hat{x}' = \hat{x}$  and  $\hat{b} = (\hat{x}, M)$ . Consequently, condition (5) reduces to  $\alpha(\hat{y} - M) \leq \alpha(f(\bar{x}) - M)$ , which implies that  $f(\hat{x}) \leq \hat{y} \leq f(\bar{x})$ .

It remains to prove condition (iii) in EVP. Let  $x \neq \hat{x}$ . If  $f(x) \geq f(\hat{x})$ , the condition holds trivially. Let  $f(x) < f(\hat{x})$ , and take  $\xi := d(x, \hat{x}) + \alpha(\hat{y} - f(x))$ . Then  $(x, f(x)) \in A \cap B_\xi(\hat{a})$ ,  $(x, M) \in B \cap B_\xi(\hat{b})$ , and by (6),

$$\alpha(f(x) - M) + \frac{\alpha\varepsilon'}{\lambda} (d(x, \hat{x}) + \alpha(\hat{y} - f(x))) = d((x, f(x)), (x, M)) + \frac{\xi\alpha\varepsilon'}{\lambda} \geq d(\hat{a}, \hat{b}) = \alpha(\hat{y} - M).$$

Thus,

$$f(x) + \frac{\varepsilon'}{\lambda} (d(x, \hat{x}) + \alpha(\hat{y} - f(x))) \geq \hat{y},$$

or equivalently,

$$f(x) + \frac{\varepsilon'}{\lambda - \alpha\varepsilon'} d(x, \hat{x}) \geq \hat{y}.$$

By the definition of  $\alpha$ , we have  $\lambda - \alpha\varepsilon' > \lambda\varepsilon'\varepsilon^{-1}$ . Hence,

$$f(x) + \frac{\varepsilon}{\lambda} d(x, \hat{x}) > f(x) + \frac{\varepsilon'}{\lambda - \alpha\varepsilon'} d(x, \hat{x}) \geq \hat{y} \geq f(\hat{x}).$$

The proof is complete.  $\square$

Condition (ii) in GEVP and AGEVP corresponds to condition (iii) in the conventional EVP, while condition (i) corresponds to the pair of conditions (i) and (ii). The  $\varepsilon$ -closeness condition (1) is going to play an important role in our analysis. It will be further discussed in Sections 4 and 5.

## 2.2 Normal Cones and Subdifferentials

In Section 6, we use dual tools – normal cones and subdifferentials, usually in the Fréchet or Clarke sense. Given a subset  $A$  of a normed vector space  $X$  and a point  $\bar{x} \in A$ , the set (cf. [23])

$$N_A^F(\bar{x}) := \left\{ x^* \in X^* : \limsup_{x \rightarrow \bar{x}, x \in A \setminus \{\bar{x}\}} \frac{\langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq 0 \right\} \quad (7)$$

is the *Fréchet normal cone* to  $A$  at  $\bar{x}$ . It is a nonempty closed convex cone, often trivial (i.e.  $N_A^F(\bar{x}) = \{0\}$ ). The *Clarke normal cone* to  $A$  at  $\bar{x}$  is defined as the set (cf. [38])

$$N_A^C(\bar{x}) := \left\{ x^* \in X^* : \langle x^*, z \rangle \leq 0 \quad \text{for all } z \in T_A^C(\bar{x}) \right\}, \quad (8)$$

where  $T_A^C(\bar{x})$  is the *Clarke tangent cone* to  $A$  at  $\bar{x}$ :

$$T_A^C(\bar{x}) := \left\{ z \in X : \forall x_k \xrightarrow{A} \bar{x}, \forall t_k \downarrow 0, \exists z_k \rightarrow z \text{ such that } x_k + t_k z_k \in A \text{ for all } k \in \mathbb{N} \right\}. \quad (9)$$

The set (8) is a nonempty weak\* closed convex cone, and  $N_A^F(\bar{x}) \subset N_A^C(\bar{x})$ . If  $A$  is a convex set, then (7) and (8) reduce to the normal cone in the sense of convex analysis (cf. e.g. [23, Proposition 1.19], [38, Proposition 2.4.4]):

$$N_A(\bar{x}) := \{x^* \in X^* : \langle x^*, x - \bar{x} \rangle \leq 0 \text{ for all } x \in A\}.$$

We will often use the generic notation  $N$  for both Fréchet and Clarke normal cones, specifying wherever necessary that either  $N := N^F$  or  $N := N^C$ .

Given a function  $f : X \rightarrow \mathbb{R}_\infty$  and a point  $\bar{x} \in \text{dom } f$ , the *Fréchet* and *Clarke subdifferentials* of  $f$  at  $\bar{x}$  can be defined via the respective normal cones to the epigraph  $\text{epi } f := \{(x, \alpha) \in X \times \mathbb{R} : f(x) \leq \alpha\}$  of  $f$  as follows:

$$\begin{aligned} \partial^F f(\bar{x}) &:= \left\{ x^* \in X^* : (x^*, -1) \in N_{\text{epi } f}^F(\bar{x}, f(\bar{x})) \right\}, \\ \partial^C f(\bar{x}) &:= \left\{ x^* \in X^* : (x^*, -1) \in N_{\text{epi } f}^C(\bar{x}, f(\bar{x})) \right\}. \end{aligned}$$

These are closed convex sets, the first one often empty, and  $\partial^F f(\bar{x}) \subset \partial^C f(\bar{x})$ . If  $f$  is convex, they both reduce to the subdifferential in the sense of convex analysis (cf. e.g. [23, Proposition 1.2], [38, Proposition 2.2.7]):

$$\partial f(\bar{x}) := \{x^* \in X^* : f(x) - f(\bar{x}) - \langle x^*, x - \bar{x} \rangle \geq 0 \text{ for all } x \in X\}.$$

By convention, we set  $N_A^F(\bar{x}) = N_A^C(\bar{x}) := \emptyset$  if  $\bar{x} \notin A$  and  $\partial^F f(\bar{x}) = \partial^C f(\bar{x}) := \emptyset$  if  $\bar{x} \notin \text{dom } f$ . It is easy to check that  $N_A^F(\bar{x}) = \partial^F i_A(\bar{x})$  and  $N_A^C(\bar{x}) = \partial^C i_A(\bar{x})$ , where  $i_A$  is the *indicator function* of  $A$ :  $i_A(x) = 0$  if  $x \in A$  and  $i_A(x) = \infty$  if  $x \notin A$ ; cf. e.g. [23, Proposition 1.18], [38, Proposition 2.4.12].

Several kinds of *subdifferential sum rules* are used in Section 6 when deducing dual space results. They are collected in the next lemma.

**Lemma 2.1 (Subdifferential sum rules)** *Suppose  $X$  is a normed vector space,  $f_1, f_2 : X \rightarrow \mathbb{R}_\infty$ , and  $\bar{x} \in \text{dom } f_1 \cap \text{dom } f_2$ .*

(i) **Convex sum rule.** *Suppose  $f_1$  and  $f_2$  are convex and  $f_1$  is continuous at a point in  $\text{dom } f_2$ . Then*

$$\partial(f_1 + f_2)(\bar{x}) = \partial f_1(\bar{x}) + \partial f_2(\bar{x}).$$

(ii) **Differentiable sum rule.** *Suppose  $f_1$  is Fréchet differentiable at  $\bar{x}$ . Then,*

$$\partial^F(f_1 + f_2)(\bar{x}) = \nabla f_1(\bar{x}) + \partial^F f_2(\bar{x}).$$

(iii) **Fuzzy sum rule.** *Suppose  $X$  is Asplund,  $f_1$  is Lipschitz continuous and  $f_2$  is lower semicontinuous in a neighbourhood of  $\bar{x}$ . Then, for any  $\varepsilon > 0$ , there exist  $x_1, x_2 \in X$  with  $\|x_i - \bar{x}\| < \varepsilon$ ,  $|f_i(x_i) - f_i(\bar{x})| < \varepsilon$  ( $i = 1, 2$ ), such that*

$$\partial^F(f_1 + f_2)(\bar{x}) \subset \partial^F f_1(x_1) + \partial^F f_2(x_2) + \varepsilon \mathbb{B}^*.$$

(iv) **Clarke–Rockafellar sum rule.** *Suppose  $f_1$  is Lipschitz continuous and  $f_2$  is lower semicontinuous in a neighbourhood of  $\bar{x}$ . Then*

$$\partial^C(f_1 + f_2)(\bar{x}) \subset \partial^C f_1(\bar{x}) + \partial^C f_2(\bar{x}).$$

The first sum rule in the lemma above is the conventional subdifferential sum rule of convex analysis; see e.g. [39, Theorem 0.3.3] and [2, Theorem 2.8.7]. Together with the second one, these are examples of *exact* sum rules. The third sum rule is known as the *fuzzy* or *approximate* sum rule (Fabian [17]) for Fréchet subdifferentials in Asplund spaces; cf., e.g., [23, Rule 2.2] and [21, Theorem 2.33]. Note that, unlike the sum rules in parts (i) and (ii) of the lemma, the subdifferentials in the right-hand side of the inclusion are computed not at the reference point, but at some points nearby. This explains the name. The fourth sum rule is formulated in terms of Clarke subdifferentials. It was established in Rockafellar [40, Theorem 2]. Similar to the previous one, it is valid generally only as inclusion. Nevertheless, it is another example of exact sum rule.

Recall that a Banach space is *Asplund* if every continuous convex function on an open convex set is Fréchet differentiable on a dense subset [41], or equivalently, if the dual of each its separable subspace is separable. We refer the reader to [21, 41, 42] for discussions about and characterizations of Asplund spaces. All reflexive, particularly, all finite dimensional Banach spaces are Asplund.

The following facts are immediate consequences of the definition of the Fréchet subdifferential and normal cone (cf. e.g. [23, Propositions 1.10 and 1.29]).

**Lemma 2.2** Suppose  $X$  is a normed vector space and  $f : X \rightarrow \mathbb{R}_\infty$ . If  $\bar{x} \in \text{dom } f$  is a point of local minimum of  $f$ , then  $0 \in \partial^F f(\bar{x})$ .

**Lemma 2.3** Suppose  $X_1$  and  $X_2$  are normed vector spaces,  $\bar{x}_1 \in A_1 \subset X_1$  and  $\bar{x}_2 \in A_2 \subset X_2$ . Then

$$N_{A_1 \times A_2}^F(\bar{x}_1, \bar{x}_2) = N_{A_1}^F(\bar{x}_1) \times N_{A_2}^F(\bar{x}_2).$$

### 2.3 Extremality, Stationarity and (Extended) Extremal Principle

In this subsection, we recall and discuss the conventional definitions of extremality, local extremality, stationarity and approximate stationarity of collections of sets, and the (extended) extremal principle. Here and in the rest of the paper, we consider  $n$  sets  $\Omega_1, \dots, \Omega_n$  ( $2 \leq n < \infty$ ) and write  $\{\Omega_1, \dots, \Omega_n\}$  to denote the collection of the sets as a single object.

The next definition collects several extremality and stationarity properties of collections of sets.

**Definition 2.1** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \cap_{i=1}^n \Omega_i$ . The collection  $\{\Omega_1, \dots, \Omega_n\}$  is

(i) **extremal** iff for any  $\varepsilon > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying

$$\bigcap_{i=1}^n (\Omega_i - a_i) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon; \quad (\text{P1})$$

(ii) **locally extremal** at  $\bar{x}$  iff there exists a number  $\rho \in ]0, \infty[$  such that, for any  $\varepsilon > 0$ , there are vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying

$$\bigcap_{i=1}^n (\Omega_i - a_i) \cap B_\rho(\bar{x}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon; \quad (\text{P2})$$

(iii) **stationary** at  $\bar{x}$  iff for any  $\varepsilon > 0$ , there exist a number  $\rho \in ]0, \varepsilon[$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying

$$\bigcap_{i=1}^n (\Omega_i - a_i) \cap B_\rho(\bar{x}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon \rho;$$

(iv) **approximately stationary** at  $\bar{x}$  iff for any  $\varepsilon > 0$ , there exist a number  $\rho \in ]0, \varepsilon[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying

$$\bigcap_{i=1}^n (\Omega_i - \omega_i - a_i) \cap (\rho \mathbb{B}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon \rho. \quad (\text{P3})$$

The formulas in Definition 2.1 and several their modifications discussed in Section 3 are central for our analysis and are going to be extensively referred to throughout the paper. We will use special P tags: (P1), (P2), ... for such formulas as well as some other formulas in Section 3 involved in primal space metric characterizations of extremality and stationarity. D tags: (D1), (D2), ... are reserved for the formulas involved in the corresponding dual space characterizations of extremality and stationarity.

Condition (i) (condition (ii)) in Definition 2.1 means that an appropriate arbitrarily small shift of the sets makes them nonintersecting (in a neighbourhood of  $\bar{x}$ ). This is a very general model embracing many optimality notions. The other two conditions correspond to more subtle properties of optimization problems, closer to stationarity.

The properties in parts (i) and (ii) of Definition 2.1 were introduced in [13] and [14], respectively; see also [21, 23]. The properties in parts (iii) and (iv) first appeared in [24] and [43], respectively; see also [32, 44]. Property (iv) was referred to in [43] as *extremality near  $\bar{x}$* . The name *approximate stationarity* was suggested in [25].

Unlike condition (ii), in conditions (iii) and (iv) the magnitudes of the ‘‘shifts’’ of the sets are related to that of the neighbourhood in which the sets become nonintersecting, namely  $\max_{1 \leq i \leq n} \|a_i\| / \rho < \varepsilon$ . Compared to (iii), in condition (iv), instead of the common point  $\bar{x}$ , each set  $\Omega_i$  is considered near its own point  $\omega_i$ .

The relationships between the properties in Definition 2.1 are straightforward. The equivalences in part (ii) of the proposition below were proved in [45, Proposition 14].

**Proposition 2.2** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \cap_{i=1}^n \Omega_i$ .

- (i) For the properties in Definition 2.1, the following implications hold true: (i)  $\Rightarrow$  (ii)  $\Rightarrow$  (iii)  $\Rightarrow$  (iv).  
(ii) If the sets are convex, then the implications in the previous item hold as equivalences: (i)  $\Leftrightarrow$  (ii)  $\Leftrightarrow$  (iii)  $\Leftrightarrow$  (iv).  
(iii) If the collection  $\{\Omega_1, \dots, \Omega_n\}$  is locally extremal at  $\bar{x}$  with  $\rho = \infty$ , then it is extremal.  
(iv) If the collection  $\{\Omega_1, \dots, \Omega_n\}$  is locally extremal at  $\bar{x}$  with some  $\rho \in ]0, \infty]$ , then the collection of  $n+1$  sets  $\Omega_1, \dots, \Omega_n, B_\rho(\bar{x})$  is extremal.

It is easy to check that all the implications in Proposition 2.2(i) can be strict; see examples in [25, 32, 45].

Thanks to Proposition 2.2(i), approximate stationarity is the weakest of the four properties in Definition 2.1. It happens to be an important type of mutual arrangement of a collection of sets in space. The next two statements provide several, respectively, primal space metric and dual space (in terms of Fréchet normals) necessary and sufficient criteria for this property. The characterizations formulated in these statements are obviously necessary for each of the other three properties in Definition 2.1.

**Theorem 2.1 (Approximate stationarity: metric criteria)** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . The following conditions are equivalent:

- (i) the collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately stationary at  $\bar{x}$ ;  
(ii) for any  $\varepsilon > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) such that

$$\varepsilon d\left(\bar{x}, \bigcap_{i=1}^n (\Omega_i - a_i)\right) > \max_{1 \leq i \leq n} d(\bar{x}, \Omega_i - a_i) \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon;$$

- (iii) for any  $\varepsilon > 0$ , there exist a point  $x \in B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) such that

$$\varepsilon d\left(x, \bigcap_{i=1}^n (\Omega_i - a_i)\right) > \max_{1 \leq i \leq n} d(x, \Omega_i - a_i) \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon.$$

The equivalence of the conditions (i) and (iii) in Theorem 2.1 is a consequence of [45, Theorem 1]; cf. [44, Theorem 1]. The equivalence of the all three conditions follows from a more general statement in Corollary 5.4 in Section 5. Each of the conditions (ii) and (iii) can be used as an equivalent definition of approximate stationarity.

The next theorem can be considered as a generalization of the classical convex *separation theorem* to collections of nonconvex sets.

**Theorem 2.2 (Extended extremal principle)** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of an Asplund space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . With  $N$  standing for the Fréchet normal cone ( $N := N^F$ ), the following conditions are equivalent:

- (i) the collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately stationary at  $\bar{x}$ ;  
(ii) for any  $\varepsilon > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that

$$\left\| \sum_{i=1}^n x_i^* \right\| < \varepsilon, \quad x_i^* \in N_{\Omega_i}(\omega_i) \quad (i = 1, \dots, n) \quad \text{and} \quad \sum_{i=1}^n \|x_i^*\| = 1; \quad (\text{D1})$$

- (iii) for any  $\varepsilon > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that

$$\sum_{i=1}^n x_i^* = 0, \quad \sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) < \varepsilon \quad \text{and} \quad \sum_{i=1}^n \|x_i^*\| = 1. \quad (\text{D2})$$

Conditions (ii) and (iii) in Theorem 2.2 have been used interchangeably (together with several their modifications) since 1980 in the concluding part of the *extremal principle* as dual necessary conditions for extremality [13] or local extremality [12, 14, 15], first in Fréchet smooth spaces (i.e. Banach spaces admitting an equivalent norm Fréchet differentiable away from zero), and since 1996 [18] in Asplund spaces. The proof of the extremal principle is based on the applications of the two fundamental variational analysis tools: the Ekeland Variational Principle (EVP) and the fuzzy sum rule for Fréchet subdifferentials (Lemma 2.1(iii)). We refer the readers to [21, Section 2.6] and [32] for more historical comments.

The fact that the generalized separation actually characterizes a weaker than (local) extremality property of approximate stationarity, and this characterization is necessary and sufficient was first established (in a slightly different form) in [43] in the setting of a Fréchet smooth Banach space and extended to Asplund spaces in [46]. The full proof of the extended extremal principle appeared in [47], while the name *Extended extremal principle* was introduced in [23]. It is worth noting that the proof of the necessity (of either condition (ii) or condition (iii)) in the extended extremal principle follows that of the conventional extremal principle and only refines some estimates, while the sufficiency is almost straightforward and is valid in arbitrary normed vector spaces.

**Remark 2.1** (i) Conditions (ii) and (iii) in Theorem 2.2 represent two kinds of widely used *generalized (approximate) separation* of a collection of sets: they claim the existence of  $n$  vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ), satisfying the normalization condition  $\sum_{i=1}^n \|x_i^*\| = 1$ , and being either normal to the respective sets at some points close (up to  $\varepsilon$ ) to  $\bar{x}$  with their sum being almost (up to  $\varepsilon$ ) 0 (condition (ii)), or almost (up to  $\varepsilon$ ) normal with their sum equal exactly 0 (condition (iii)).

The equivalence of the conditions (ii) and (iii) in Theorem 2.2 is not difficult to check directly (even in the setting of an arbitrary normed vector space and without assuming the closedness of the sets) using elementary arguments which involve small perturbations of the vectors and then scaling the perturbed vectors to ensure the normalization condition. Such arguments have been used in many proofs and are scattered across a number of publications. They were made explicit (in the case of two sets) in [32, Lemma 1]. In the next subsection, we formulate a more general statement, which will be used also in Section 6.

- (ii) The inequality in (D2) is only meaningful when  $\varepsilon \leq 1$ , because otherwise it is a direct consequence of the equalities. If condition (D2) holds with  $\varepsilon = 1$ , it also holds with some  $\varepsilon < 1$ . Thanks to these observations, when applying Theorem 2.2 or its extensions, one can always assume that  $\varepsilon < 1$ .
- (iii) Theorem 2.2 (as well as the conventional extremal principle) with conditions (ii) and (iii) in their current form cannot be extended beyond Asplund spaces. However, replacing in these conditions Fréchet normal cones with normal cones corresponding to other subdifferentials possessing reasonable (approximate or exact) sum rules in the respective *trustworthy* [19] spaces, one can employ basically the same routine to show the necessity of the amended conditions in these spaces; see e.g. [19–22] and the discussion in the Introduction. Thanks to Lemma 2.1(iv) and (i), in general Banach spaces one can use Clarke normal cones or even conventional normal cones in the sense of convex analysis if the sets are convex. Note that the sufficiency of the conditions (ii) and (iii) in Theorem 2.2 is only valid for the Fréchet normal cones, though in general normed vector spaces. Thus, with such extensions, we only have (i)  $\Rightarrow$  (ii)  $\Leftrightarrow$  (iii), unless the sets are convex.

## 2.4 Perturbations and Scaling of Vectors

Below we present several assertions containing elementary arguments which are used in proving the equivalence of the conditions (ii) and (iii) in Theorem 2.2 and similar facts. We start with a general statement, which will be used also in Section 6.

**Lemma 2.4** *Suppose  $K_1, \dots, K_n$  are cones in a normed vector space,  $\varepsilon > 0$ ,  $\rho > 0$  and  $\lambda > 0$ . Suppose also that vectors  $z_1, \dots, z_n$  satisfy*

$$\lambda \sum_{i=1}^n d(z_i, K_i) + \rho \left\| \sum_{i=1}^n z_i \right\| < \varepsilon, \quad \sum_{i=1}^n \|z_i\| = 1. \quad (10)$$

- (i) *If  $\varepsilon + \lambda \leq \rho$ , then there exist vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) satisfying the following conditions:*

$$\sum_{i=1}^n \hat{z}_i = 0, \quad \sum_{i=1}^n d(\hat{z}_i, K_i) < \frac{\varepsilon}{\lambda}, \quad \sum_{i=1}^n \|\hat{z}_i\| = 1. \quad (11)$$

- (ii) *If  $\varepsilon + \rho \leq \lambda$  then there exist vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) satisfying the following conditions:*

$$\left\| \sum_{i=1}^n \hat{z}_i \right\| < \frac{\varepsilon}{\rho}, \quad \hat{z}_i \in K_i \quad (i = 1, \dots, n), \quad \sum_{i=1}^n \|\hat{z}_i\| = 1. \quad (12)$$

- (iii) *Moreover, if the underlying space is dual to a normed vector space  $X$ , and*

$$\sum_{i=1}^n \langle z_i, x_i \rangle \geq \tau \max_{1 \leq i \leq n} \|x_i\| \quad (13)$$

*for some vectors  $x_i \in X$  ( $i = 1, \dots, n$ ), not all zero, and a number  $\tau \in ]0, 1]$ , then the vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) in parts (i) or (ii) satisfy*

$$\sum_{i=1}^n \langle \hat{z}_i, x_i \rangle > \hat{\tau} \max_{1 \leq i \leq n} \|x_i\|, \quad (14)$$

*where  $\hat{\tau} := \frac{\tau\rho - \varepsilon}{\rho + \varepsilon}$  under the assumptions in part (i), and  $\hat{\tau} := \frac{\tau\lambda - \varepsilon}{\lambda + \varepsilon}$  under the assumptions in part (ii).*

**Proof.** (i) Let  $\varepsilon + \lambda \leq \rho$ . Set  $z := \sum_{i=1}^n z_i$  and  $v_i := z_i - \frac{1}{n}z$  ( $i = 1, \dots, n$ ). Then  $\sum_{i=1}^n v_i = 0$  and, by (10),  $\|z\| < \frac{\varepsilon}{\rho} < 1$ ,

$$\sum_{i=1}^n \|v_i\| \leq \sum_{i=1}^n \|z_i\| + \|z\| < 1 + \frac{\varepsilon}{\rho} \quad \text{and} \quad \sum_{i=1}^n \|v_i\| \geq \sum_{i=1}^n \|z_i\| - \|z\| = 1 - \|z\| > 0.$$

Set  $\hat{z}_i := v_i / \sum_{i=1}^n \|v_i\|$ . Then  $\sum_{i=1}^n \hat{z}_i = 0$ ,  $\sum_{i=1}^n \|\hat{z}_i\| = 1$  and

$$\begin{aligned} \sum_{i=1}^n d(\hat{z}_i, K_i) &\leq \frac{\sum_{i=1}^n d(z_i, K_i) + \|z\|}{\sum_{i=1}^n \|v_i\|} = \frac{\lambda \sum_{i=1}^n d(z_i, K_i) + \rho \|z\| - (\rho - \lambda) \|z\|}{\lambda \sum_{i=1}^n \|v_i\|} \\ &< \frac{\varepsilon - (\rho - \lambda) \|z\|}{\lambda(1 - \|z\|)} \leq \frac{\varepsilon - \varepsilon \|z\|}{\lambda(1 - \|z\|)} = \frac{\varepsilon}{\lambda}, \end{aligned}$$

i.e., all conditions in (11) are satisfied.

(ii) Let  $\varepsilon + \rho \leq \lambda$ . By (10), there exist vectors  $v_i \in K_i$  ( $i = 1, \dots, n$ ) such that

$$\lambda \sum_{i=1}^n \|z_i - v_i\| + \rho \left\| \sum_{i=1}^n z_i \right\| < \varepsilon.$$

In particular,  $\sum_{i=1}^n \|z_i - v_i\| < \frac{\varepsilon}{\lambda} < 1$ . Hence,

$$\begin{aligned} \sum_{i=1}^n \|v_i\| &\leq \sum_{i=1}^n \|z_i\| + \sum_{i=1}^n \|z_i - v_i\| = 1 + \sum_{i=1}^n \|z_i - v_i\|, \\ \sum_{i=1}^n \|v_i\| &\geq \sum_{i=1}^n \|z_i\| - \sum_{i=1}^n \|z_i - v_i\| = 1 - \sum_{i=1}^n \|z_i - v_i\| > 0, \\ \left\| \sum_{i=1}^n v_i \right\| &\leq \sum_{i=1}^n \|z_i - v_i\| + \left\| \sum_{i=1}^n z_i \right\| < \frac{1}{\rho} \left( \varepsilon - (\lambda - \rho) \sum_{i=1}^n \|z_i - v_i\| \right) \leq \frac{\varepsilon}{\rho} \left( 1 - \sum_{i=1}^n \|z_i - v_i\| \right). \end{aligned}$$

Set  $\hat{z}_i := v_i / \sum_{i=1}^n \|v_i\|$ . Then  $\hat{z}_i \in K_i$  ( $i = 1, \dots, n$ ),  $\sum_{i=1}^n \|\hat{z}_i\| = 1$  and

$$\left\| \sum_{i=1}^n \hat{z}_i \right\| = \frac{\left\| \sum_{i=1}^n v_i \right\|}{\sum_{i=1}^n \|v_i\|} < \frac{\varepsilon(1 - \sum_{i=1}^n \|z_i - v_i\|)}{\rho(1 - \sum_{i=1}^n \|z_i - v_i\|)} = \frac{\varepsilon}{\rho},$$

i.e., all conditions in (12) are satisfied.

(iii) Suppose that the underlying space is dual to a normed vector space  $X$  and condition (13) is satisfied for some vectors  $x_i \in X$  ( $i = 1, \dots, n$ ), not all zero, and a number  $\tau \in ]0, 1]$ . Then, using the notations introduced above, we have:

$$\sum_{i=1}^n \langle \hat{z}_i, x_i \rangle \geq \frac{\sum_{i=1}^n \langle z_i, x_i \rangle - (\sum_{i=1}^n \|z_i - v_i\|) \max_{1 \leq i \leq n} \|x_i\|}{\sum_{i=1}^n \|v_i\|} \geq \frac{\tau - \sum_{i=1}^n \|z_i - v_i\|}{\sum_{i=1}^n \|v_i\|} \max_{1 \leq i \leq n} \|x_i\|.$$

Employing the estimates in part (i), we obtain

$$\frac{\tau - \sum_{i=1}^n \|z_i - v_i\|}{\sum_{i=1}^n \|v_i\|} = \frac{\tau - \|z\|}{\sum_{i=1}^n \|v_i\|} > \frac{\tau - \frac{\varepsilon}{\rho}}{1 + \frac{\varepsilon}{\rho}} = \frac{\tau\rho - \varepsilon}{\rho + \varepsilon},$$

while the estimates in part (ii) give

$$\frac{\tau - \sum_{i=1}^n \|z_i - v_i\|}{\sum_{i=1}^n \|v_i\|} \geq \frac{\tau - \sum_{i=1}^n \|z_i - v_i\|}{1 + \sum_{i=1}^n \|z_i - v_i\|} > \frac{\tau - \frac{\varepsilon}{\lambda}}{1 + \frac{\varepsilon}{\lambda}} = \frac{\tau\lambda - \varepsilon}{\rho + \lambda}.$$

Thus, in both cases we arrive at (14).  $\square$

The next two corollaries present two important special cases of Lemma 2.4.

**Corollary 2.1** Suppose  $K_1, \dots, K_n$  are cones in a normed vector space,  $\varepsilon \in ]0, 1[$  and vectors  $z_1, \dots, z_n$  satisfy

$$\left\| \sum_{i=1}^n z_i \right\| < \varepsilon, \quad z_i \in K_i \quad (i = 1, \dots, n), \quad \sum_{i=1}^n \|z_i\| = 1.$$

Then there exist vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) satisfying the following conditions:

$$\sum_{i=1}^n \hat{z}_i = 0, \quad \sum_{i=1}^n d(\hat{z}_i, K_i) < \frac{\varepsilon}{1 - \varepsilon}, \quad \sum_{i=1}^n \|\hat{z}_i\| = 1.$$

Moreover, if the underlying space is dual to a normed vector space  $X$ , and condition (13) is satisfied for some vectors  $x_i \in X$  ( $i = 1, \dots, n$ ), not all zero, and a number  $\tau \in ]0, 1]$ , then the vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) satisfy condition (14) with  $\hat{\tau} := \frac{\tau - \varepsilon}{1 + \varepsilon}$ .

**Proof.** Apply Lemma 2.4(i) with  $\rho = 1$  and  $\lambda = 1 - \varepsilon$ . □

**Corollary 2.2** Suppose  $K_1, \dots, K_n$  are cones in a normed vector space,  $\varepsilon \in ]0, 1[$  and vectors  $z_1, \dots, z_n$  satisfy

$$\sum_{i=1}^n z_i = 0, \quad \sum_{i=1}^n d(z_i, K_i) < \varepsilon, \quad \sum_{i=1}^n \|z_i\| = 1.$$

Then there exist vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) satisfying the following conditions:

$$\left\| \sum_{i=1}^n \hat{z}_i \right\| < \frac{\varepsilon}{1-\varepsilon}, \quad \hat{z}_i \in K_i \ (i = 1, \dots, n), \quad \sum_{i=1}^n \|\hat{z}_i\| = 1. \quad (15)$$

Moreover, if the underlying space is dual to a normed vector space  $X$ , and condition (13) is satisfied for some vectors  $x_i \in X$  ( $i = 1, \dots, n$ ), not all zero, and a number  $\tau \in ]0, 1]$ , then the vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) satisfy condition (14) with  $\hat{\tau} := \frac{\tau-\varepsilon}{1+\varepsilon}$ .

**Proof.** Apply Lemma 2.4(ii) with  $\rho = 1 - \varepsilon$  and  $\lambda = 1$ . □

As an immediate consequence of Corollaries 2.1 and 2.2, we obtain the following important assertion.

**Corollary 2.3** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . Conditions (ii) and (iii) in Theorem 2.2 are equivalent.

When  $n = 2$ , the main estimate in Corollary 2.1 can be improved.

**Proposition 2.3** Suppose  $K_1$  and  $K_2$  are cones in a normed vector space,  $\varepsilon > 0$ , and vectors  $z_1$  and  $z_2$  satisfy

$$\|z_1 + z_2\| < \varepsilon, \quad z_1 \in K_1, \ z_2 \in K_2, \quad \|z_1\| + \|z_2\| = 1.$$

Then, there exist vectors  $\hat{z}_1$  and  $\hat{z}_2$  satisfying the following conditions:

$$\hat{z}_1 + \hat{z}_2 = 0, \quad \hat{z}_1 \in K_1, \quad d(\hat{z}_2, K_2) < \varepsilon, \quad \|\hat{z}_1\| + \|\hat{z}_2\| = 1.$$

**Proof.** Without loss of generality, we can assume that  $\|z_2\| \leq \frac{1}{2} \leq \|z_1\|$ . Set  $\hat{z}_1 := \frac{z_1}{2\|z_1\|}$  and  $\hat{z}_2 := -\hat{z}_1$ . Then  $\hat{z}_1 + \hat{z}_2 = 0$ ,  $\|\hat{z}_1\| = \|\hat{z}_2\| = \frac{1}{2}$ ,  $\hat{z}_1 \in K_1$  and

$$d(\hat{z}_2, K_2) \leq \left\| \hat{z}_2 - \frac{z_2}{2\|z_1\|} \right\| = \frac{\|z_1 + z_2\|}{2\|z_1\|} < \frac{\varepsilon}{2\|z_1\|} \leq \varepsilon.$$

This completes the proof. □

Next we formulate an asymmetric modification of Corollary 2.2 which will be used in Section 6.

**Proposition 2.4** Suppose  $K_1, \dots, K_n$  are cones in a normed vector space,  $\varepsilon \in ]0, 1[$  and vectors  $z_1, \dots, z_n$  satisfy conditions

$$\sum_{i=1}^n z_i = 0, \quad \sum_{i=1}^n d(z_i, K_i) < \varepsilon, \quad \sum_{i=1}^{n-1} \|z_i\| = 1.$$

Then, there exist vectors  $\hat{z}_i$  ( $i = 1, \dots, n$ ) satisfying the following conditions:

$$\sum_{i=1}^n \hat{z}_i = 0, \quad \hat{z}_i \in K_i \ (i = 1, \dots, n-1), \quad d(\hat{z}_n, K_n) < \frac{\varepsilon}{1-\varepsilon} \quad \text{and} \quad \sum_{i=1}^{n-1} \|\hat{z}_i\| = 1. \quad (16)$$

**Proof.** Take  $y_i \in K_i$  ( $i = 1, \dots, n$ ) such that  $\sum_{i=1}^n \|z_i - y_i\| < \varepsilon$ . Then  $\|\sum_{i=1}^n y_i\| < \varepsilon$  and  $\sum_{i=1}^{n-1} \|y_i\| > 1 - \varepsilon$ . It follows that  $d(-\sum_{i=1}^{n-1} y_i, K_n) \leq d(-\sum_{i=1}^{n-1} y_i, y_n) < \varepsilon$ . Hence, vectors  $\hat{z}_i := y_i / \sum_{i=1}^{n-1} \|y_i\|$  ( $i = 1, \dots, n-1$ ) and  $\hat{z}_n := -\sum_{i=1}^{n-1} \hat{z}_i$  satisfy all the conditions in (16). □

### 3 Modifications of the Extremality and Stationarity Properties

There exist several modifications of the properties discussed in Subsection 2.3, scattered in the literature. Below we briefly discuss some of them which are going to be important for our subsequent study.

The properties in Definition 2.1 involve translations of all the sets. It is easy to see that in all the properties it is sufficient to consider translations of all but one sets leaving the remaining set unchanged. This simple observation leads to asymmetric conditions in the next proposition which can be useful, especially in the case  $n = 2$ .

**Proposition 3.1** *Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . The collection  $\{\Omega_1, \dots, \Omega_n\}$  is*

(i) *extremal if and only if, for any  $\varepsilon > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying*

$$\bigcap_{i=1}^{n-1} (\Omega_i - a_i) \cap \Omega_n = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n-1} \|a_i\| < \varepsilon; \quad (\text{P4})$$

(ii) *locally extremal at  $\bar{x}$  if and only if there exists a number  $\rho \in ]0, \infty]$  such that, for any  $\varepsilon > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying*

$$\bigcap_{i=1}^{n-1} (\Omega_i - a_i) \cap \Omega_n \cap B_\rho(\bar{x}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n-1} \|a_i\| < \varepsilon; \quad (\text{P5})$$

(iii) *stationary at  $\bar{x}$  if and only if, for any  $\varepsilon > 0$ , there exist a number  $\rho \in ]0, \varepsilon[$  and vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying*

$$\bigcap_{i=1}^{n-1} (\Omega_i - a_i) \cap \Omega_n \cap B_\rho(\bar{x}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n-1} \|a_i\| < \varepsilon\rho;$$

(iv) *approximately stationary at  $\bar{x}$  if and only if, for any  $\varepsilon > 0$ , there exist a number  $\rho \in ]0, \varepsilon[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  ( $i = 1, \dots, n$ ) and vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying*

$$\bigcap_{i=1}^{n-1} (\Omega_i - \omega_i - a_i) \cap (\Omega_n - \omega_n) \cap (\rho\mathbb{B}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n-1} \|a_i\| < \varepsilon\rho. \quad (\text{P6})$$

From now on we mostly focus on characterizations of approximate stationarity.

Observe that, by scaling the vectors, the normalization condition  $\sum_{i=1}^n \|x_i^*\| = 1$  in the dual generalized separation properties (D1) and (D2) can be dropped if the inequalities there are amended to

$$\sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) < \varepsilon \sum_{i=1}^n \|x_i^*\| \quad \text{and} \quad \left\| \sum_{i=1}^n x_i^* \right\| < \varepsilon \sum_{i=1}^n \|x_i^*\|,$$

respectively. Note that each of the amended inequalities still implies that  $\sum_{i=1}^n \|x_i^*\| > 0$ . Moreover, the normalization condition  $\sum_{i=1}^n \|x_i^*\| = 1$  involving  $n$  vectors can be replaced by the similar asymmetric condition involving  $n-1$  vectors:  $\sum_{i=1}^{n-1} \|x_i^*\| = 1$ . This observation can be especially useful in the case  $n = 2$ .

Observe further that parameter  $\varepsilon$  in the Definition 2.1(iv) of approximate stationarity and its reformulation in Proposition 3.1(iv) as well as the metric and dual characterizations in Theorems 2.1(ii) and (iii) and 2.2(ii) and (iii) plays multiple roles. To get a deeper insight into the approximate stationarity property, it makes sense to split the parameter  $\varepsilon$  into two components. From now on, we will use the letter  $\alpha$  to denote the component controlling the size of the shifts of the sets. This parameter is going to be crucial for quantifying the corresponding transversality property, playing the role of the rate/modulus of the property.

Based on the above observations, we now formulate a list of primal and dual equivalent characterizations of approximate stationarity, complementing Definition 2.1(iv), Proposition 3.1(iv), Theorem 2.1(ii) and (iii) and Theorem 2.2(ii) and (iii), which will be used in the sequel.

**Proposition 3.2** *Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . The following conditions are equivalent:*

(i) *the collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately stationary at  $\bar{x}$ ;*

(ii) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist a number  $\rho \in ]0, \varepsilon[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying

$$\bigcap_{i=1}^n (\Omega_i - \omega_i - a_i) \cap (\rho\mathbb{B}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \alpha\rho; \quad (\text{P7})$$

(iii) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist a number  $\rho \in ]0, \varepsilon[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  ( $i = 1, \dots, n$ ) and vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying

$$\bigcap_{i=1}^{n-1} (\Omega_i - \omega_i - a_i) \cap (\Omega_n - \omega_n) \cap (\rho\mathbb{B}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n-1} \|a_i\| < \alpha\rho; \quad (\text{P8})$$

(iv) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) such that

$$\alpha d\left(\bar{x}, \bigcap_{i=1}^n (\Omega_i - a_i)\right) > \max_{1 \leq i \leq n} d(\bar{x}, \Omega_i - a_i) \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon; \quad (\text{P9})$$

(v) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist a point  $x \in B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) such that

$$\alpha d\left(x, \bigcap_{i=1}^n (\Omega_i - a_i)\right) > \max_{1 \leq i \leq n} d(x, \Omega_i - a_i) \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \varepsilon. \quad (\text{P10})$$

With  $N$  standing for either Clarke ( $N := N^C$ ) or Fréchet ( $N := N^F$ ) normal cone, the following conditions are equivalent to conditions (ii) and (iii) in Theorem 2.2:

(vi) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that

$$\left\| \sum_{i=1}^n x_i^* \right\| < \alpha, \quad x_i^* \in N_{\Omega_i}(\omega_i) \quad (i = 1, \dots, n) \quad \text{and} \quad \sum_{i=1}^n \|x_i^*\| = 1; \quad (\text{D3})$$

(vii) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that

$$\sum_{i=1}^n x_i^* = 0, \quad \sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) < \alpha \quad \text{and} \quad \sum_{i=1}^n \|x_i^*\| = 1; \quad (\text{D4})$$

(viii) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that

$$\left\| \sum_{i=1}^n x_i^* \right\| < \alpha, \quad x_i^* \in N_{\Omega_i}(\omega_i) \quad (i = 1, \dots, n) \quad \text{and} \quad \sum_{i=1}^{n-1} \|x_i^*\| = 1; \quad (\text{D5})$$

(ix) for any  $\varepsilon > 0$  and  $\alpha > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that

$$\sum_{i=1}^n x_i^* = 0, \quad \sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) < \alpha \quad \text{and} \quad \sum_{i=1}^{n-1} \|x_i^*\| = 1. \quad (\text{D6})$$

If  $X$  is Asplund and  $N = N^F$ , then all conditions (i)–(ix) are equivalent.

**Remark 3.1** The maximum in each of the conditions in Definition 2.1, Theorem 2.1, Proposition 3.1 and the first (primal space) part of Proposition 3.2 can be replaced by the sum. Moreover, any norm on  $\mathbb{R}^n$  (or  $\mathbb{R}^{n-1}$  in the case of Proposition 3.1 and Proposition 3.2(iii)) can be used instead. The sum of the norms in the normalization conditions in (D1) and (D2) and the second (dual space) part of Proposition 3.2 stands for the corresponding dual norm and can be replaced by the maximum, or any other norm on  $\mathbb{R}^n$  or  $\mathbb{R}^{n-1}$ .

The “ $\alpha$ -version” of the approximate stationarity based on its equivalent representation in Proposition 3.2(i) is going to be used in the subsequent study.

**Definition 3.1 (Approximate  $\alpha$ -stationarity)** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  and  $\alpha > 0$ . The collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately  $\alpha$ -stationary at  $\bar{x}$  iff for any  $\varepsilon > 0$ , there exist a number  $\rho \in ]0, \varepsilon[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying conditions (P7).

In view of the above discussion, the next assertion is straightforward.

**Proposition 3.3** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . The collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately stationary at  $\bar{x}$  if and only if it is approximately  $\alpha$ -stationary at  $\bar{x}$  for all  $\alpha > 0$ .

In our deeper analysis of the core arguments in proofs of metric and dual characterizations of the extremality and stationarity in Sections 5 and 6, we will study properties which correspond to fixing, besides  $\alpha$ , also other parameters involved in the primal and dual properties discussed above.

Theorems 2.1 and 2.2 as well as the equivalent characterizations in Proposition 3.2 can be ‘reversed’ into statements providing primal and dual space criteria for the absence of the approximate stationarity, which turns out to be an important *regularity/transversality* property of collections of sets, which plays an important role in constraint qualifications, qualification conditions in subdifferential/coderivative calculus and convergence analysis of computational algorithms [5, 6, 25, 36, 44, 45, 48–52].

**Definition 3.2 (Transversality)** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . The collection  $\{\Omega_1, \dots, \Omega_n\}$  is transversal at  $\bar{x}$  iff there exist numbers  $\alpha > 0$  and  $\varepsilon > 0$  such that

$$\bigcap_{i=1}^n (\Omega_i - \omega_i - a_i) \cap (\rho \mathbb{B}) \neq \emptyset \quad (17)$$

for all  $\rho \in ]0, \varepsilon[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying  $\max_{1 \leq i \leq n} \|a_i\| < \alpha \rho$ .

The property in the above definition was first considered explicitly in [45]. It has been renamed multiple times (*regularity*, *strong regularity*, *property (UR)<sub>S</sub>*, *uniform regularity*, and finally *transversality*). A table illustrating the evolution of the terminology can be found in [26, Section 2]. The transversality property of collections of sets happens to be a direct counterpart of the *metric regularity* property of set-valued mappings; cf. [25, 44, 45].

The number  $\alpha$  involved in the definition provides a quantitative characterization of the property. The supremum of all such numbers (with the convention that the supremum of the empty subset of  $\mathbb{R}_+$  equals 0), denoted  $\text{tr}[\Omega_1, \dots, \Omega_n](\bar{x})$ , is the *modulus of transversality* of the collection  $\Omega_1, \dots, \Omega_n$  at  $\bar{x}$ . Thus, the case  $\text{tr}[\Omega_1, \dots, \Omega_n](\bar{x}) = 0$  corresponds to approximate stationarity.

In view of Proposition 3.2, the next two statements are direct consequences of Theorems 2.1 and 2.2, respectively.

**Theorem 3.1 (Transversality: metric criteria)** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . The following conditions are equivalent:

- (i) the collection  $\{\Omega_1, \dots, \Omega_n\}$  is transversal at  $\bar{x}$ ;
- (ii) there exist numbers  $\alpha > 0$  and  $\varepsilon > 0$  such that

$$\alpha d \left( \bar{x}, \bigcap_{i=1}^n (\Omega_i - a_i) \right) \leq \max_{1 \leq i \leq n} d(\bar{x}, \Omega_i - a_i)$$

for all vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ ;

- (iii) there exist numbers  $\alpha > 0$  and  $\varepsilon > 0$  such that

$$\alpha d \left( x, \bigcap_{i=1}^n (\Omega_i - a_i) \right) \leq \max_{1 \leq i \leq n} d(x, \Omega_i - a_i)$$

for all points  $x \in B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ .

The equivalence of conditions (i) and (iii) in Theorem 3.1 recaptures [45, Theorem 1]; cf. [44, Theorem 1] and [26, Definition 2]. Adding condition (ii) to the list of equivalent conditions sheds additional light on the transversality property.

**Theorem 3.2 (Transversality: dual criteria)** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of an Asplund space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . With  $N$  standing for the Fréchet normal cone ( $N := N^F$ ), the following conditions are equivalent:

- (i) the collection  $\{\Omega_1, \dots, \Omega_n\}$  is transversal at  $\bar{x}$ ;
- (ii) there exist numbers  $\alpha > 0$  and  $\varepsilon > 0$  such that  $\|\sum_{i=1}^n x_i^*\| > \alpha$  for all points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in N_{\Omega_i}(\omega_i)$  ( $i = 1, \dots, n$ ) satisfying  $\sum_{i=1}^n \|x_i^*\| = 1$ ;
- (iii) there exist numbers  $\alpha > 0$  and  $\varepsilon > 0$  such that  $\sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) > \alpha$  for all points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying  $\sum_{i=1}^n x_i^* = 0$  and  $\sum_{i=1}^n \|x_i^*\| = 1$ .

In view of Theorem 3.2, transversality of a collection of sets is equivalent to the absence of the generalized separation.

- Remark 3.2** (i) In view of Corollaries 2.1 and 2.2, conditions (ii) and (iii) in Theorem 3.2 are equivalent.  
(ii) The supremums of all numbers  $\alpha$  in parts (ii) and (iii) of Theorem 3.1 and part (ii) of Theorem 3.2 equal  $\text{tr}[\Omega_1, \dots, \Omega_n](\bar{x})$ ; see [45, Theorem 1], [25, Theorem 4(vi)] and Corollary 5.4 below. The supremum of all numbers  $\alpha$  in part (iii) of Theorem 3.2 can be different from  $\text{tr}[\Omega_1, \dots, \Omega_n](\bar{x})$ , but its relationship with  $\text{tr}[\Omega_1, \dots, \Omega_n](\bar{x})$  can be easily established using elementary arguments discussed in the next subsection.

*Example 3.1* In the space  $\mathbb{R}^2$  equipped with the maximum norm (hence, the dual norm is the sum norm), consider the two perpendicular lines:  $\Omega_1 := \{(t, 0) : t \in \mathbb{R}\}$  and  $\Omega_2 := \{(0, t) : t \in \mathbb{R}\}$ . Then we have  $\bar{x} := (0, 0) \in \Omega_1 \cap \Omega_2$ ,  $N_{\Omega_1}^F(x) = \{(0, t) : t \in \mathbb{R}\}$  for any  $x \in \Omega_1$  and  $N_{\Omega_2}^F(x) = \{(t, 0) : t \in \mathbb{R}\}$  for any  $x \in \Omega_2$ .

If  $x_1^* = (0, t_1)$  and  $x_2^* = (t_2, 0)$  are normal vectors to  $\Omega_1$  and  $\Omega_2$ , respectively, then  $\|x_1^*\| + \|x_2^*\| = \|x_1^* + x_2^*\| = |t_1| + |t_2|$ . Hence, the supremum of all  $\alpha$  in part (ii) of Theorem 3.2 is 1 (and is equal to  $\text{tr}[\Omega_1, \Omega_2]$ ).

If  $x_1^*, x_2^* \in (\mathbb{R}^2)^*$ ,  $x_1^* + x_2^* = 0$  and  $\|x_1^*\| + \|x_2^*\| = 1$ , then  $x_1^* = -x_2^* = (t_1, t_2)$  for some numbers  $t_1$  and  $t_2$  satisfying  $|t_1| + |t_2| = \frac{1}{2}$ , and the distances from  $x_1^*$  and  $x_2^*$  to the corresponding normal cones equal  $|t_1|$  and  $|t_2|$ , respectively. Hence, the supremum of all  $\alpha$  in part (iii) of Theorem 3.2 is  $\frac{1}{2}$ .

**Remark 3.3** Since the approximate stationarity and transversality properties are complementary to each other, it would be natural to refer to the negation of the approximate  $\alpha$ -stationarity property, i.e., the existence of a number  $\varepsilon > 0$  such that condition (17) holds for all  $\rho \in ]0, \varepsilon[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying  $\max_{1 \leq i \leq n} \|a_i\| < \alpha\rho$ , as  $\alpha$ -transversality at  $\bar{x}$ . We do not consider this property in the current paper.

## 4 Distances between $n$ Sets

When studying mutual arrangement of collections of sets in space, particularly their extremality, stationarity and regularity properties, we need to be able to estimate the ‘distance’ between the sets, i.e. how ‘far apart’ they are, or, at least, whether they have a common point. In the case of two sets, the conventional distance

$$d(\Omega_1, \Omega_2) := \inf_{\omega_1 \in \Omega_1, \omega_2 \in \Omega_2} d(\omega_1, \omega_2) \quad (18)$$

does the job. The general case of  $n \geq 2$  sets is not that straightforward. In this section we discuss several candidates for the role of ‘distance’.

We start with discussing distances between  $n \geq 2$  points.

### 4.1 Distances between $n$ Points

Our aim in this subsection is to consider ways of defining  $n$ -point distances estimating quantitatively the overall ‘closeness’ of a collection of  $n \geq 2$  points in a metric space  $(X, d)$ .

Given  $n \geq 2$  points  $\omega_1, \dots, \omega_n$  in  $X$ , one can use one of the following two quantities:

$$d_1(\omega_1, \dots, \omega_n) := \max_{1 \leq i \leq n-1} d(\omega_i, \omega_n), \quad (19)$$

$$d_2(\omega_1, \dots, \omega_n) := \inf_{x \in X} \max_{1 \leq i \leq n} d(\omega_i, x). \quad (20)$$

When  $n = 2$ , (19) reduces to the conventional distance  $d(\omega_1, \omega_2)$ . However, when  $n > 2$  this distance is not symmetric: the last point in the list plays a special role, and the quantity itself depends on the choice of the last point. In contrast to (19), in definition (20) all points play the same role, and it involves minimization over an additional parameter  $x$ . Formulas (19) and (20) produce in general different numbers, whatever the choice of the last point in (19) is.

In the setting of a normed vector space, the following symmetric distance can be of interest:

$$d_3(\omega_1, \dots, \omega_n) := \max_{1 \leq i \leq n} \left\| \omega_i - \frac{1}{n} \sum_{j=1}^n \omega_j \right\|. \quad (21)$$

*Example 4.1* Consider three points in  $\mathbb{R}$ : 0, 1 and 5. By (19), we have  $d_1(0, 5, 1) = \max\{|0 - 1|, |5 - 1|\} = 4$ , while  $d_1(0, 1, 5) = d_1(1, 5, 0) = 5$ . The infimum in definition (20) is achieved at  $x = 2.5$  and equals  $d_2(0, 1, 5) = \max\{|0 - 2.5|, |1 - 2.5|, |5 - 2.5|\} = 2.5$ . The average  $\frac{1}{3}(0 + 1 + 5)$  equals 2, and formula (21) gives  $d_3(0, 1, 5) = \max\{|0 - 2|, |1 - 2|, |5 - 2|\} = 3$ . Thus, for these three points all three definitions (19), (20) and (21) give different numbers.

Note the obvious connection between the distances  $d_1$  and  $d_2$ :

$$d_2(\omega_1, \dots, \omega_n) := \inf_{x \in X} d_1(\omega_1, \dots, \omega_n, x).$$

Observe also that  $d_1(\omega_1, \dots, \omega_n)$  is actually the usual (maximum) distance between the points  $(\omega_1, \dots, \omega_{n-1})$  and  $(\omega_n, \dots, \omega_n)$  in  $X^{n-1}$ , while  $d_2(\omega_1, \dots, \omega_n)$  represents the distance from the points  $(\omega_1, \dots, \omega_n) \in X^n$  to the ‘diagonal’ subspace  $\{(\omega, \dots, \omega) \in X^n : \omega \in X\}$ . If the points  $\omega_1, \dots, \omega_n$  do not all coincide, then the quantity computed in accordance with formula (20) will remain strictly positive if the infimum there is taken not over the whole space, but over any its subset. This can be a way of defining ‘localized distances’.

Some properties of the quantities (19) and (20) are collected in the next proposition.

**Proposition 4.1** *Suppose  $\omega_1, \dots, \omega_n$  ( $n \geq 2$ ) are points in a metric space  $X$ .*

- (i)  $d_2(\omega_1, \dots, \omega_n) \leq d_1(\omega_1, \dots, \omega_n) \leq 2d_2(\omega_1, \dots, \omega_n)$ .
- (ii) *If  $X$  is a normed vector space with the distance induced by the norm and  $n = 2$ , then  $d_1(\omega_1, \omega_2) = 2d_2(\omega_1, \omega_2)$ , i.e. the last inequality in (i) holds as equality.*

*Suppose additionally that  $X$  is a normed vector space.*

- (iii)  $d_2(\omega_1, \dots, \omega_n) \leq d_3(\omega_1, \dots, \omega_n) \leq 2d_2(\omega_1, \dots, \omega_n)$ .
- (iv) *If  $n = 2$ , then  $d_2(\omega_1, \omega_2) = d_3(\omega_1, \omega_2)$ , i.e. the first inequality in (iii) holds as equality.*

**Proof.** (i) The first inequality follows immediately from the definitions:

$$d_2(\omega_1, \dots, \omega_n) = \inf_{x \in X} \max_{1 \leq i \leq n} d(\omega_i, x) \leq \max_{1 \leq i \leq n} d(\omega_i, \omega_n) = \max_{1 \leq i \leq n-1} d(\omega_i, \omega_n) = d_1(\omega_1, \dots, \omega_n).$$

To prove the second inequality, first fix an  $x \in X$ .

$$d_1(\omega_1, \dots, \omega_n) = \max_{1 \leq i \leq n-1} d(\omega_i, \omega_n) \leq \max_{1 \leq i \leq n-1} (d(\omega_i, x) + d(\omega_n, x)) \leq 2 \max_{1 \leq i \leq n} d(\omega_i, x).$$

Taking the infimum over  $x \in X$  in the right-hand side of the above inequality, we arrive at the second inequality in (i).

- (ii) Let  $X$  be a normed vector space with the distance induced by the norm and  $n = 2$ . Then

$$\begin{aligned} d_1(\omega_1, \omega_2) &= \|\omega_1 - \omega_2\| = 2 \max \left\{ \left\| \frac{\omega_1 + \omega_2}{2} - \omega_1 \right\|, \left\| \frac{\omega_1 + \omega_2}{2} - \omega_2 \right\| \right\} \\ &\geq 2 \inf_{x \in X} \max \{ \|x - \omega_1\|, \|x - \omega_2\| \} = 2d_2(\omega_1, \omega_2). \end{aligned}$$

Combining this with the second inequality in (i) proves (ii).

- (iii) Suppose that  $X$  is a normed vector space. The first inequality follows immediately from the definitions. To prove the second inequality, first fix an  $x \in X$ . Then

$$\begin{aligned} d_3(\omega_1, \dots, \omega_n) &= \max_{1 \leq i \leq n} \left\| \omega_i - \frac{1}{n} \sum_{j=1}^n \omega_j \right\| \leq \max_{1 \leq i \leq n} \|\omega_i - x\| + \left\| \frac{1}{n} \sum_{j=1}^n \omega_j - x \right\| \\ &\leq \max_{1 \leq i \leq n} \|\omega_i - x\| + \frac{1}{n} \sum_{j=1}^n \|\omega_j - x\| \leq 2 \max_{1 \leq i \leq n} \|\omega_i - x\|. \end{aligned}$$

Taking the infimum over  $x \in X$  in the right-hand side of the second inequality, we arrive at the last inequality in (iii).

- (iv) Let  $n = 2$ . Using definitions (21) and (19), and the equality in (ii), we obtain:

$$d_3(\omega_1, \omega_2) = \max \left\{ \left\| \omega_1 - \frac{\omega_1 + \omega_2}{2} \right\|, \left\| \omega_2 - \frac{\omega_1 + \omega_2}{2} \right\| \right\} = \frac{1}{2} \|\omega_1 - \omega_2\| = \frac{1}{2} d_1(\omega_1, \omega_2) = d_2(\omega_1, \omega_2).$$

□

It follows from part (ii) of Proposition 4.1 that quantity (20) does not reduce to the conventional distance when  $n = 2$ : in the setting of a normed vector space it equals  $\frac{1}{2}d(\omega_1, \omega_2)$ . In part (ii) of Proposition 4.1, the infimum in formula (20) is computed explicitly. Unfortunately, when  $n > 2$ , this seems impossible in general even in the setting of a normed vector space; see the discussion in [53, Section 6].

**Remark 4.1** In the setting of a normed vector space, if  $n = 2$  and  $\omega_1 \neq \omega_2$ , then, in view of Proposition 4.1(ii) and (iv), the first inequality in Proposition 4.1(i) and the second inequality in Proposition 4.1(iii) are strict. If  $n > 2$ , then all the inequalities in Proposition 4.1(i) and (iii) can be strict. This fact is illustrated by Example 4.1.

## 4.2 Distances between $n$ Sets

Now we employ the distances between collections of points discussed in the previous subsection to quantify ‘closeness’ of collections of sets. Given  $n \geq 2$  subsets  $\Omega_1, \dots, \Omega_n$  of a metric space  $X$  and an  $n$ -point distance  $d$ , the distance between  $\Omega_1, \dots, \Omega_n$  is defined in the usual way:

$$d(\Omega_1, \dots, \Omega_n) := \inf_{\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n} d(\omega_1, \dots, \omega_n). \quad (22)$$

Applying construction (22) to the  $n$ -point distances (19), (20) and (21), we obtain the following definitions of particular distances between  $n$  sets, respectively:

$$d_1(\Omega_1, \dots, \Omega_n) := \inf_{\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n} \max_{1 \leq i \leq n-1} d(\omega_i, \omega_n), \quad (23)$$

$$d_2(\Omega_1, \dots, \Omega_n) := \inf_{\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n, x \in X} \max_{1 \leq i \leq n} d(\omega_i, x), \quad (24)$$

$$d_3(\Omega_1, \dots, \Omega_n) := \inf_{\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n} \max_{1 \leq i \leq n} \left\| \omega_i - \frac{1}{n} \sum_{j=1}^n \omega_j \right\|. \quad (25)$$

When  $n = 2$ , definition (23) reduces to the conventional distance (18). Definition (25) is meaningful in the setting of a normed vector space only.

Proposition 4.1 leads to similar relations for the distances between  $n$  sets.

**Proposition 4.2** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a metric space  $X$ .

(i)  $d_2(\Omega_1, \dots, \Omega_n) \leq d_1(\Omega_1, \dots, \Omega_n) \leq 2d_2(\Omega_1, \dots, \Omega_n)$ .

(ii) If  $n = 2$ , then  $d_1(\Omega_1, \Omega_2) = 2d_2(\Omega_1, \Omega_2)$ .

(iii)  $d_1(\Omega_1, \dots, \Omega_n) > 0 \iff d_2(\Omega_1, \dots, \Omega_n) > 0$ .

Suppose additionally that  $X$  is a normed vector space.

(iv)  $d_2(\Omega_1, \dots, \Omega_n) \leq d_3(\Omega_1, \dots, \Omega_n) \leq 2d_2(\Omega_1, \dots, \Omega_n)$ .

(v) If  $n = 2$ , then  $d_2(\Omega_1, \Omega_2) = d_3(\Omega_1, \Omega_2)$ .

(vi)  $d_1(\Omega_1, \dots, \Omega_n) > 0 \iff d_2(\Omega_1, \dots, \Omega_n) > 0 \iff d_3(\Omega_1, \dots, \Omega_n) > 0$ .

**Proof.** Conditions (i), (ii) and (iv), (v) are consequences of the corresponding conditions in Proposition 4.1. Condition (iii) is a consequence of condition (i). Condition (vi) is a consequence of conditions (iii) and (iv).  $\square$

Another distance can be of interest and is going to be used in the sequel. Given a subset  $\Omega_{n+1}$  of  $X$ , define

$$d_{\Omega_{n+1}}(\Omega_1, \dots, \Omega_n) := \inf_{\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n, x \in \Omega_{n+1}} \max_{1 \leq i \leq n} d(\omega_i, x).$$

Now observe that  $d_{\Omega_{n+1}}$  as well as the symmetric distance  $d_2$  (24) are particular cases of the asymmetric distance  $d_1$  (23) applied to  $n+1$  sets:

$$d_{\Omega_{n+1}}(\Omega_1, \dots, \Omega_n) = d_1(\Omega_1, \dots, \Omega_n, \Omega_{n+1}) \quad \text{and} \quad d_2(\Omega_1, \dots, \Omega_n) = d_1(\Omega_1, \dots, \Omega_n, X). \quad (26)$$

This observation makes the asymmetric distance  $d_1$  a rather general quantitative measure of closeness of collections of sets. Apart from the most straightforward case  $\Omega_{n+1} := X$  used in the above example, another useful particular case is given by  $\Omega_{n+1} := B_\rho(\bar{x})$  where  $\bar{x} \in X$  is a fixed point (related to the sets  $\Omega_1, \dots, \Omega_n$ ) and  $\rho > 0$ . This allows one to examine closeness of sets in a neighbourhood of the given point.

- Remark 4.2** (i) As it was observed earlier, the distance  $d_1$  (23) depends in general on the order of the sets. However, thanks to Proposition 4.2(iii), if it is strictly positive for some permutation of the sets, it remains strictly positive for any other permutation.
- (ii) The maximum operation in all the above definitions of distances between  $n$  points and  $n$  sets corresponds to the maximum norm in either  $\mathbb{R}^{n-1}$  or  $\mathbb{R}^n$ . It can be replaced in all these definitions and subsequent statements by any other finite dimensional norm producing different but in a sense equivalent ‘distances’. The  $p$ -norm version of the quantity (23) was considered in [31] under the name ( $p$ -weighted) *nonintersect index*. In the current paper, for the sake of simplicity of presentation only the maximum norm is considered.

The next proposition and its corollary characterize a set of points in the given collection of sets, which are almost closest (up to  $\varepsilon$ ) points of these sets with respect to the chosen  $n$ -point distance. (When  $n = 2$  such points play a key role in the geometric versions of the Ekeland variational principle considered in Section 2.) It also introduces a two-step procedure, which is going to be used in the sequel. Given a collection of sets  $\Omega_1, \dots, \Omega_n$  with empty intersection and a collection of points  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ), we

- 1) consider another collection of sets  $\Omega_1 - \omega_1, \dots, \Omega_n - \omega_n$ , whose intersection is obviously nonempty, and
- 2) construct ‘small’ (up to  $\varepsilon$ ) translation vectors  $a_1, \dots, a_n$  such that the translated sets  $\Omega_1 - \omega_1 - a_1, \dots, \Omega_n - \omega_n - a_n$  have empty intersection again.

Thus, the proposition translates  $\varepsilon$ -closeness of a collection of points, which served as the key assumption when proving unified separation theorems in [31], into the language of  $\varepsilon$ -translations of the sets employed in Definition 2.1 of extremality/stationarity properties, the corresponding Proposition 3.1 and parts (ii) and (iii) of Proposition 3.2.

For simplicity and in view of the observed above universality of the distance  $d_1$  (23), we first consider this distance.

**Proposition 4.3** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\bigcap_{i=1}^n \Omega_i = \emptyset$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) and  $\varepsilon > 0$ . If

$$d_1(\omega_1, \dots, \omega_n) < d_1(\Omega_1, \dots, \Omega_n) + \varepsilon, \quad (27)$$

then  $M := d_1(\omega_1, \dots, \omega_n) > 0$  and condition (P8) is satisfied, where  $a_i := \frac{\varepsilon'}{M}(\omega_n - \omega_i)$  ( $i = 1, \dots, n-1$ ) and  $\varepsilon'$  is either any number in  $]0, \varepsilon[$  satisfying  $M - d_1(\Omega_1, \dots, \Omega_n) < \varepsilon' \leq M$  if  $d_1(\Omega_1, \dots, \Omega_n) > 0$ , or  $\varepsilon' = M$  if  $d_1(\Omega_1, \dots, \Omega_n) = 0$ .

**Proof.** In view of  $\bigcap_{i=1}^n \Omega_i = \emptyset$ , we have  $M > 0$ . Let condition (27) be satisfied.

Suppose first that  $d_1(\Omega_1, \dots, \Omega_n) > 0$ . Choose a positive number  $\varepsilon' < \varepsilon$  such that  $M - d_1(\Omega_1, \dots, \Omega_n) < \varepsilon' \leq M$ . Then  $\max_{1 \leq i \leq n-1} \|a_i\| = \varepsilon' < \varepsilon$ . Suppose that the first condition in (P8) does not hold. Then there exists a point  $x \in \bigcap_{i=1}^{n-1} (\Omega_i - \omega_i - a_i) \cap (\Omega_n - \omega_n)$ , and consequently,  $\hat{\omega}_i := \omega_i + a_i + x \in \Omega_i$  ( $i = 1, \dots, n-1$ ) and  $\hat{\omega}_n := \omega_n + x \in \Omega_n$ . Thus,

$$\|\hat{\omega}_i - \hat{\omega}_n\| = \left\| \omega_i - \omega_n - \frac{\varepsilon'}{M}(\omega_i - \omega_n) \right\| = \left(1 - \frac{\varepsilon'}{M}\right) \|\omega_i - \omega_n\| \quad (i = 1, \dots, n-1),$$

and consequently,

$$d_1(\Omega_1, \dots, \Omega_n) \leq d_1(\hat{\omega}_1, \dots, \hat{\omega}_n) = \left(1 - \frac{\varepsilon'}{M}\right) d_1(\omega_1, \dots, \omega_n) = M - \varepsilon'.$$

This contradicts the choice of  $\varepsilon'$ . Hence, conditions (P8) hold true.

In the case  $d_1(\Omega_1, \dots, \Omega_n) = 0$ , set  $a_i := \omega_n - \omega_i$  ( $i = 1, \dots, n-1$ ). Then, by (27),  $\max_{1 \leq i \leq n-1} \|a_i\| = d_1(\omega_1, \dots, \omega_n) < \varepsilon$  and

$$\bigcap_{i=1}^{n-1} (\Omega_i - \omega_i - a_i) \cap (\Omega_n - \omega_n) = \bigcap_{i=1}^n (\Omega_i - \omega_n) = \bigcap_{i=1}^n \Omega_i - \omega_n = \emptyset,$$

i.e., conditions (P8) hold true.  $\square$

Applying Proposition 4.3 to the collection of  $n+1$  sets  $\Omega_1, \dots, \Omega_n, X$ , we arrive at the following statement in terms of the distance  $d_2$  (24).

**Corollary 4.1** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\bigcap_{i=1}^n \Omega_i = \emptyset$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ),  $x \in X$  and  $\varepsilon > 0$ . If

$$\max_{1 \leq i \leq n} d(\omega_i, x) < d_2(\Omega_1, \dots, \Omega_n) + \varepsilon, \quad (28)$$

then  $M := \max_{1 \leq i \leq n} \|\omega_i - x\| > 0$  and conditions (P1) hold true with  $\Omega'_i := \Omega_i - \omega_i$  in place of  $\Omega_i$  and  $a_i := \frac{\varepsilon'}{M}(x - \omega_i)$  ( $i = 1, \dots, n$ ), where  $\varepsilon'$  is either any number in  $]0, \varepsilon[$  satisfying  $M - d_2(\Omega_1, \dots, \Omega_n) < \varepsilon' \leq M$  if  $d_2(\Omega_1, \dots, \Omega_n) > 0$ , or  $\varepsilon' = M$  if  $d_2(\Omega_1, \dots, \Omega_n) = 0$ .

## 5 Localizations and Translations of the Sets

In this section, we continue studying mutual arrangement of  $n$  sets in space started in Section 4.

The next lemma extends the Asymmetric Geometric Ekeland Variational Principle (AGEVP) from Section 2 to the case of  $n \geq 2$  sets, and as such it is also equivalent to the Ekeland Variational Principle. Just like in AGEVP, balls of two different radii are used in the concluding part of the lemma, one of them employed in the localizations of the first  $n - 1$  sets and the other one for the remaining single set. The latter set is going to play a special role in the subsequent analysis.

The lemma translates  $\varepsilon$ -closeness of a given collection of points into  $\xi\varepsilon$ -closeness (with an additional parameter  $\xi$ ) of another collection of points with respect to a collection of certain localizations of the sets (depending on  $\xi$ ).

**Lemma 5.1** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a complete metric space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ), and  $\varepsilon > 0$ . If condition (27) is satisfied, then, for all numbers  $\lambda, \rho > 0$ , there exist  $\hat{\omega}_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n - 1$ ) and  $\hat{\omega}_n \in \Omega_n \cap B_\rho(\omega_n)$  such that*

- (i)  $d_1(\hat{\omega}_1, \dots, \hat{\omega}_n) \leq d_1(\omega_1, \dots, \omega_n)$ ;
- (ii)  $d_1(\Omega_1 \cap B_{\xi\lambda}(\hat{\omega}_1), \dots, \Omega_{n-1} \cap B_{\xi\lambda}(\hat{\omega}_{n-1}), \Omega_n \cap B_{\xi\rho}(\hat{\omega}_n)) + \xi\varepsilon > d_1(\hat{\omega}_1, \dots, \hat{\omega}_n)$  for all  $\xi > 0$ .

**Proof.** The product space  $X^{n-1}$  considered with the maximum metric is complete. Set  $A := \Omega_1 \times \dots \times \Omega_{n-1}$ , and  $B := \{(x, \dots, x) : x \in \Omega_n\} \subset X^{n-1}$ ,  $a := (\omega_1, \dots, \omega_{n-1}) \in A$ ,  $b := (\omega_n, \dots, \omega_n) \in B$ . Then  $d_1(\Omega_1, \dots, \Omega_n) = d(A, B)$  and  $d_1(\omega_1, \dots, \omega_n) = d(a, b)$ .

Applying AGEVP, we find points  $\hat{a} = (\hat{\omega}_1, \dots, \hat{\omega}_{n-1}) \in A \cap B_\lambda(a)$  and  $\hat{b} = (\hat{\omega}_n, \dots, \hat{\omega}_n) \in B \cap B_\rho(b)$  satisfying conditions (i) and (ii) in AGEVP. Recalling that the maximum metric is used in  $X^{n-1}$ , it follows that  $\hat{\omega}_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n - 1$ ),  $\hat{\omega}_n \in B \cap B_\rho(\omega_n)$  and conditions (i) and (ii) above are satisfied.  $\square$

In view of the equalities in (26), the following corollary involving the symmetric distance  $d_2$  is immediate.

**Corollary 5.1** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a complete metric space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ),  $x \in X$  and  $\varepsilon > 0$ . If condition (28) is satisfied, then, for all  $\lambda, \rho > 0$ , there exist  $\hat{\omega}_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) and  $\hat{x} \in B_\rho(x)$  such that*

- (i)  $\max_{1 \leq i \leq n} d(\hat{\omega}_i, \hat{x}) \leq \max_{1 \leq i \leq n} d(\omega_i, x)$ ;
- (ii)  $d_{B_{\xi\rho}(\hat{x})}(\Omega_1 \cap B_{\xi\lambda}(\hat{\omega}_1), \dots, \Omega_n \cap B_{\xi\lambda}(\hat{\omega}_n)) + \xi\varepsilon > \max_{1 \leq i \leq n} d(\hat{\omega}_i, \hat{x})$  for all  $\xi > 0$ .

The next proposition is a consequence of Lemma 5.1 in the Banach space setting. It characterizes  $\varepsilon$ -closest points of a collection of sets with empty intersection and involves localizations of the sets and small (up to  $\xi\varepsilon$ ) translations of the first  $n - 1$  localizations. The proposition transforms the nonlocal non-intersection condition  $\bigcap_{i=1}^n \Omega_i = \emptyset$  into a non-intersection condition of translated localizations of the sets. It also exposes the special role played by the last set in the list.

**Proposition 5.1** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ), and  $\varepsilon > 0$ . If  $\bigcap_{i=1}^n \Omega_i = \emptyset$  and condition (27) is satisfied, then, for any  $\lambda, \rho > 0$ , there are points  $\hat{\omega}_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n - 1$ ),  $\hat{\omega}_n \in \Omega_n \cap B_\rho(\omega_n)$  such that, for any  $\xi > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n - 1$ ) satisfying*

$$\bigcap_{i=1}^{n-1} \left( ((\Omega_i - \hat{\omega}_i) \cap (\xi\lambda)\mathbb{B}) - a_i \right) \cap (\Omega_n - \hat{\omega}_n) \cap (\xi\rho)\mathbb{B} = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n-1} \|a_i\| < \xi\varepsilon. \quad (29)$$

**Proof.** Applying Lemma 5.1, we find points  $\hat{\omega}_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n - 1$ ) and  $\hat{\omega}_n \in \Omega_n \cap B_\rho(\omega_n)$  satisfying condition (ii) in that lemma. Given any number  $\xi > 0$ , we can now apply Proposition 4.3 with the function  $d_1$  and sets  $\Omega_1 \cap B_{\xi\lambda}(\hat{\omega}_1), \dots, \Omega_{n-1} \cap B_{\xi\lambda}(\hat{\omega}_{n-1}), \Omega_n \cap B_{\xi\rho}(\hat{\omega}_n)$ , points  $\hat{\omega}_i$  ( $i = 1, \dots, n$ ) and number  $\xi\varepsilon$  in place of sets  $\Omega_1, \dots, \Omega_n$ , points  $\omega_i$  ( $i = 1, \dots, n$ ) and number  $\varepsilon$ , respectively, to find vectors  $a_i \in X$  ( $i = 1, \dots, n - 1$ ) such that condition (29) is satisfied.  $\square$

**Remark 5.1** (i) Under the conditions of Proposition 5.1,  $0 \in \bigcap_{i=1}^n (\Omega_i - \omega_i)$ ,  $0 \in \bigcap_{i=1}^n (\Omega_i - \hat{\omega}_i)$ , and the expressions involved in the first condition in (29) correspond to localizations of the sets  $\Omega_i - \hat{\omega}_i$  ( $i = 1, \dots, n$ ) near 0, or equivalently, localizations of the original sets  $\Omega_i$  near  $\hat{\omega}_i$  ( $i = 1, \dots, n$ ).

- (ii) There are certain similarities between Propositions 4.3 and 5.1 in terms of both assumptions and conclusions. There are also important differences. The assumptions of Propositions 4.3 are weaker: the space is not assumed to be complete and the sets are not assumed to be closed. The concluding non-intersection condition in (P8) (Propositions 4.3) is formulated for the given sets and in terms of the given collection of points, while Proposition 5.1 establishes existence of another collection of points, and the corresponding condition in (29) is formulated for localizations of the sets near these points. The translations of the sets are constructed in Propositions 4.3 explicitly and are entirely determined by the given collection of points. At the same time, in Proposition 5.1 the size of the translations and localizations of the sets as well as the distance of the new points from the given ones are controlled by additional parameters. These parameters, which appear in Proposition 5.1, represent the major feature of this statement compared to Propositions 4.3. They provide an additional degree (degrees) of freedom for the applications of the result.

Applying Proposition 5.1 to the collection of  $n + 1$  sets  $\Omega_1, \dots, \Omega_n, X$ , we arrive at the following statement.

**Corollary 5.2** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ),  $x \in X$  and  $\varepsilon > 0$ . If  $\bigcap_{i=1}^n \Omega_i = \emptyset$  and condition (28) is satisfied, then, for any numbers  $\lambda, \rho > 0$ , there are points  $\hat{\omega}_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) such that, for any  $\xi > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying*

$$\bigcap_{i=1}^n \left( ((\Omega_i - \hat{\omega}_i) \cap (\xi\lambda)\mathbb{B}) - a_i \right) \cap (\xi\rho)\mathbb{B} = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a_i\| < \xi\varepsilon. \quad (30)$$

Observe that conditions (29) and (30) are exactly conditions (P5) and (P2), respectively, applied to the localizations of the sets  $\Omega_1 - \hat{\omega}_1, \dots, \Omega_n - \hat{\omega}_n$  with  $\bar{x} := 0$ , and  $\xi\varepsilon$  and  $\xi\rho$  in place of  $\varepsilon$  and  $\rho$ , respectively.

Proposition 5.1 assumes that the sets have empty intersection. Next we demonstrate that it can be also applied to collections of sets having a common point. Specifically, we consider the special case (P2) in the Definition 2.1(ii) of local extremality, where the last set in the list (of  $n + 1$  sets) is a ball centered at this point. Since the radius of the ball is allowed to be infinite, the global setting is covered too.

**Proposition 5.2** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ ,  $\varepsilon > 0$  and  $\rho \in ]0, \infty]$ . If conditions (P2) are satisfied for some vectors  $a_i \in X$  ( $i = 1, \dots, n$ ), then, for any  $\lambda > 0$ , there are points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and a number  $\delta \in ]0, 1[$  such that, for all  $\xi \in ]0, \delta[$ ,*

$$\bigcap_{i=1}^n \left( ((\Omega_i - \omega_i) \cap (\xi\lambda)\mathbb{B}) - a'_i \right) \cap (\xi\rho)\mathbb{B} = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a'_i\| < \xi\varepsilon \quad (31)$$

for some vectors  $a'_i \in X$  ( $i = 1, \dots, n$ ).

Moreover, if  $\lambda \geq \rho + \varepsilon$ , then

$$\bigcap_{i=1}^n (\Omega_i - \omega_i - a'_i) \cap (\xi\rho)\mathbb{B} = \emptyset; \quad (32)$$

if  $\lambda + \varepsilon \leq \rho$ , then

$$\bigcap_{i=1}^n \left( ((\Omega_i - \omega_i) \cap (\xi\lambda)\mathbb{B}) - a'_i \right) = \emptyset. \quad (33)$$

**Proof.** Let  $\rho' \in ]0, \rho[$ . By (P2),  $\bigcap_{i=1}^{n+1} \Omega'_i = \emptyset$ , where  $\Omega'_i := \Omega_i - a_i$  ( $i = 1, \dots, n$ ) and  $\Omega'_{n+1} := \bar{B}_{\rho'}(\bar{x})$ , and

$$d_1(\bar{x} - a_1, \dots, \bar{x} - a_n, \bar{x}) = \max_{1 \leq i \leq n} \|a_i\| < \varepsilon.$$

Applying Proposition 5.1, we can find points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in X$  with  $\|x\| < \rho'$  such that, for any  $\xi > 0$ , there exist vectors  $a'_i \in X$  ( $i = 1, \dots, n$ ) satisfying

$$\bigcap_{i=1}^n \left( ((\Omega_i - \omega_i) \cap (\xi\lambda)\mathbb{B}) - a'_i \right) \cap \bar{B}_{\rho'}(x) \cap (\xi\rho)\mathbb{B} = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n} \|a'_i\| < \xi\varepsilon.$$

Set  $\delta := (\rho' - \|x\|)/\rho$ . Then,  $\delta \in ]0, 1[$  and, for any  $\xi \in ]0, \delta[$ , we have  $(\xi\rho)\mathbb{B} \subset \bar{B}_{\rho'}(x)$ , and consequently, (31) holds true. If  $\lambda \geq \rho + \varepsilon$ , then  $(\xi\rho)\mathbb{B} \subset (\xi\lambda)\mathbb{B} - a'_i$  for all  $i = 1, \dots, n$ , and consequently, (31) implies (32). Similarly, if  $\lambda + \varepsilon \leq \rho$ , then  $(\xi\lambda)\mathbb{B} - a'_i \subset (\xi\rho)\mathbb{B}$  for all  $i = 1, \dots, n$ , and consequently, (31) implies (33).  $\square$

**Remark 5.2** Conditions (P2) in Proposition 5.2 are formulated for a fixed point  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  and fixed  $\varepsilon > 0$  and  $\rho \in ]0, \infty]$ . It presumes a certain balance between the values of  $\varepsilon$  and  $\rho$ : the larger the value of  $\rho$  is, the larger value of  $\varepsilon$  is needed to ensure the existence of vectors  $a'_i \in X$  ( $i = 1, \dots, n$ ) satisfying (P2). In contrast, condition (31) involves two additional parameters: an arbitrary  $\lambda > 0$  and a sufficiently small  $\xi \in ]0, 1[$ . Fixed  $\varepsilon$  and  $\rho$  are replaced by  $\xi\varepsilon$  and  $\xi\rho$ , respectively, preserving their ratio, while the sets are replaced by their localizations controlled by  $\xi\lambda$ . This advancement comes at a price: instead of a single common fixed point  $\bar{x}$ , we now have to deal with a collection of individual points  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ), whose distance from  $\bar{x}$  is controlled by  $\lambda$ . Now the balance between this distance and the size of the localizations of the sets becomes important: choosing a smaller  $\lambda$  ensures that the individual points  $\omega_i$  are closer to  $\bar{x}$  while at the same time reducing the size of the localizations of the sets and, thus, weakening condition (31).

Given a collection of sets  $\Omega_1, \dots, \Omega_n$  ( $n \geq 2$ ), a point  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  and a  $\rho \in ]0, \infty]$ , define (cf. [24, 45]):

$$\theta_\rho[\Omega_1, \dots, \Omega_n](\bar{x}) = \sup \left\{ r > 0 : \bigcap_{i=1}^n (\Omega_i - a_i) \cap B_\rho(\bar{x}) \neq \emptyset \text{ for all } a_i \in r\mathbb{B} \right\}.$$

This nonnegative quantity tells us how far the sets can be pushed apart until their intersection becomes empty with respect to the fixed  $\rho$ -neighbourhood of  $\bar{x}$ .

**Corollary 5.3** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$  and  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ . If  $\rho > 0$  and  $\varepsilon > \theta_\rho[\Omega_1, \dots, \Omega_n](\bar{x})$ , then

$$\inf_{\substack{\omega_i \in \Omega_i \cap B_{\rho+\varepsilon}(\bar{x}) \\ (i=1, \dots, n)}} \limsup_{\alpha \downarrow 0} \alpha^{-1} \theta_\alpha[\Omega_1 - \omega_1, \dots, \Omega_n - \omega_n](0) \leq \rho^{-1} \theta_\rho[\Omega_1, \dots, \Omega_n](\bar{x}).$$

Moreover, if  $\bar{x} \in \text{bd} \bigcap_{i=1}^n \Omega_i$ , then

$$\liminf_{\substack{\omega_i \rightarrow \bar{x}, \omega_i \in \Omega_i \\ (i=1, \dots, n)}} \limsup_{\alpha \downarrow 0} \alpha^{-1} \theta_\alpha[\Omega_1 - \omega_1, \dots, \Omega_n - \omega_n](0) \leq \liminf_{\rho \downarrow 0} \rho^{-1} \theta_\rho[\Omega_1, \dots, \Omega_n](\bar{x}).$$

**Proof.** The first assertion is a direct consequence of Proposition 5.2. The second assertion is a consequence of the first one since  $\bar{x} \in \text{bd} \bigcap_{i=1}^n \Omega_i$  implies that  $\theta_\rho[\Omega_1, \dots, \Omega_n](\bar{x}) \rightarrow 0$  as  $\rho \downarrow 0$ ; cf. [45, Proposition 3].  $\square$

The next proposition presents a metric counterpart of the conditions (P2). It contains the key ingredients of the metric criteria of approximate stationarity and transversality in Theorems 2.1 and 3.1, respectively.

**Proposition 5.3** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\varepsilon > 0$ ,  $\rho > 0$  and  $\alpha := \frac{\varepsilon}{\rho}$ . Then

- (i) conditions (P2) imply conditions (P9);
- (ii) if conditions (P9) are satisfied, then there exist a number  $\rho' \in ]0, \rho[$ , points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $a'_i \in X$  ( $i = 1, \dots, n$ ) such that conditions (P7) are satisfied with  $\rho'$  and  $a'_i$  in place of  $\rho$  and  $a_i$ .

**Proof.** (i) If conditions (P2) are satisfied, then  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$  and

$$\alpha d \left( \bar{x}, \bigcap_{i=1}^n (\Omega_i - a_i) \right) > \alpha \rho = \varepsilon > \max_{1 \leq i \leq n} \|a_i\| \geq \max_{1 \leq i \leq n} d(\bar{x}, \Omega_i - a_i).$$

(ii) Let conditions (P9) be satisfied. Then

$$\max_{1 \leq i \leq n} d(\bar{x}, \Omega_i - a_i) \leq \max_{1 \leq i \leq n} \|a_i\| < \varepsilon = \alpha \rho,$$

and there exists a number  $\rho' \in ]0, \rho[$  such that

$$\alpha d \left( \bar{x}, \bigcap_{i=1}^n (\Omega_i - a_i) \right) > \alpha \rho' > \max_{1 \leq i \leq n} d(\bar{x}, \Omega_i - a_i).$$

It follows from the first inequality above that  $\bigcap_{i=1}^n (\Omega_i - a_i) \cap B_{\rho'}(\bar{x}) = \emptyset$ , while due to the second inequality, there exist  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) such that  $\max_{1 \leq i \leq n} \|a'_i\| < \alpha \rho'$ , where  $a'_i := a_i + \bar{x} - \omega_i$  ( $i = 1, \dots, n$ ).  $\square$

**Corollary 5.4** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  and  $\alpha > 0$ . The following conditions are equivalent:

- (i) the collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately  $\alpha$ -stationary at  $\bar{x}$ ;
- (ii) for any  $\varepsilon > 0$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) such that conditions (P9) are satisfied;
- (iii) for any  $\varepsilon > 0$ , there exist vectors  $x \in B_\varepsilon(\bar{x})$  and  $a_i \in X$  ( $i = 1, \dots, n$ ) such that conditions (P10) are satisfied.

**Proof.** (i)  $\Rightarrow$  (ii). Let condition (i) be satisfied and a number  $\varepsilon > 0$  be given. Set  $\varepsilon' := \varepsilon/(\alpha + 1)$ . Then conditions (P7) hold with some  $\rho \in ]0, \varepsilon'[$ ,  $\omega_i \in \Omega_i \cap B_{\varepsilon'}(\bar{x})$ , and  $a_i \in X$  ( $i = 1, \dots, n$ ). By Proposition 5.3(i), applied to the sets  $\Omega_i - \omega_i$  ( $i = 1, \dots, n$ ) having the common point 0, we have

$$\alpha d \left( 0, \bigcap_{i=1}^n (\Omega_i - \omega_i - a_i) \right) > \max_{1 \leq i \leq n} d(0, \Omega_i - \omega_i - a_i).$$

It follows that the first inequality in (P9) is satisfied with  $a'_i := a_i + \omega_i - \bar{x}$  in place of  $a_i$  ( $i = 1, \dots, n$ ), and  $\max_{1 \leq i \leq n} \|a'_i\| < \alpha\rho + \varepsilon' < (\alpha + 1)\varepsilon' = \varepsilon$ . Hence, condition (ii) is satisfied.

(ii)  $\Rightarrow$  (i). Let condition (ii) be satisfied and a number  $\varepsilon > 0$  be given. Set  $\varepsilon' := \min\{\alpha, 1\}\varepsilon$ , and find vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) such that conditions (P9) are satisfied with  $\varepsilon'$  in place of  $\varepsilon$ . By Proposition 5.3(ii), there are  $\rho \in ]0, \varepsilon'/\alpha[$ ,  $\omega_i \in \Omega_i \cap B_{\varepsilon'}(\bar{x})$ ,  $a'_i \in X$  ( $i = 1, \dots, n$ ) such that conditions (P7) hold with  $a'_i$  in place of  $a_i$ . Since  $\varepsilon' \leq \varepsilon$  and  $\varepsilon'/\alpha \leq \varepsilon$ , condition (i) is satisfied.

(ii)  $\Rightarrow$  (iii) is obvious.

(iii)  $\Rightarrow$  (ii). Let condition (iii) be satisfied and a number  $\varepsilon > 0$  be given. Set  $\varepsilon' := \varepsilon/2$ . Then conditions (P10) hold with some  $x \in B_{\varepsilon'}(\bar{x})$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ) and  $\varepsilon'$  in place of  $\varepsilon$ . Set  $a'_i := a_i + x - \bar{x}$  ( $i = 1, \dots, n$ ). Then conditions (P9) hold true with  $a'_i$  in place of  $a_i$ . Hence, condition (ii) is satisfied.  $\square$

In view of Proposition 3.3, Corollary 5.4 immediately yields Theorems 2.1 and 3.1.

## 6 Dual Characterizations

This section presents a series of ‘generalized separation’ statements, providing dual characterizations of certain typical ‘extremal’ arrangements of collections of sets, discussed in the preceding sections, and traces the relationships between them. These statements contain core arguments, which can be found in various existing versions of the (extended) extremal principle, as well as some new extensions.

The definition of the approximate stationarity (Definition 2.1(iv)) and its conventional dual characterizations in the extended extremal principle (Theorem 2.2(ii) and (iii)) are all formulated “for any  $\varepsilon > 0$  there exist ...”. At the same time, the proofs of the extremal principle and its extensions establish connections between the values of  $\varepsilon$  and other parameters involved in the assumptions and the conclusions. These connections, usually hidden in the proofs, are of importance for more subtle extremality statements. We expose them in the statements below.

All the separation statements in this section are consequences of the next general theorem, providing dual characterizations of the slightly weakened version of the asymmetric extremality property contained in Proposition 3.1(i). It involves the  $d_1$  distance between  $n$  sets defined in (23) and actually combines two statements: for closed sets in general Banach spaces and specifically in Asplund spaces. So far it has been common to formulate (and prove!) such statements separately; cf. [21, 31].

**Theorem 6.1** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n-1$ ), and  $\varepsilon > 0$ . Suppose also that*

$$\bigcap_{i=1}^{n-1} (\Omega_i - a_i) \cap \Omega_n = \emptyset, \quad (34)$$

$$\max_{1 \leq i \leq n-1} \|a_i\| < d_1(\Omega_1 - a_1, \dots, \Omega_{n-1} - a_{n-1}, \Omega_n) + \varepsilon \quad (35)$$

(or simply  $\max_{1 \leq i \leq n-1} \|a_i\| < \varepsilon$ ). Then,

(i) for any  $\lambda > 0$  and  $\rho > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n-1$ ),  $\omega_n \in \Omega_n \cap B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that

$$\sum_{i=1}^n x_i^* = 0, \quad \sum_{i=1}^{n-1} \|x_i^*\| = 1, \quad (36)$$

$$\lambda \sum_{i=1}^{n-1} d(x_i^*, N_{\Omega_i}(\omega_i)) + \rho d(x_n^*, N_{\Omega_n}(\omega_n)) < \varepsilon, \quad (37)$$

$$\sum_{i=1}^{n-1} \langle x_i^*, \omega_n + a_i - \omega_i \rangle = \max_{1 \leq i \leq n-1} \|\omega_n + a_i - \omega_i\|, \quad (38)$$

where  $N$  in (37) stands for the Clarke normal cone ( $N := N^C$ );

(ii) if  $X$  is Asplund, then, for any  $\lambda > 0$ ,  $\rho > 0$  and  $\tau \in ]0, 1[$ , there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n-1$ ),  $\omega_n \in \Omega_n \cap B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (36), (37), where  $N$  in (37) stands for the Fréchet normal cone ( $N := N^F$ ), and

$$\sum_{i=1}^{n-1} \langle x_i^*, \omega_n + a_i - \omega_i \rangle > \tau \max_{1 \leq i \leq n-1} \|\omega_n + a_i - \omega_i\|. \quad (39)$$

**Proof.** Let  $\lambda > 0$  and  $\rho > 0$ . In view of (35), we can choose positive numbers  $\varepsilon_1$  and  $\varepsilon_2$  satisfying

$$\max_{1 \leq i \leq n-1} \|a_i\| - d_1(\Omega_1 - a_1, \dots, \Omega_{n-1} - a_{n-1}, \Omega_n) < \varepsilon_1 < \varepsilon_2 < \varepsilon \quad (40)$$

Note that  $\max_{1 \leq i \leq n-1} \|a_i\| = d_1(\bar{x} - a_1, \dots, \bar{x} - a_{n-1}, \bar{x})$ . We can apply Lemma 5.1 to find points  $\hat{\omega}_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n-1$ ) and  $\hat{\omega}_n \in \Omega_n \cap B_\rho(\bar{x})$  such that

$$\max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\| < d_1(\Omega_1 \cap B_{\alpha\lambda}(\hat{\omega}_1) - a_1, \dots, \Omega_{n-1} \cap B_{\alpha\lambda}(\hat{\omega}_{n-1}) - a_{n-1}, \Omega_n \cap B_{\alpha\rho}(\hat{\omega}_n)) + \alpha\varepsilon_1 \quad (41)$$

for all  $\alpha > 0$ . Consider the three functions  $f_1, f_2, f_3 : X^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ :

$$f_1(u_1, \dots, u_n) := \max_{1 \leq i \leq n-1} \|u_i - a_i - u_n\|, \quad (42)$$

$$f_2(u_1, \dots, u_n) := \varepsilon_2 \max \left\{ \lambda^{-1} \max_{1 \leq i \leq n-1} \|u_i - \hat{\omega}_i\|, \rho^{-1} \|u_n - \hat{\omega}_n\| \right\}, \quad (43)$$

$$f_3(u_1, \dots, u_n) := \begin{cases} 0 & \text{if } u_i \in \Omega_i \text{ (} i = 1, \dots, n\text{),} \\ \infty & \text{otherwise.} \end{cases} \quad (44)$$

Observe that, in view of (34),  $f_1(\hat{\omega}_1, \dots, \hat{\omega}_n) = \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\| > 0$ . Moreover,

$$f_1(u_1, \dots, u_n) - f_1(\hat{\omega}_1, \dots, \hat{\omega}_n) + f_2(u_1, \dots, u_n) \geq 0 \quad \text{for all } u_i \in \Omega_i \text{ (} i = 1, \dots, n\text{)}.$$

Indeed, assume that there are  $u_i \in \Omega_i$  ( $i = 1, \dots, n$ ) such that the inequality does not hold. Then  $(u_1, \dots, u_n) \neq (\hat{\omega}_1, \dots, \hat{\omega}_n)$ , and consequently,  $f_2(u_1, \dots, u_n) > 0$ . Set  $\alpha := f_2(u_1, \dots, u_n)/\varepsilon_1$ . We have  $\|u_i - \hat{\omega}_i\| \leq \frac{\varepsilon_1}{\varepsilon_2} \alpha \lambda < \alpha \lambda$  ( $i = 1, \dots, n-1$ ),  $\|u_n - \hat{\omega}_n\| \leq \frac{\varepsilon_1}{\varepsilon_2} \alpha \rho < \alpha \rho$ , and

$$\max_{1 \leq i \leq n-1} \|u_i - a_i - u_n\| - \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\| + \alpha\varepsilon_1 < 0,$$

which contradicts (41). Thus,  $(\hat{\omega}_1, \dots, \hat{\omega}_n)$  is a point of minimum of the sum  $f_1 + f_2 + f_3$ , and consequently (Lemma 2.2),  $0 \in \partial(f_1 + f_2 + f_3)(\hat{\omega}_1, \dots, \hat{\omega}_n)$ .

Functions  $f_1$  and  $f_2$  are convex and Lipschitz continuous. It is easy to check that the subdifferentials of  $f_1$ ,  $f_2$  and  $f_3$  possess the following properties:

1) The function  $f_1$  (42) is a composition function:  $f_1(u_1, \dots, u_n) = g(A(u_1, \dots, u_n) - (a_1, \dots, a_{n-1}))$ , where  $A$  is the linear operator from  $X^n$  to  $X^{n-1}$ :  $A(u_1, \dots, u_n) := (u_1 - u_n, \dots, u_{n-1} - u_n)$ , and  $g$  is the maximum norm on  $X^{n-1}$ :  $g(u_1, \dots, u_{n-1}) := \max_{1 \leq i \leq n-1} \|u_i\|$ . The corresponding dual norm has the form  $(v_1^*, \dots, v_{n-1}^*) \mapsto \sum_{i=1}^{n-1} \|v_i^*\|$ . It is easy to check that the adjoint operator  $A^* : (X^*)^{n-1} \rightarrow (X^*)^n$  is of the form  $A^*(v_1^*, \dots, v_{n-1}^*) = (v_1^*, \dots, v_{n-1}^*, -\sum_{j=1}^{n-1} v_j^*)$ . When

$f_1(u_1, \dots, u_n) > 0$ , the subdifferential  $\partial g(u_1 - a_1 - u_n, \dots, u_{n-1} - a_{n-1} - u_n)$  is the set of  $(v_{11}^*, \dots, v_{1,n-1}^*) \in (X^*)^{n-1}$  satisfying (see e.g. [2, Corollary 2.4.16])

$$\sum_{i=1}^{n-1} \|v_{1i}^*\| = 1 \quad \text{and} \quad \sum_{i=1}^{n-1} \langle v_{1i}^*, u_i - a_i - u_n \rangle = \max_{1 \leq i \leq n-1} \|u_i - a_i - u_n\|. \quad (45)$$

Thus, in view of the convex chain rule (see e.g. [2, Theorem 2.8.3]), if  $f_1(u_1, \dots, u_n) > 0$ , then the subdifferential  $\partial f_1(u_1, \dots, u_n)$  is the set of all vectors  $(v_{11}^*, \dots, v_{1,n-1}^*, -\sum_{j=1}^{n-1} v_{1j}^*) \in (X^*)^n$ , where vectors  $v_{1i}^* \in X^*$  ( $i = 1, \dots, n-1$ ) satisfy (45).

2) The function  $f_2$  (43) is a positive multiple of the norm on  $X^n$  (translated by  $(\hat{\omega}_1, \dots, \hat{\omega}_n)$ ). Its subgradients  $(v_{21}^*, \dots, v_{2n}^*)$  at any point satisfy

$$\lambda \sum_{i=1}^{n-1} \|v_{2i}^*\| + \rho \|v_{2n}^*\| \leq \varepsilon_2. \quad (46)$$

3) The function  $f_3$  (44) is the indicator function of the set  $\Omega_1 \times \dots \times \Omega_n$ . Its subdifferential has a simple representation:  $\partial f_3(u_1, \dots, u_n) = \prod_{i=1}^n N_{\Omega_i}(u_i)$  for all  $u_i \in \Omega_i$  ( $i = 1, \dots, n$ ) (Lemma 2.3).

From this point the proof splits into two cases.

(i) We apply the Clarke–Rockafellar subdifferential sum rule (Lemma 2.1(iv)) to find elements of the three subdifferentials:  $(v_{11}^*, \dots, v_{1,n-1}^*) \in \partial g(\hat{\omega}_1 - a_1 - \hat{\omega}_n, \dots, \hat{\omega}_{n-1} - a_n - \hat{\omega}_n)$ ,  $(v_{21}^*, \dots, v_{2n}^*) \in \partial f_2(\hat{\omega}_1, \dots, \hat{\omega}_n)$  and  $(v_{31}^*, \dots, v_{3n}^*) \in \partial f_3(\hat{\omega}_1, \dots, \hat{\omega}_n)$  such that

$$v_{1i}^* + v_{2i}^* + v_{3i}^* = 0 \quad (i = 1, \dots, n-1) \quad \text{and} \quad -\left(\sum_{j=1}^{n-1} v_{1j}^*\right) + v_{2n}^* + v_{3n}^* = 0.$$

Then  $v_{3i}^* \in N_{\Omega_i}(\hat{\omega}_i)$  ( $i = 1, \dots, n$ ) and conditions (45) and (46) are satisfied. The conclusion of part (i) of the theorem holds true with  $\hat{\omega}_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ),  $x_i^* := -v_{1i}^*$  ( $i = 1, \dots, n-1$ ) and  $x_n^* := \sum_{j=1}^{n-1} v_{1j}^*$ .

(ii) Let  $X$  be Asplund and  $\tau \in ]0, 1[$ . We can apply the fuzzy sum rule (Lemma 2.1(iii)) to the sum  $(f_1 + f_2) + f_3$  followed by the conventional convex sum rule (Lemma 2.1(i)) applied to  $f_1 + f_2$ . Choose a  $\xi > 0$  satisfying the following conditions:

$$\xi < \varepsilon - \varepsilon_2, \quad \xi < \lambda - \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - \bar{x}\|, \quad \xi < \rho - \|\hat{\omega}_n - \bar{x}\|, \quad (47)$$

$$(10 - 2\tau)\xi < (1 - \tau) \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\|. \quad (48)$$

Since  $\tau \in ]0, 1[$ , the last inequality implies in particular that

$$2\xi < \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\|. \quad (49)$$

Applying the fuzzy sum rule, we find two points  $(x_1, \dots, x_n)$ ,  $(\omega_1, \dots, \omega_n) \in X^n$  such that  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) and

$$\max_{1 \leq i \leq n} \|x_i - \hat{\omega}_i\| < \xi, \quad \max_{1 \leq i \leq n} \|\omega_i - \hat{\omega}_i\| < \xi, \quad (50)$$

and elements of the three subdifferentials:  $(v_{11}^*, \dots, v_{1,n-1}^*) \in \partial g(x_1 - a_1 - x_n, \dots, x_{n-1} - a_n - x_n)$ ,  $(v_{21}^*, \dots, v_{2n}^*) \in \partial f_2(x_1, \dots, x_n)$  and  $(v_{31}^*, \dots, v_{3n}^*) \in \partial f_3(\omega_1, \dots, \omega_n)$  such that

$$\sum_{i=1}^{n-1} \|v_{1i}^* + v_{2i}^* + v_{3i}^*\| + \left\| -\left(\sum_{j=1}^{n-1} v_{1j}^*\right) + v_{2n}^* + v_{3n}^* \right\| < \frac{\xi}{\max\{\lambda, \rho\}}.$$

The last condition implies

$$\lambda \sum_{i=1}^{n-1} \|v_{1i}^* + v_{2i}^* + v_{3i}^*\| + \rho \left\| -\left(\sum_{j=1}^{n-1} v_{1j}^*\right) + v_{2n}^* + v_{3n}^* \right\| < \xi. \quad (51)$$

In view of (47), (48), (49) and (50), we have the following estimates:

$$\|\omega_i - \bar{x}\| \leq \|\hat{\omega}_i - \bar{x}\| + \|\omega_i - \hat{\omega}_i\| < \lambda \quad (i = 1, \dots, n-1), \quad \|\omega_n - \bar{x}\| < \|\hat{\omega}_n - \bar{x}\| + \|\omega_n - \hat{\omega}_n\| < \rho,$$

(hence,  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n-1$ ) and  $\omega_n \in \Omega_n \cap B_\rho(\bar{x})$ )

$$\begin{aligned}
\|\omega_i - x_i\| &\leq \|\omega_i - \hat{\omega}_i\| + \|x_i - \hat{\omega}_i\| < 2\xi \quad (i = 1, \dots, n), \\
\max_{1 \leq i \leq n-1} \|x_i - a_i - x_n\| &\geq \max_{1 \leq i \leq n-1} (\|\hat{\omega}_i - a_i - \hat{\omega}_n\| - \|x_i - \hat{\omega}_i\| - \|x_n - \hat{\omega}_n\|) \\
&> \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\| - 2\xi > 0, \\
\max_{1 \leq i \leq n-1} \|\omega_i - a_i - \omega_n\| &\geq \max_{1 \leq i \leq n-1} (\|\hat{\omega}_i - a_i - \hat{\omega}_n\| - \|\omega_i - \hat{\omega}_i\| - \|\omega_n - \hat{\omega}_n\|) \\
&> \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\| - 2\xi > 0, \\
(1 - \tau) \max_{1 \leq i \leq n-1} \|\omega_i - a_i - \omega_n\| &> (1 - \tau) \left( \max_{1 \leq i \leq n-1} \|\hat{\omega}_i - a_i - \hat{\omega}_n\| - 2\xi \right) \\
&> (10 - 2\tau)\xi - 2(1 - \tau)\xi = 8\xi,
\end{aligned} \tag{52}$$

conditions (45) and (46) are satisfied and  $v_{3i}^* \in N_{\Omega_i}(\omega_i)$  ( $i = 1, \dots, n$ ). Denote  $x_i^* := -v_{1i}^*$  ( $i = 1, \dots, n-1$ ) and  $x_n^* := \sum_{j=1}^{n-1} v_{1j}^*$ . Then  $\sum_{i=1}^n x_i^* = 0$  and  $\sum_{i=1}^n \|x_i^*\| = 1$ . Using (45) and (52), we obtain the following inequalities:

$$\begin{aligned}
\sum_{i=1}^{n-1} \langle x_i^*, \omega_n + a_i - \omega_i \rangle &\geq \sum_{i=1}^{n-1} \langle v_{1i}^*, x_i - a_i - x_n \rangle - \sum_{i=1}^{n-1} \|v_{1i}^*\| (\|\omega_i - x_i\| + \|\omega_n - x_n\|) \\
&> \max_{1 \leq i \leq n-1} \|x_i - a_i - x_n\| - 4\xi \\
&\geq \max_{1 \leq i \leq n-1} (\|\omega_i - a_i - \omega_n\| - \|\omega_i - x_i\| - \|\omega_n - x_n\|) - 4\xi \\
&> \max_{1 \leq i \leq n-1} \|\omega_i - a_i - \omega_n\| - 8\xi.
\end{aligned} \tag{54}$$

Adding (53) and (54), we arrive at (39). Making use of (46), (51) and (47), we obtain the following estimates:

$$\lambda \sum_{i=1}^{n-1} \|x_i^* - v_{3i}^*\| + \rho \|x_n^* - v_{3n}^*\| < \lambda \sum_{i=1}^{n-1} \|v_{2i}^*\| + \rho \|v_{2n}^*\| + \xi < \varepsilon_2 + \xi < \varepsilon.$$

The last inequality yields (37).  $\square$

**Remark 6.1** (i) Conditions (34) and (35) are implied by conditions (P4). Hence, in view of Proposition 3.1(i), Theorem 6.1 provides dual necessary characterizations of extremality.

- (ii) Conditions (38) and (39) relate the dual vectors  $x_i^*$  and the primal space vectors  $\omega_n + a_i - \omega_i$  ( $i = 1, \dots, n-1$ ). Such conditions, though not common in the conventional formulations of the extremal/generalized separation statements, seem to provide important additional characterizations of the properties. Conditions of this kind first appeared explicitly in the generalized separation theorems in [31], where the authors also provided motivations for employing such conditions. As one can see from the proof above, conditions (38) and (39) originate in computing the convex subdifferential of the norm in  $X^{n-1}$  at a nonzero point; see (45). Subdifferentiating a norm (in either  $X^{n-1}$  or  $X^n$ ) at a nonzero point is a necessary step in the proofs of all existing versions of the extremal principle and its extensions, starting with the very first one in [12, Theorem 6.1], with conditions like (45) hidden in the proofs. In several statements in the rest of this section, following [31], we make such conditions exposed.
- (iii) Lemma 5.1 substitutes in the proof of Theorem 6.1 the conventional Ekeland variational principle.
- (iv) If condition (35) in Theorem 6.1 is replaced by a stronger one:

$$\max_{1 \leq i \leq n-1} \|a_i\| = d_1(\Omega_1 - a_1, \dots, \Omega_{n-1} - a_{n-1}, \Omega_n),$$

which means that the infimum of  $\max_{1 \leq i \leq n-1} \|\omega_i - a_i - \omega_n\|$  over  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) is attained at  $\omega_1 = \dots = \omega_n = \bar{x}$ , then the application of Lemma 5.1 in the proof can be dropped, leading to an improvement in part (i): conditions (37) and (38) can be replaced by  $x_i^* \in N_{\Omega_i}^C(\bar{x})$  ( $i = 1, \dots, n$ ) and  $\sum_{i=1}^{n-1} \langle x_i^*, a_i \rangle = \max_{1 \leq i \leq n-1} \|a_i\|$ , respectively. A similar fact was observed in [31, Theorem 3.1'].

The assumption  $\cap_{i=1}^n \Omega_i \neq \emptyset$  in Theorem 6.1 is not restrictive. The common point  $\bar{x} \in \cap_{i=1}^n \Omega_i$  of the collection of sets can be replaced by a collection of individual points  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ). The next statement provides dual characterizations of the slightly weakened version of the asymmetric extremality property contained in Proposition 4.3.

**Corollary 6.1** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ),  $a_i \in X$  ( $i = 1, \dots, n-1$ ), and  $\varepsilon > 0$ . Suppose also that

$$\bigcap_{i=1}^{n-1} (\Omega_i - \omega_i - a_i) \cap (\Omega_n - \omega_n) = \emptyset, \quad (55)$$

$$\max_{1 \leq i \leq n-1} \|a_i\| < d_1(\Omega_1 - \omega_1 - a_1, \dots, \Omega_{n-1} - \omega_{n-1} - a_{n-1}, \Omega_n - \omega_n) + \varepsilon \quad (56)$$

(or simply  $\max_{1 \leq i \leq n-1} \|a_i\| < \varepsilon$ ). Then,

(i) for any  $\lambda > 0$  and  $\rho > 0$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n-1$ ),  $\omega'_n \in \Omega_n \cap B_\rho(\omega_n)$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (36) and (37) with  $N := N^C$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), and

$$\sum_{i=1}^{n-1} \langle x_i^*, \omega'_n - \omega_n + a_i - \omega'_i + \omega_i \rangle = \max_{1 \leq i \leq n-1} \|\omega'_n - \omega_n + a_i - \omega'_i + \omega_i\|;$$

(ii) if  $X$  is Asplund, then, for any  $\lambda > 0$ ,  $\rho > 0$  and  $\tau \in ]0, 1[$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n-1$ ),  $\omega'_n \in \Omega_n \cap B_\rho(\omega_n)$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (36) and (37) with  $N := N^F$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), and

$$\sum_{i=1}^{n-1} \langle x_i^*, \omega'_n - \omega_n + a_i - \omega'_i + \omega_i \rangle > \tau \max_{1 \leq i \leq n-1} \|\omega'_n - \omega_n + a_i - \omega'_i + \omega_i\|.$$

**Proof.** The sets  $\Omega'_i := \Omega_i - \omega_i$  ( $i = 1, \dots, n$ ) satisfy  $0 \in \bigcap_{i=1}^n \Omega'_i$  and conditions (34) and (35). The conclusion follows from Theorem 6.1 after noticing that  $N_{\Omega'_i}(\omega'_i - \omega_i) = N_{\Omega_i}(\omega'_i)$  ( $i = 1, \dots, n$ ).  $\square$

Theorem 6.1 is a particular case of Corollary 6.1 with  $\omega_i = \bar{x}$  ( $i = 1, \dots, n$ ).

The following theorem is an immediate consequence of Theorem 6.1. It combines two *unified separation theorems* due to Zheng and Ng [31, Theorems 3.1 and 3.4].

**Theorem 6.2** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bigcap_{i=1}^n \Omega_i = \emptyset$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ),  $\varepsilon > 0$  and condition (27) is satisfied. Then,

(i) for any  $\lambda > 0$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (D6) with  $N$  standing for the Clarke normal cone ( $N := N^C$ ),  $\alpha := \frac{\varepsilon}{\lambda}$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), and

$$\sum_{i=1}^{n-1} \langle x_i^*, \omega'_n - \omega'_i \rangle = \max_{1 \leq i \leq n-1} \|\omega'_n - \omega'_i\|; \quad (57)$$

(ii) if  $X$  is Asplund, then, for any numbers  $\lambda > 0$  and  $\tau \in ]0, 1[$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (D6) with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ),  $\alpha := \frac{\varepsilon}{\lambda}$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), and

$$\sum_{i=1}^{n-1} \langle x_i^*, \omega'_n - \omega'_i \rangle > \tau \max_{1 \leq i \leq n-1} \|\omega'_n - \omega'_i\|. \quad (58)$$

**Proof.** Observe that the sets  $\Omega'_i := \Omega_i - \omega_i$  ( $i = 1, \dots, n$ ) and vectors  $a_i := \omega_n - \omega_i$  ( $i = 1, \dots, n-1$ ) satisfy  $0 \in \bigcap_{i=1}^n \Omega'_i$  and

$$\bigcap_{i=1}^{n-1} (\Omega'_i - a_i) \cap \Omega'_n = \bigcap_{i=1}^n (\Omega_i - \omega_n) = \bigcap_{i=1}^n \Omega_i - \omega_n = \emptyset,$$

$$\max_{1 \leq i \leq n-1} \|a_i\| = d_1(\omega_1, \dots, \omega_n) < d_1(\Omega_1, \dots, \Omega_n) + \varepsilon = d_1(\Omega'_1 - a_1, \dots, \Omega'_{n-1} - a_{n-1}, \Omega'_n) + \varepsilon.$$

Applying Theorem 6.1 with  $\rho = \lambda$ , we arrive at the conclusions.  $\square$

**Remark 6.2** (i) In [31], instead of the  $d_1$  distance in condition (27), a slightly more general  $p$ -weighted nonintersect index was used with the corresponding  $q$ -weighted sums replacing the usual ones in (D6), (57) and (58). This corresponds to considering  $l_p$  norms on product spaces and the corresponding  $l_q$  dual norms. In the current paper, for simplicity only the maximum norm on product spaces is considered together with the corresponding sum norm in the dual space; cf. Remark 4.2(ii.)

- (ii) Theorem 6.2 is a consequence of Theorem 6.1, which in turn is a consequence of the Ekeland variational principle. Thanks to [30, Theorem 3.1], part (i) of Theorem 6.2 is equivalent to the Ekeland variational principle. Hence, the conclusion of Theorem 6.1 is also equivalent to the Ekeland variational principle (and to completeness of the space  $X$ ).

The next theorem is a ‘symmetric’ version of Theorem 6.1.

**Theorem 6.3** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\rho > 0$  and  $\varepsilon > 0$ . Suppose also that*

$$\bigcap_{i=1}^n (\Omega_i - a_i) \cap B_\rho(\bar{x}) = \emptyset, \quad (59)$$

$$\max_{1 \leq i \leq n} \|a_i\| < d_{B_\rho(\bar{x})}(\Omega_1 - a_1, \dots, \Omega_n - a_n) + \varepsilon \quad (60)$$

(or simply  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ ). Then,

- (i) *for any  $\lambda > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that*

$$\lambda \sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) + \rho \left\| \sum_{i=1}^n x_i^* \right\| < \varepsilon, \quad \sum_{i=1}^n \|x_i^*\| = 1, \quad (61)$$

$$\sum_{i=1}^n \langle x_i^*, x + a_i - \omega_i \rangle = \max_{1 \leq i \leq n} \|x + a_i - \omega_i\|, \quad (62)$$

where  $N$  in (61) stands for the Clarke normal cone ( $N := N^C$ );

- (ii) *if  $X$  is Asplund, then, for any  $\lambda > 0$  and  $\tau \in ]0, 1[$ , there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (61), where  $N$  stands for the Fréchet normal cone ( $N := N^F$ ), and*

$$\sum_{i=1}^n \langle x_i^*, x + a_i - \omega_i \rangle > \tau \max_{1 \leq i \leq n} \|x + a_i - \omega_i\|. \quad (63)$$

**Proof.** Choose an  $\varepsilon' \in ]0, \varepsilon[$  such that condition (60) holds true with  $\varepsilon'$  in place of  $\varepsilon$ , and then choose a  $\rho' \in ]0, \rho[$  such that  $\rho - \rho' < \varepsilon - \varepsilon'$ . It is sufficient to apply Theorem 6.1 to the collection of  $n + 1$  closed sets  $\Omega_1, \dots, \Omega_n$  and  $\Omega_{n+1} := \bar{B}_{\rho'}(\bar{x})$  with  $\varepsilon'$  and  $\rho'$  in place of  $\varepsilon$  and  $\rho$ , respectively. Notice that  $\Omega_{n+1} \cap B_{\rho'}(\bar{x}) = B_{\rho'}(\bar{x})$  and  $N_{\Omega_{n+1}(\bar{x})}(x) = 0$  for any  $x \in B_{\rho'}(x)$ . One only needs to check the inequality in (61), which is straightforward:

$$\lambda \sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) + \rho \left\| \sum_{i=1}^n x_i^* \right\| < \varepsilon' + (\rho - \rho') \left\| \sum_{i=1}^n x_i^* \right\| \leq \varepsilon' + (\rho - \rho') < \varepsilon.$$

The proof is complete.  $\square$

**Remark 6.3** (i) Conditions (59) and (60) are implied by conditions (P2). Hence, Theorem 6.3 provides dual necessary characterizations of the local extremality.

- (ii) The inequality in (61) combines two constraints on the vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ): they must be close to the respective normal cones and their sum must be small. This inequality obviously implies two separate inequalities:

$$\sum_{i=1}^n d(x_i^*, N_{\Omega_i}(\omega_i)) < \frac{\varepsilon}{\lambda} \quad \text{and} \quad \left\| \sum_{i=1}^n x_i^* \right\| < \frac{\varepsilon}{\rho},$$

while the converse implication is not true in general. Both these constraints are involved in each of the two generalized separation conditions in the (extended) extremal principle discussed in Section 2 (parts (ii) and (iii) of Theorem 2.2). However, each of the generalized separation conditions in Theorem 2.2 requires actually a stronger version of one of the constraints: either the vectors must belong to the respective normal cones (part (ii) of Theorem 2.2) or their sum must be exactly zero (part (iii) of Theorem 2.2). Fortunately, as the next two corollaries show, the required stronger versions of (one of) the constraints are consequences of the combined condition (61) and Lemma 2.4.

**Corollary 6.2** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\varepsilon > 0$ ,  $\lambda > 0$ ,  $\varepsilon + \lambda \leq \rho$ , and conditions (59) and (60) are satisfied (the latter condition can be replaced by the simpler and stronger one:  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ ). Then,*

- (i) there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ), satisfying conditions (D4) with  $N$  standing for the Clarke normal cone ( $N := N^C$ ) and  $\alpha := \frac{\varepsilon}{\lambda}$ , and condition (63) with  $\tau := \frac{\rho - \varepsilon}{\rho + \varepsilon}$ ;
- (ii) if  $X$  is Asplund, then, for any  $\tau \in ]0, \frac{\rho - \varepsilon}{\rho + \varepsilon}[$ , there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying condition (D4) with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ) and  $\alpha := \frac{\varepsilon}{\lambda}$ , and condition (63).

**Proof.** Set  $\tau' := \tau(1 + \frac{\varepsilon}{\rho}) + \frac{\varepsilon}{\rho}$ , and observe that  $\tau' = 1$  in part (i), and  $\tau' \in ]0, 1[$  in part (ii). Applying Theorem 6.1, we find points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that condition (61) with the respective normal cone is satisfied and

$$\sum_{i=1}^n \langle x_i^*, x + a_i - \omega_i \rangle \geq \tau' \max_{1 \leq i \leq n} \|x + a_i - \omega_i\|. \quad (64)$$

The assertion follows from Lemma 2.4(i) and (iii) after noticing that  $\frac{\tau' \rho - \varepsilon}{\rho + \varepsilon} = \tau$ .  $\square$

**Corollary 6.3** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\varepsilon > 0$ ,  $\rho > 0$ ,  $\varepsilon + \rho \leq \lambda$ , and conditions (59) and (60) are satisfied (the latter condition can be replaced by the simpler and stronger one:  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ ). Then,

- (i) there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ), satisfying conditions (D3) with  $N$  standing for the Clarke normal cone ( $N := N^C$ ) and  $\alpha := \frac{\varepsilon}{\rho}$ , and condition (63) with  $\tau := \frac{\lambda - \varepsilon}{\lambda + \varepsilon}$ ;
- (ii) if  $X$  is Asplund, then, for any  $\tau \in ]0, \frac{\lambda - \varepsilon}{\lambda + \varepsilon}[$ , there exist points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (D3) with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ) and  $\alpha := \frac{\varepsilon}{\rho}$ , and condition (63).

**Proof.** Set  $\tau' := \tau(1 + \frac{\varepsilon}{\lambda}) + \frac{\varepsilon}{\lambda}$ , and observe that  $\tau' = 1$  in part (i), and  $\tau' \in ]0, 1[$  in part (ii). Applying Theorem 6.1, we find points  $\omega_i \in \Omega_i \cap B_\lambda(\bar{x})$  ( $i = 1, \dots, n$ ) and  $x \in B_\rho(\bar{x})$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (61) (with the respective normal cone) and (64) are satisfied. The assertion follows from Lemma 2.4(ii) and (iii) after noticing that  $\frac{\tau' \lambda - \varepsilon}{\lambda + \varepsilon} = \tau$ .  $\square$

The versions of Theorem 6.3 and Corollaries 6.2 and 6.3 with a common point  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  of a collection of sets replaced by a collection of individual points  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ), presented in the next three corollaries, follow immediately.

**Corollary 6.4** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\omega_i \in \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\rho > 0$  and  $\varepsilon > 0$ . Suppose also that

$$\bigcap_{i=1}^n (\Omega_i - \omega_i - a_i) \cap (\rho \mathbb{B}) = \emptyset, \quad (65)$$

$$\max_{1 \leq i \leq n} \|a_i\| < d_{\rho \mathbb{B}}(\Omega_1 - \omega_1 - a_1, \dots, \Omega_n - \omega_n - a_n) + \varepsilon \quad (66)$$

(or simply  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ ). Then,

- (i) for any  $\lambda > 0$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) and  $x \in \rho \mathbb{B}$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (61) with  $N$  standing for the Clarke normal cone ( $N := N^C$ ) and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), and

$$\sum_{i=1}^n \langle x_i^*, x + a_i - \omega'_i + \omega_i \rangle = \max_{1 \leq i \leq n} \|x + a_i - \omega'_i + \omega_i\|;$$

- (ii) if  $X$  is Asplund, then, for any  $\lambda > 0$  and  $\tau \in ]0, 1[$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) and  $x \in \rho \mathbb{B}$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (61) with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ) and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), and

$$\sum_{i=1}^n \langle x_i^*, x + a_i - \omega'_i + \omega_i \rangle > \tau \max_{1 \leq i \leq n} \|x + a_i - \omega'_i + \omega_i\|. \quad (67)$$

**Corollary 6.5** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\omega_i \in \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\varepsilon > 0$ ,  $\lambda > 0$ ,  $\varepsilon + \lambda \leq \rho$ , and conditions (65) and (66) are satisfied (the latter condition can be replaced by the simpler and stronger one:  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ ). Then,

- (i) there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) and  $x \in \rho\mathbb{B}$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D4) are satisfied with  $N$  standing for the Clarke normal cone ( $N := N^C$ ),  $\alpha := \frac{\varepsilon}{\lambda}$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), as well as condition (67) with  $\tau := \frac{\rho - \varepsilon}{\rho + \varepsilon}$ ;
- (ii) if  $X$  is Asplund, then, for any  $\tau \in ]0, \frac{\rho - \varepsilon}{\rho + \varepsilon}[$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) and  $x \in \rho\mathbb{B}$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D4) are satisfied with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ),  $\alpha := \frac{\varepsilon}{\lambda}$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), as well as condition (67).

**Corollary 6.6** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\omega_i \in \Omega_i$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\varepsilon > 0$ ,  $\rho > 0$ ,  $\varepsilon + \rho \leq \lambda$ , and conditions (65) and (66) are satisfied (the latter condition can be replaced by the simpler and stronger one:  $\max_{1 \leq i \leq n} \|a_i\| < \varepsilon$ ). Then,

- (i) there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) and  $x \in \rho\mathbb{B}$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D3) are satisfied with  $N$  standing for the Clarke normal cone ( $N := N^C$ ),  $\alpha := \frac{\varepsilon}{\rho}$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), as well as condition (67) with  $\tau := \frac{\lambda - \varepsilon}{\lambda + \varepsilon}$ ;
- (ii) if  $X$  is Asplund, then, for any  $\tau \in ]0, \frac{\lambda - \varepsilon}{\lambda + \varepsilon}[$ , there exist points  $\omega'_i \in \Omega_i \cap B_\lambda(\omega_i)$  ( $i = 1, \dots, n$ ) and  $x \in \rho\mathbb{B}$ , and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D3) are satisfied with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ),  $\alpha := \frac{\varepsilon}{\rho}$  and  $\omega'_i$  in place of  $\omega_i$  ( $i = 1, \dots, n$ ), as well as condition (67).

**Remark 6.4** In the above three corollaries, the assumption of the existence of a common point  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  of a collection of sets, used in Theorem 6.3 and Corollaries 6.2 and 6.3, is relaxed to that of a collection of individual points  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) (which always exist as long as all the sets are nonempty). On the other hand, if such a point  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  exists, it can be used along with the collection  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) to provide additional useful estimates. Indeed, if  $\|\omega_i - \bar{x}\| \leq \xi$  ( $i = 1, \dots, n$ ) for some  $\xi \geq 0$  (one can take, e.g.,  $\xi := \max_{1 \leq i \leq n} \|\omega_i - \bar{x}\|$ ), then each of the above three corollaries immediately gives  $\|\omega'_i - \bar{x}\| < \xi + \lambda$  ( $i = 1, \dots, n$ ). This simple observation plays an important role in the proof of the extended extremal principle. It is used also in the next statement, which is a consequence of Corollaries 6.5 and 6.6.

**Proposition 6.1** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$ ,  $\xi \geq 0$ ,  $\omega_i \in \Omega_i \cap B_\xi(\bar{x})$ ,  $a_i \in X$  ( $i = 1, \dots, n$ ),  $\alpha \in ]0, 1[$ ,  $\rho > 0$ ,  $\xi + \rho(1 - \alpha) < \delta$ , and conditions (P7) are satisfied. Then,

- (i) there exist points  $\omega'_i \in B_\delta(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D3) are satisfied with  $N$  standing for the Clarke normal cone ( $N := N^C$ ) and  $\omega'_i$  in place of  $\omega_i$ , as well as condition (67) with  $\tau := \frac{1}{1 + 2\alpha}$ ;
- (ii) there exist points  $\omega'_i \in B_\delta(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D4) are satisfied with  $N$  standing for the Clarke normal cone ( $N := N^C$ ), and  $\alpha' := \frac{\alpha}{1 - \alpha}$  and  $\omega'_i$  in place of  $\alpha$  and  $\omega_i$ , respectively, as well as condition (67) with  $\tau := \frac{1 - \alpha}{1 + \alpha}$ .

Suppose  $X$  is Asplund and  $N$  stands for the Fréchet normal cone ( $N := N^F$ ). Then,

- (iii) for any  $\tau \in ]0, \frac{1}{1 + 2\alpha}[$ , there exist points  $\omega'_i \in B_\delta(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D3) are satisfied with  $\omega'_i$  in place of  $\omega_i$ , as well as condition (67);
- (iv) for any  $\tau \in ]0, \frac{1 - \alpha}{1 + \alpha}[$ , there exist points  $\omega'_i \in B_\delta(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D4) are satisfied with  $\alpha' := \frac{\alpha}{1 - \alpha}$  and  $\omega'_i$  in place of  $\alpha$  and  $\omega_i$ , respectively, as well as condition (67).

**Proof.** In view of Remark 6.4, assertions (i) and (iii) are consequences of Corollary 6.5 with  $\varepsilon := \alpha\rho$  and  $\lambda := (1 - \alpha)\rho$ , while assertions (ii) and (iv) are consequences of Corollary 6.6 with  $\varepsilon := \alpha\rho$  and  $\lambda := (1 + \alpha)\rho$ .  $\square$

**Remark 6.5** The assertion in Proposition 6.1(iii) improves [22, Theorem 3.1], which was used in [22] as the main tool when extending the extremal principle to infinite collections of sets.

The above proposition yields dual characterizations of approximate  $\alpha$ -stationarity (Definition 3.1).

**Corollary 6.7** Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \bigcap_{i=1}^n \Omega_i$  and  $\alpha \in ]0, 1[$ . Suppose also that the collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately  $\alpha$ -stationary at  $\bar{x}$ . Then,

- (i) for any  $\varepsilon > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D3) are satisfied with  $N$  standing for the Clarke normal cone ( $N := N^C$ );
- (ii) for any  $\varepsilon > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D4) are satisfied with  $N$  standing for the Clarke normal cone ( $N := N^C$ ), and  $\alpha' := \frac{\alpha}{1 - \alpha}$  in place of  $\alpha$ .

Suppose  $X$  is Asplund and  $N$  stands for the Fréchet normal cone ( $N := N^F$ ). Then,

- (iii) for any  $\varepsilon > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D3) are satisfied;
- (iv) for any  $\varepsilon > 0$ , there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D4) are satisfied with  $\alpha' := \frac{\alpha}{1-\alpha}$  in place of  $\alpha$ .

**Proof.** In each of the assertions, let a number  $\varepsilon > 0$  be given. For parts (i) and (iii), choose a number  $\xi \in ]0, \frac{\varepsilon}{2+\alpha}[$ , while for parts (ii) and (iv), choose a number  $\xi \in ]0, \frac{\varepsilon}{2-\alpha}[$ . By Definition 3.1, there exist a number  $\rho \in ]0, \xi[$ , points  $\omega_i \in \Omega_i \cap B_\xi(\bar{x})$  and vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) such that conditions (P7) are satisfied. In parts (i) and (iii), we have  $\xi + \rho(1 + \alpha) < \xi(2 + \alpha) < \varepsilon$ , and in parts (ii) and (iv), we have  $\xi + \rho(1 - \alpha) < \xi(2 - \alpha) < \varepsilon$ . The conclusions follow from Proposition 6.1.  $\square$

**Remark 6.6** The infinitesimal statements in Corollary 6.7 are crucial for the extended extremal principle and its extensions to infinite collections of sets. For instance, in view of Proposition 3.3, Corollary 6.7(iii) and (iv) immediately yield the implications, respectively, (i)  $\Rightarrow$  (ii) and (i)  $\Rightarrow$  (iii) in Theorem 2.2.

The dual characterizations of the ‘extremal’ arrangements of collections of sets given in the statements in the first part of this section are themselves in a sense extremal properties of collections of sets. They can be partially reversed in the setting of a general normed vector space and with Fréchet normal cones. We start with an ‘asymmetric’ statement where the last set on the list plays a special role.

**Proposition 6.2** Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) and  $\varepsilon > 0$ . If vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfy

$$x_i^* \in N_{\Omega_i}^F(\omega_i) \quad (i = 1, \dots, n-1), \quad d(x_n^*, N_{\Omega_n}^F(\omega_n)) < \varepsilon \quad (68)$$

and conditions (36), then there exists a  $\delta > 0$  such that, for any  $\rho \in ]0, \delta[$  and  $\tau \in ]0, 1[$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying conditions (P6) and

$$\sum_{i=1}^{n-1} \langle x_i^*, a_i \rangle > \tau \varepsilon \rho. \quad (69)$$

**Proof.** By (68), we can choose a vector  $x^* \in N_{\Omega_n}^F(\omega_n)$  and a positive number  $\varepsilon' < \varepsilon$  such that

$$\|x_n^* - x^*\| < \varepsilon'. \quad (70)$$

Choose also numbers  $\varepsilon_1 > 0$  and  $\varepsilon_2 > 0$  such that

$$n\varepsilon_1 + (n-1)\varepsilon_2 < \varepsilon - \varepsilon' \quad \text{and} \quad (n-1)\varepsilon_2 < \varepsilon(1 - \tau). \quad (71)$$

By the definition of the Fréchet normal cone, there exists a number  $\delta > 0$  such that

$$\langle x_i^*, \omega - \omega_i \rangle \leq \frac{\varepsilon_1}{\varepsilon + 1} \|\omega - \omega_i\| \quad \text{for all} \quad \omega \in \Omega_i \cap B_{(\varepsilon+1)\delta}(\omega_i) \quad (i = 1, \dots, n-1), \quad (72)$$

$$\langle x^*, \omega - \omega_n \rangle \leq \varepsilon_1 \|\omega - \omega_n\| \quad \text{for all} \quad \omega \in \Omega_n \cap B_\delta(\omega_n). \quad (73)$$

Let  $\rho \in ]0, \delta[$ . Choose vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) such that

$$\|a_i\| < \varepsilon \rho \quad \text{and} \quad \langle x_i^*, a_i \rangle > \varepsilon \rho \|x_i^*\| - \varepsilon_2 \rho \quad (i = 1, \dots, n-1). \quad (74)$$

Then, by (36) and (71),

$$\sum_{i=1}^{n-1} \langle x_i^*, a_i \rangle > \varepsilon \rho - (n-1)\varepsilon_2 \rho > \tau \varepsilon \rho.$$

Hence, the inequality in (P6) and condition (69) are satisfied. Suppose that the equality in (P6) does not hold. Then there exist  $\omega'_i \in \Omega_i$  ( $i = 1, \dots, n$ ) and an  $x \in \rho \mathbb{B}$  such that

$$\omega'_1 - \omega_1 - a_1 = \dots = \omega'_{n-1} - \omega_{n-1} - a_{n-1} = \omega'_n - \omega_n = x.$$

By (74),  $\|\omega'_i - \omega_i\| = \|x + a_i\| \leq \|x\| + \|a_i\| < (\varepsilon + 1)\rho < (\varepsilon + 1)\delta$  ( $i = 1, \dots, n-1$ ) and  $\|\omega'_n - \omega_n\| = \|x\| \leq \rho < \delta$ . Hence, by (72), (73) and (74),

$$\begin{aligned}\langle x_i^*, x \rangle &= \langle x_i^*, \omega'_i - \omega_i \rangle - \langle x_i^*, a_i \rangle < -\varepsilon\rho \|x_i^*\| + (\varepsilon_1 + \varepsilon_2)\rho, \\ \langle x^*, x \rangle &= \langle x^*, \omega'_n - \omega_n \rangle < \varepsilon_1\rho,\end{aligned}$$

and consequently, using (36) and (71),

$$\langle x^* - x_n^*, x \rangle = \sum_{i=1}^{n-1} \langle x_i^*, x \rangle + \langle x^*, x \rangle < -\varepsilon\rho + n\varepsilon_1\rho + (n-1)\varepsilon_2\rho < -\varepsilon'\rho.$$

On the other hand, by (70),  $\langle x^* - x_n^*, x \rangle > -\varepsilon'\rho$ . A contradiction.  $\square$

The corresponding ‘symmetric’ statement follows immediately after applying Proposition 6.2 to the collection of  $n+1$  sets  $\Omega_1, \dots, \Omega_n$  and  $X$ .

**Corollary 6.8** *Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) and  $\varepsilon > 0$ . If vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfy conditions (D1), where  $N$  stands for the Fréchet normal cone ( $N := N^F$ ), then there is a  $\delta > 0$  such that, for any  $\rho \in ]0, \delta[$  and  $\tau \in ]0, 1[$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying condition (P3) and*

$$\sum_{i=1}^n \langle x_i^*, a_i \rangle > \tau\varepsilon\rho. \quad (75)$$

In view of Proposition 3.1(iv), the above two statements produce dual sufficient characterizations for approximate stationarity.

**Corollary 6.9** *Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$  and  $\bar{x} \in \cap_{i=1}^n \Omega_i$ . If for any  $\varepsilon > 0$  there exist points  $\omega_i \in \Omega_i \cap B_\varepsilon(\bar{x})$  ( $i = 1, \dots, n$ ) and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying either conditions (68) or conditions (D1), then the collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately stationary at  $\bar{x}$ .*

**Remark 6.7** Similar to the dual necessary characterizations of extremality/stationarity properties discussed in the first part of this section, the sufficient conditions in Proposition 6.2 and Corollary 6.8 contain conditions (69) and (75), respectively, relating the given dual vectors  $x_i^*$  ( $i = 1, \dots, n$ ) and the primal space translation vectors  $a_i$  ( $i = 1, \dots, n$ ) guaranteed by the statements. In view of Remark 6.1(ii), such conditions seem to be an intrinsic feature of the extremality/stationarity properties, independently on whether one goes from primal space conditions to dual space ones or the other way round.

Combining Corollaries 6.7 and 6.8, we can formulate a full dual characterization of approximate  $\alpha$ -stationarity when either the space is Asplund or the sets are convex.

**Corollary 6.10** *Suppose  $\Omega_1, \dots, \Omega_n$  are closed subsets of a Banach space  $X$ ,  $\bar{x} \in \cap_{i=1}^n \Omega_i$  and  $\alpha > 0$ . Suppose also that either  $X$  is Asplund or  $\Omega_1, \dots, \Omega_n$  are convex. The collection  $\{\Omega_1, \dots, \Omega_n\}$  is approximately  $\alpha$ -stationary at  $\bar{x}$  if and only if, for any  $\varepsilon > 0$ , there exist points  $\omega_i \in B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) such that conditions (D3) are satisfied with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ).*

*Moreover, under the above conditions, if  $X$  is Asplund and  $\tau \in ]0, \frac{1}{1+2\alpha}[$ , or  $\Omega_1, \dots, \Omega_n$  are convex and  $\tau := \frac{1}{1+2\alpha}$ , then, for any  $\varepsilon > 0$ , points  $\omega_i \in B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) can be chosen to satisfy also condition (63), while, for any  $\hat{\tau} \in ]0, 1[$ , vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) in Definition 3.1 of the approximate  $\alpha$ -stationarity can be chosen to satisfy additionally  $\sum_{i=1}^n \langle x_i^*, a_i \rangle > \hat{\tau}\alpha\rho$ .*

**Proof.** The ‘only if’ part together with the condition (63) in the ‘moreover’ part follow from parts (i) and (iii) of Corollary 6.7, taking into account that for convex sets the Clarke and Fréchet normal cones coincide. Conversely, given any  $\varepsilon > 0$ ,  $\hat{\tau} \in ]0, 1[$ , points  $\omega_i \in B_\varepsilon(\bar{x})$  and vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfying conditions (D3), Corollary 6.8 with  $\alpha$  and  $\hat{\tau}$  in place of  $\varepsilon$  and  $\tau$ , respectively, yields the approximate  $\alpha$ -stationarity and condition  $\sum_{i=1}^n \langle x_i^*, a_i \rangle > \hat{\tau}\alpha\rho$ .  $\square$

**Remark 6.8** In view of Corollary 2.3 and Remark 3.2, the first part of Corollary 6.10 yields the statements of Theorems 2.2 and 3.2.

The assumption  $x_i^* \in N_{\Omega_i}^F(\omega_i)$  ( $i = 1, \dots, n-1$ ) in Proposition 6.2 can be relaxed (at the expense of weakening the estimates in (P6) and (69)). The next statement is a consequence of Propositions 6.2 and 2.4.

**Corollary 6.11** *Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) and  $\varepsilon \in ]0, 1[$ . If vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfy conditions (D6) with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ), then there is a  $\delta > 0$  such that, for any  $\rho \in ]0, \delta[$  and  $\tau \in ]0, 1[$ , there exist vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying conditions (P6) and (69) with  $\varepsilon' := \varepsilon/(1-\varepsilon)$  in place of  $\varepsilon$ .*

Since  $\varepsilon/(1-\varepsilon) < \varepsilon$ , the conclusions of Corollary 6.11 are weaker than those of Proposition 6.2. Observe that conditions (P6) involve a localization of the  $n$ -th set (near  $\omega_n \in \Omega_n$ ). The estimates can be improved by considering localizations of all the sets.

**Proposition 6.3** *Suppose  $\Omega_1, \dots, \Omega_n$  are subsets of a normed vector space  $X$ ,  $\omega_i \in \Omega_i$  ( $i = 1, \dots, n$ ) and  $\varepsilon > 0$ . If vectors  $x_i^* \in X^*$  ( $i = 1, \dots, n$ ) satisfy conditions (D6) with  $N$  standing for the Fréchet normal cone ( $N := N^F$ ), then there is a  $\delta > 0$  such that, for any  $\rho \in ]0, \delta[$  and  $\tau \in ]0, 1[$ , there are vectors  $a_i \in X$  ( $i = 1, \dots, n-1$ ) satisfying*

$$\bigcap_{i=1}^{n-1} \left( (\Omega_i - \omega_i) \cap (\rho\mathbb{B}) - a_i \right) \cap (\Omega_n - \omega_n) \cap (\rho\mathbb{B}) = \emptyset \quad \text{and} \quad \max_{1 \leq i \leq n-1} \|a_i\| < \varepsilon\rho, \quad (76)$$

and condition (69).

The proof below is a modification of that of Proposition 6.2.

**Proof.** Choose vectors  $y_i^* \in N_{\Omega_i}^F(\omega_i)$  ( $i = 1, \dots, n$ ) and a positive number  $\varepsilon' < \varepsilon$  such that

$$\sum_{i=1}^n \|x_i^* - y_i^*\| < \varepsilon'. \quad (77)$$

Then choose numbers  $\varepsilon_1 > 0$  and  $\varepsilon_2 > 0$  such that  $\varepsilon_1 + \varepsilon_2 < \varepsilon - \varepsilon'$  and  $\varepsilon_2 < (1-\tau)\varepsilon$ . By the definition of the Fréchet normal cone, there is a  $\delta > 0$  such that

$$\langle y_i^*, \omega - \omega_i \rangle \leq \frac{\varepsilon_1}{n} \|\omega - \omega_i\| \quad \text{for all } \omega \in \Omega_i \cap B_\delta(\omega_i) \quad (i = 1, \dots, n). \quad (78)$$

Let  $\rho \in ]0, \delta[$ . Choose vectors  $a_i \in X$  ( $i = 1, \dots, n$ ) satisfying

$$\|a_i\| < \varepsilon\rho \quad \text{and} \quad \langle x_i^*, a_i \rangle > \varepsilon\rho \|x_i^*\| - \frac{\varepsilon_2\rho}{n-1} \quad (i = 1, \dots, n-1).$$

By (D6), we have

$$\sum_{i=1}^{n-1} \langle x_i^*, a_i \rangle > \varepsilon\rho - \varepsilon_2\rho. \quad (79)$$

The inequality in (76) and condition (69) follow. Suppose that the equality in (76) is not satisfied. Then there exist points  $\omega'_i \in \Omega_i \cap B_\rho(\omega_i)$  ( $i = 1, \dots, n$ ) and an  $x \in \rho\mathbb{B}$  such that

$$\omega'_1 - \omega_1 - a_1 = \dots = \omega'_{n-1} - \omega_{n-1} - a_{n-1} = \omega'_n - \omega_n = x. \quad (80)$$

Hence, making use of (80), (D6), (79), (78) and (77), we have

$$\begin{aligned} 0 &= \sum_{i=1}^{n-1} \langle x_i^*, (\omega'_n - \omega_n) - (\omega'_i - \omega_i - a_i) \rangle = - \sum_{i=1}^{n-1} \langle x_i^*, \omega'_i - \omega_i - a_i \rangle - \langle x_n^*, \omega'_n - \omega_n \rangle \\ &= \sum_{i=1}^{n-1} \langle x_i^*, a_i \rangle - \sum_{i=1}^n \langle x_i^*, \omega'_i - \omega_i \rangle = \sum_{i=1}^{n-1} \langle x_i^*, a_i \rangle - \sum_{i=1}^n \langle y_i^*, \omega'_i - \omega_i \rangle + \sum_{i=1}^n \langle y_i^* - x_i^*, \omega'_i - \omega_i \rangle \\ &> \varepsilon\rho - \varepsilon_2\rho - \frac{\varepsilon_1}{n} \sum_{i=1}^n \|\omega'_i - \omega_i\| - \varepsilon' \max_{1 \leq i \leq n} \|\omega'_i - \omega_i\| \geq (\varepsilon - \varepsilon' - \varepsilon_1 - \varepsilon_2)\rho > 0. \end{aligned}$$

This contradiction proves the proposition.  $\square$

## 7 Conclusions

We have exposed, analyzed and refined the core arguments used in various proofs of the extremal principle and its extensions as well as in primal and dual characterizations of the approximate stationarity and transversality of collections of sets, presenting a unifying theory, encompassing all existing approaches to obtaining ‘extremal’ statements. For that, we have examined and clarified quantitative relationships between the parameters involved in the respective definitions and statements. Some new characterizations of extremality properties have been obtained.

**Acknowledgements** The research was supported by the Australian Research Council, project DP160100854. Hoa T. Bui is supported by an Australian Government Research Training Program (RTP) Stipend and RTP Fee-Offset Scholarship through Federation University Australia. Alexander Y. Kruger benefited from the support of the FMJH Program PGMO and from the support of EDF.

We wish to thank PhD student Nguyen Duy Cuong from Federation University Australia for careful reading of the manuscript and helping us eliminate numerous typos, and the anonymous referees for their constructive comments and suggestions.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

1. Rockafellar, R.T.: *Convex Analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J. (1970)
2. Zălinescu, C.: *Convex Analysis in General Vector Spaces*. World Scientific Publishing Co. Inc., River Edge, NJ (2002). DOI 10.1142/9789812777096
3. Penot, J.P.: *Analysis—from Concepts to Applications*. Universitext. Springer, [Cham] (2016)
4. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. *SIAM Rev.* **38**(3), 367–426 (1996). DOI 10.1137/S0036144593251710
5. Kruger, A.Y., Luke, D.R., Thao, N.H.: Set regularities and feasibility problems. *Math. Program., Ser. B* **168**(1-2), 279–311 (2018). DOI 10.1007/s10107-016-1039-x
6. Drusvyatskiy, D., Ioffe, A.D., Lewis, A.S.: Transversality and alternating projections for nonconvex sets. *Found. Comput. Math.* **15**(6), 1637–1651 (2015). DOI 10.1007/s10208-015-9279-3
7. Bauschke, H.H., Dao, M.N., Noll, D., Phan, H.M.: Proximal point algorithm, Douglas-Rachford algorithm and alternating projections: a case study. *J. Convex Anal.* **23**(1), 237–261 (2016)
8. Kruger, A.Y., Thao, N.H.: Regularity of collections of sets and convergence of inexact alternating projections. *J. Convex Anal.* **23**(3), 823–847 (2016)
9. Noll, D., Rondepierre, A.: On local convergence of the method of alternating projections. *Found. Comput. Math.* **16**(2), 425–455 (2016). DOI 10.1007/s10208-015-9253-0
10. Dubovitskii, A.Y., Miljutin, A.A.: Extremum problems in the presence of restrictions. *USSR Comp. Maths. Math. Phys.* **5**, 1–80 (1965)
11. Kruger, A.Y., Mordukhovich, B.S.: New necessary optimality conditions in problems of nondifferentiable programming. In: *Numerical Methods of Nonlinear Programming*, pp. 116–119. Kharkov (1979). In Russian. Available at <https://asterius.ballarat.edu.au/kruger/research/publications.html>
12. Kruger, A.Y., Mordukhovich, B.S.: Generalized normals and derivatives and necessary conditions for an extremum in problems of nondifferentiable programming. II. *VINITI no. 494-80*, 60 pp. Minsk (1980). In Russian. Available at <https://asterius.ballarat.edu.au/kruger/research/publications.html>
13. Kruger, A.Y., Mordukhovich, B.S.: Extremal points and the Euler equation in nonsmooth optimization problems. *Dokl. Akad. Nauk BSSR* **24**(8), 684–687 (1980). In Russian. Available at <https://asterius.ballarat.edu.au/kruger/research/publications.html>
14. Kruger, A.Y.: Generalized differentials of nonsmooth functions. *VINITI no. 1332-81*. Minsk (1981). 67 pp. In Russian. Available at <https://asterius.ballarat.edu.au/kruger/research/publications.html>
15. Kruger, A.Y.: Generalized differentials of nonsmooth functions and necessary conditions for an extremum. *Sibirsk. Mat. Zh.* **26**(3), 78–90 (1985). (In Russian; English transl.: *Siberian Math. J.* 26 (1985), 370–379)
16. Kruger, A.Y.:  $\varepsilon$ -semidifferentials and  $\varepsilon$ -normal elements. *VINITI no. 1331-81*. Minsk (1981). 76 pp. In Russian. Available at <https://asterius.ballarat.edu.au/kruger/research/publications.html>
17. Fabian, M.: Subdifferentiability and trustworthiness in the light of a new variational principle of Borwein and Preiss. *Acta Univ. Carolinae* **30**, 51–56 (1989)
18. Mordukhovich, B.S., Shao, Y.: Extremal characterizations of Asplund spaces. *Proc. Amer. Math. Soc.* **124**(1), 197–205 (1996)
19. Ioffe, A.D.: Fuzzy principles and characterization of trustworthiness. *Set-Valued Anal.* **6**, 265–276 (1998)
20. Borwein, J.M., Jofré, A.: A nonconvex separation property in Banach spaces. *Math. Methods Oper. Res.* **48**(2), 169–179 (1998)
21. Mordukhovich, B.S.: *Variational Analysis and Generalized Differentiation. I: Basic Theory, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 330. Springer, Berlin (2006)
22. Kruger, A.Y., López, M.A.: Stationarity and regularity of infinite collections of sets. *J. Optim. Theory Appl.* **154**(2), 339–369 (2012)
23. Kruger, A.Y.: On Fréchet subdifferentials. *J. Math. Sci.* **116**(3), 3325–3358 (2003)

24. Kruger, A.Y.: Weak stationarity: eliminating the gap between necessary and sufficient conditions. *Optimization* **53**(2), 147–164 (2004)
25. Kruger, A.Y.: About stationarity and regularity in variational analysis. *Taiwanese J. Math.* **13**(6A), 1737–1785 (2009)
26. Kruger, A.Y., Luke, D.R., Thao, N.H.: About subtransversality of collections of sets. *Set-Valued Var. Anal.* **25**(4), 701–729 (2017). DOI 10.1007/s11228-017-0436-5
27. Zheng, X.Y., Ng, K.F.: Perturbation analysis of error bounds for systems of conic linear inequalities in Banach spaces. *SIAM J. Optim.* **15**(4), 1026–1041 (2005)
28. Zheng, X.Y., Ng, K.F.: The Lagrange multiplier rule for multifunctions in Banach spaces. *SIAM J. Optim.* **17**(4), 1154–1175 (2006). DOI 10.1137/060651860
29. Li, G., Ng, K.F., Zheng, X.Y.: Unified approach to some geometric results in variational analysis. *J. Funct. Anal.* **248**(2), 317–343 (2007)
30. Li, G., Tang, C., Yu, G., Wei, Z.: On a separation principle for nonconvex sets. *Set-Valued Anal.* **16**, 851–860 (2008)
31. Zheng, X.Y., Ng, K.F.: A unified separation theorem for closed sets in a Banach space and optimality conditions for vector optimization. *SIAM J. Optim.* **21**(3), 886–911 (2011). DOI 10.1137/100811155
32. Bui, H.T., Kruger, A.Y.: About extensions of the extremal principle. *Vietnam J. Math.* **46**(2), 215–242 (2018). DOI 10.1007/s10013-018-0278-y
33. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*. Springer, Berlin (1998)
34. Penot, J.P.: Calculus without Derivatives, *Graduate Texts in Mathematics*, vol. 266. Springer, New York (2013). DOI 10.1007/978-1-4614-4538-8
35. Dontchev, A.L., Rockafellar, R.T.: *Implicit Functions and Solution Mappings. A View from Variational Analysis*, 2 edn. Springer Series in Operations Research and Financial Engineering. Springer, New York (2014)
36. Ioffe, A.D.: *Variational Analysis of Regular Mappings. Theory and Applications*. Springer Monographs in Mathematics. Springer (2017)
37. Ekeland, I.: On the variational principle. *J. Math. Anal. Appl.* **47**, 324–353 (1974)
38. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. John Wiley & Sons Inc., New York (1983)
39. Ioffe, A.D., Tikhomirov, V.M.: *Theory of Extremal Problems, Studies in Mathematics and its Applications*, vol. 6. North-Holland Publishing Co., Amsterdam (1979)
40. Rockafellar, R.T.: Directionally Lipschitzian functions and subdifferential calculus. *Proc. London Math. Soc.* (3) **39**(2), 331–355 (1979)
41. Phelps, R.R.: *Convex Functions, Monotone Operators and Differentiability, Lecture Notes in Mathematics*, vol. 1364, second edn. Springer-Verlag, Berlin (1993)
42. Borwein, J.M., Zhu, Q.J.: *Techniques of Variational Analysis*. Springer, New York (2005)
43. Kruger, A.Y.: About extremality of systems of sets. *Dokl. Nats. Akad. Nauk Belarusi* **42**(1), 24–28 (1998). In Russian. Available at <https://asterius.ballarat.edu.au/akruger/research/publications.html>
44. Kruger, A.Y.: About regularity of collections of sets. *Set-Valued Anal.* **14**(2), 187–206 (2006)
45. Kruger, A.Y.: Stationarity and regularity of set systems. *Pac. J. Optim.* **1**(1), 101–126 (2005)
46. Kruger, A.Y.: Strict  $(\varepsilon, \delta)$ -semidifferentials and extremality of sets and functions. *Dokl. Nats. Akad. Nauk Belarusi* **44**(2), 19–22 (2000). In Russian. Available at <https://asterius.ballarat.edu.au/akruger/research/publications.html>
47. Kruger, A.Y.: Strict  $(\varepsilon, \delta)$ -subdifferentials and extremality conditions. *Optimization* **51**(3), 539–554 (2002)
48. Kruger, A.Y., Thao, N.H.: About uniform regularity of collections of sets. *Serdica Math. J.* **39**, 287–312 (2013)
49. Kruger, A.Y., Thao, N.H.: Quantitative characterizations of regularity properties of collections of sets. *J. Optim. Theory Appl.* **164**(1), 41–67 (2015). DOI 10.1007/s10957-014-0556-0
50. Lewis, A.S., Luke, D.R., Malick, J.: Local linear convergence for alternating and averaged nonconvex projections. *Found. Comput. Math.* **9**(4), 485–513 (2009). DOI 10.1007/s10208-008-9036-y
51. Bauschke, H.H., Luke, D.R., Phan, H.M., Wang, X.: Restricted normal cones and the method of alternating projections: theory. *Set-Valued Var. Anal.* **21**(3), 431–473 (2013). DOI 10.1007/s11228-013-0239-2
52. Hesse, R., Luke, D.R.: Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM J. Optim.* **23**(4), 2397–2419 (2013). DOI 10.1137/120902653
53. Bui, H.T., Lindstrom, S.B., Roshchina, V.: Variational Analysis Down Under 2018 open problem session. *J. Optim. Theory Appl.* DOI 10.1007/s10957-018-1399-x. This issue