

---

## Extremal Principle in Variational Analysis

It is well known that the convex separation principle plays a fundamental role in many aspects of nonlinear analysis, optimization, and their applications. Actually the whole convex analysis revolves around using separation theorems for convex sets. In problems with nonconvex data separation theorems are applied to convex approximations. This is a conventional way to derive necessary optimality conditions in constrained optimization: first build *tangential convex* approximations of the problem data around an optimal solution in *primal* spaces and then apply convex separation theorems to get supporting elements in dual spaces (Lagrange multipliers, adjoint arcs, prices, etc.). For problems of nonsmooth optimization this approach inevitably leads to the usage of *convex* sets of normals and subgradients, whose calculus is also based on convex separation theorems.

This chapter is devoted to another principle in variational analysis, called the *extremal principle*, which can be viewed as a variational counterpart of the convex separation principle in nonconvex settings. The extremal principle provides necessary conditions for local extremal points of set systems in terms of generalized normals to nonconvex sets with *no use* of tangential approximations and convex separation. It is the base for subsequent applications in this book to nonconvex calculus, optimization, and related topics.

We mainly consider three versions of the extremal principle in Banach spaces formulated, respectively, in terms of  $\varepsilon$ -normals, Fréchet normals, and basic normals from Chap. 1. It will be shown, by direct variational arguments and the method of separable reduction, that the class of *Asplund spaces* is the most suitable framework for the validity and applications of these results. We also establish relationships between the extremal principle and other basic results in variational analysis, obtain a number of variational characterizations of Asplund spaces in terms of the normal and subgradient constructions studied above, and derive their simplified representations important in what follows. Finally, we discuss some abstract versions of the extremal principle in terms of axiomatically defined normal and subdifferential structures in appropriate Banach spaces.

## 2.1 Set Extremality and Nonconvex Separation

In this section we introduce a general concept of set extremality and study its relationships with conventional notions of optimal solutions in constrained optimization and separation of sets. We formulate three basic versions of the extremal principle and prove the strongest one in finite-dimensional spaces. As usual, our standard framework is Banach spaces unless otherwise stated.

### 2.1.1 Extremal Systems of Sets

We start with the definition of extremal systems of sets that may belong to linear topological spaces.

**Definition 2.1 (local extremality of set systems).** *Let  $\Omega_1, \dots, \Omega_n$  be nonempty subsets of a space  $X$  for  $n \geq 2$ , and let  $\bar{x}$  be a common point of these sets. We say that  $\bar{x}$  is a LOCAL EXTREMAL POINT of the set system  $\{\Omega_1, \dots, \Omega_n\}$  if there are sequences  $\{a_{ik}\} \subset X$ ,  $i = 1, \dots, n$ , and a neighborhood  $U$  of  $\bar{x}$  such that  $a_{ik} \rightarrow 0$  as  $k \rightarrow \infty$  and*

$$\bigcap_{i=1}^n (\Omega_i - a_{ik}) \cap U = \emptyset \text{ for all large } k \in \mathbb{N}.$$

*In this case  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  is said to be an EXTREMAL SYSTEM in  $X$ .*

Loosely speaking, the local extremality of sets at a common point means that they can be locally “pushed apart” by a small perturbation (translation) of even one of them. For  $n = 2$  the local extremality of  $\{\Omega_1, \Omega_2, \bar{x}\}$  can be equivalently described as follows: there exists a neighborhood  $U$  of  $\bar{x}$  such that for any  $\varepsilon > 0$  there is  $a \in \varepsilon B$  with  $(\Omega_1 + a) \cap \Omega_2 \cap U = \emptyset$ . Note that the condition  $\Omega_1 \cap \Omega_2 = \{\bar{x}\}$  doesn’t necessarily imply that  $\bar{x}$  is a local extremal point of  $\{\Omega_1, \Omega_2\}$ . A simple example is given by  $\Omega_1 := \{(v, v) \mid v \in \mathbb{R}\}$  and  $\Omega_2 := \{(v, -v) \mid v \in \mathbb{R}\}$ .

It is clear that every *boundary point*  $\bar{x}$  of a closed set  $\Omega$  is a local extremal point of the pair  $\{\Omega, \bar{x}\}$ . In general, this geometric concept of extremality covers conventional notions of optimal solutions to various problems of scalar and vector optimization. In particular, let  $\bar{x}$  be a local solution to the following problem of *constrained optimization*:

$$\text{minimize } \varphi(x) \text{ subject to } x \in \Omega \subset X.$$

Then one can easily check that  $(\bar{x}, \varphi(\bar{x}))$  is a local extremal point of the set system  $\{\Omega_1, \Omega_2\}$  in  $X \times \mathbb{R}$  with  $\Omega_1 = \text{epi } \varphi$  and  $\Omega_2 = \Omega \times \{\varphi(\bar{x})\}$ . Indeed, we satisfy the requirements of Definition 2.1 with  $a_{1k} = (0, v_k)$ ,  $a_{2k} = 0$ , and  $U = O \times \mathbb{R}$ , where  $v_k \uparrow 0$  and where  $O$  is a neighborhood of the local minimizer  $\bar{x}$ . In the subsequent parts of the book the reader will find many other examples of extremal systems in problems related to optimization, variational principles, generalized differential calculus, and applications to welfare economics.

The next simple property of extremal systems is useful in what follows.

**Proposition 2.2 (interiors of sets in extremal systems).** *For every extremal system  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  in  $X$  one has*

$$(\text{int } \Omega_1) \cap \dots \cap (\text{int } \Omega_{n-1}) \cap \Omega_n \cap U = \emptyset, \quad (2.1)$$

where  $U$  is a neighborhood of the local extremal point  $\bar{x}$ .

**Proof.** Assuming the contrary, pick any point  $x$  from the intersection in (2.1) and take arbitrary sequences  $a_{ik} \rightarrow 0$ ,  $i = 1, \dots, n$ , in  $X$ . Since  $x \in \text{int } \Omega_i \cap U$  for  $i = 1, \dots, n-1$ , we have  $x - a_{nk} \in U$  and  $x + a_{ik} - a_{nk} \in \Omega_i$  for  $i = 1, \dots, n-1$  and  $k \in \mathbb{N}$  large enough. Thus  $x - a_{nk} \in (\Omega_i - a_{ik}) \cap U$  for all  $i = 1, \dots, n$  and large  $k$ , which contradicts the set extremality.  $\triangle$

Now we establish relationships between the concept of set extremality from Definition 2.1 and the conventional separation property for a finite number of sets that may be nonconvex. Recall that sets  $\Omega_i \subset X$ ,  $i = 1, \dots, n$ , are said to be *separated* if there exist vectors  $x_i^* \in X^*$ , not equal to zero simultaneously, and numbers  $\alpha_i$  such that

$$\begin{aligned} \langle x_i^*, x \rangle &\leq \alpha_i \quad \text{for all } x \in \Omega_i, \quad i = 1, \dots, n, \\ x_1^* + \dots + x_n^* &= 0, \quad \alpha_1 + \dots + \alpha_n \leq 0. \end{aligned}$$

Note that if the sets  $\Omega_i$  are separated and have a common point, then the last condition must hold as equality.

**Proposition 2.3 (extremality and separation).** *Let  $\Omega_1, \dots, \Omega_n$  ( $n \geq 2$ ) be subsets of  $X$  that have at least one common point. The following hold:*

(i) *If these sets are separated, then the system  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  is extremal for every common point  $\bar{x}$  of these sets.*

(ii) *The converse is true if all  $\Omega_i$  are convex and  $\text{int } \Omega_i \neq \emptyset$  for  $i = 1, \dots, n-1$ .*

**Proof.** Assume that  $\Omega_i$  are separated with  $x_n^* \neq 0$ , which doesn't restrict the generality. Pick any  $a \in X$  with  $\langle x_n^*, a \rangle > 0$  and put  $a_k := a/k$  for all  $k \in \mathbb{N}$ . Let us show that

$$\Omega_1 \cap \dots \cap \Omega_{n-1} \cap (\Omega_n - a_k) = \emptyset, \quad k \in \mathbb{N},$$

which obviously implies the extremality of  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  for every common point  $\bar{x}$ . Assuming the contrary and taking any  $x$  from the latter intersection, one has by the separation property that

$$\langle x_i^*, x \rangle \leq \alpha_i, \quad i = 1, \dots, n-1, \quad \text{and} \quad \langle x_n^*, x + a_k \rangle \leq \alpha_n, \quad k \in \mathbb{N}.$$

Summing up, we arrive at  $\alpha_1 + \dots + \alpha_n \geq \frac{1}{k} \langle x_n^*, a \rangle > 0$ , a contradiction. Thus (i) holds. The converse assertion (ii) follows from Proposition 2.2 and the separation theorem for convex sets.  $\triangle$

Note that, for *convex sets in finite dimensions*, Proposition 2.3(ii) holds with *no interiority* assumption on  $\Omega_i$ ,  $i = 1, \dots, n-1$ . This follows from the extremal principle established below in Theorem 2.8. Hence for  $\dim X < \infty$  the *extremality and separation of convex sets are unconditionally equivalent*. One will also see that the extremal principle allows us to relax interiority assumptions on convex sets  $\Omega_i$ ,  $i = 1, \dots, n-1$ , ensuring the validity of Proposition 2.3(ii) in infinite dimensions.

**Corollary 2.4 (extremality criterion for convex sets).** *Let  $\Omega_i$ ,  $i = 1, \dots, n$ , be convex sets in  $X$  having at least one point in common. Assume that  $\text{int } \Omega_i \neq \emptyset$  for  $i = 1, \dots, n-1$ . Then condition (2.1) with  $U = X$  is necessary and sufficient for extremality of the system  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$ , where  $\bar{x}$  is any common point of these sets.*

**Proof.** Follows from Propositions 2.2 and 2.3(i), since condition (2.1) ensures the separation (and hence extremality) property of  $n$  convex sets with nonempty interiors of all but one of them.  $\triangle$

Note that the convexity of  $\Omega_i$  is essential for the extremality criterion in Corollary 2.4. A counterexample is provided by the sets

$$\Omega_1 := \mathbb{R}_+^2 \cup \mathbb{R}_-^2, \Omega_2 := \{(x_1, x_2) \mid x_1 \leq 0, x_2 \geq 0\} \cup \{(x_1, x_2) \mid x_1 \geq 0, x_2 \leq 0\}.$$

### 2.1.2 Versions of the Extremal Principle and Supporting Properties

In this subsection we define three basic versions of the extremal principle in Banach spaces and show that they can be treated as a kind of local separation of nonconvex sets around extremal points. We also discuss their relationships with supporting properties of nonconvex sets expressed in terms of generalized normals from Definition 1.1.

**Definition 2.5 (versions of the extremal principle).** *Let  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  be an extremal system in  $X$ . We say that:*

(i)  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  *satisfies the  $\varepsilon$ -EXTREMAL PRINCIPLE if for every  $\varepsilon > 0$  there are  $x_i \in \Omega_i \cap (\bar{x} + \varepsilon B)$  and  $x_i^* \in X^*$  such that*

$$x_i^* \in \widehat{N}_\varepsilon(x_i; \Omega_i), \quad i = 1, \dots, n, \quad (2.2)$$

$$x_1^* + \dots + x_n^* = 0, \quad \|x_1^*\| + \dots + \|x_n^*\| = 1. \quad (2.3)$$

(ii)  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  *satisfies the APPROXIMATE EXTREMAL PRINCIPLE if for every  $\varepsilon > 0$  there are  $x_i \in \Omega_i \cap (\bar{x} + \varepsilon B)$  and*

$$x_i^* \in \widehat{N}(x_i; \Omega_i) + \varepsilon B^*, \quad i = 1, \dots, n, \quad (2.4)$$

*such that (2.3) holds.*

(iii)  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  satisfies the EXACT EXTREMAL PRINCIPLE if there are basic normals

$$x_i^* \in N(\bar{x}; \Omega_i), \quad i = 1, \dots, n, \quad (2.5)$$

such that (2.3) holds.

We say that the corresponding version of the extremal principle holds in the space  $X$  if it holds for every extremal system  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  in  $X$ , where all the sets  $\Omega_i$  are (locally) closed around  $\bar{x}$ .

It is clear that the number 1 in the nontriviality condition of (2.3) can be replaced with any other positive number, which should be independent of  $\varepsilon$  in versions (i) and (ii). Note that  $\varepsilon$  in “ $\varepsilon$ -extremal principle” is just a part of the notation (and not a subject to change unlike anywhere else), which emphasizes the difference between (2.2) and (2.4). Since one always has  $\hat{N}(x; \Omega) + \varepsilon B^* \subset \hat{N}_\varepsilon(x; \Omega)$ , the  $\varepsilon$ -extremal principle follows from the approximate extremal principle for any extremal system in a Banach space  $X$ . We’ll see below that these two versions of the extremal principle are actually equivalent if they apply to *every* extremal system in  $X$ .

Thus the relations of the extremal principle provide necessary conditions for local extremal points of set systems and can be viewed as *generalized Euler equations* in an abstract geometric setting. They also can be treated as proper variational counterparts of local separation for nonconvex sets. To see this, we first consider the exact extremal principle for two sets. Then (2.3) and (2.5) reduce to: there is  $x^* \in X^*$  with

$$0 \neq x^* \in N(\bar{x}; \Omega_1) \cap (-N(\bar{x}; \Omega_2)). \quad (2.6)$$

When both  $\Omega_1$  and  $\Omega_2$  are *convex*, (2.6) means

$$\langle x^*, u_1 \rangle \leq \langle x^*, u_2 \rangle \quad \text{for all } u_1 \in \Omega_1 \text{ and } u_2 \in \Omega_2,$$

which is exactly the classical separation property for two convex sets. Similarly, relations (2.3) and (2.5) for  $n$  convex sets ( $n > 2$ ) give the conventional separation property considered in the preceding subsection.

Note that, in contrast to the classical separation, the extremal principle applies only to *local extremal points* of set systems. As shown in Proposition 2.3, it is *always the case* for every common point of sets separated in the classical sense. Therefore, any sufficient condition for convex separation implies set extremality. The above discussion allows us to view the extremal principle as a local *variational* extension of the classical separation to nonconvex sets. It is important to emphasize that in many situations occurring in applications, even in the case of convex sets, the local extremality of points in question can be checked *automatically* from the problem statement, and we don’t need to care about any interiority-like conditions, etc. This supports a variational approach to such problems (which may be not of a variational nature) based on the extremal principle; see below.

Considering “fuzzy” versions (i) and (ii) of the extremal principle for systems of two sets, we reduce them to the following relations: for every  $\varepsilon > 0$  there are  $x_i \in \Omega_i \cap (\bar{x} + \varepsilon \mathcal{B})$ ,  $i = 1, 2$ , and  $x^* \in X^*$  with  $\|x^*\| = 1$  such that, respectively,

$$\begin{aligned} x^* &\in \widehat{N}_\varepsilon(x_1; \Omega_1) \cap (-\widehat{N}_\varepsilon(x_2; \Omega_2)) , \\ x^* &\in (\widehat{N}(x_1; \Omega_1) + \varepsilon \mathcal{B}^*) \cap (-\widehat{N}(x_2; \Omega_2) + \varepsilon \mathcal{B}^*) . \end{aligned}$$

For convex sets they coincide, due to Proposition 1.3, and provide an *approximate separation* of  $\Omega_1$  and  $\Omega_2$  near  $\bar{x}$ . Likewise, relations (2.2)–(2.4) of the extremal principle in the general case under consideration can be viewed as a local variational counterpart of the approximate local separation for nonconvex sets.

Next let us consider a special case of extremal systems generated by *boundary points*  $\bar{x}$  of locally closed sets  $\Omega \subset X$ , i.e., extremal systems of the type  $\{\Omega, \{\bar{x}\}, \bar{x}\}$  in the notation of Definition 2.1. Then the exact extremal principle gives the *nontriviality property* for the basic normal cone:

$$N(\bar{x}; \Omega) \neq \{0\} \text{ if and only if } \bar{x} \in \text{bd } \Omega . \quad (2.7)$$

Note that the “only if” part follows immediately from Definition 1.1 for any closed set  $\Omega \subset X$ , and the “if” part is an easy consequence of the exact extremal principle whenever it holds in  $X$ . When  $\Omega$  is convex, condition (2.7) reduces to the classical *supporting hyperplane* theorem; so in general (2.7) can be viewed as a local extension of this result to nonconvex sets. Applying the other versions of the extremal principle, we get some approximate supporting properties of nonconvex sets in terms of  $\varepsilon$ -normals and Fréchet normals at points near  $\bar{x}$ .

**Proposition 2.6 (approximate supporting properties of nonconvex sets).** *Given a proper closed set  $\Omega \subset X$  and a point  $\bar{x} \in \text{bd } \Omega$ , one has the following:*

(i) *If the  $\varepsilon$ -extremal principle holds for  $\{\Omega, \{\bar{x}\}, \bar{x}\}$ , then whenever  $\varepsilon > 0$  and  $M > \varepsilon$  there is  $x \in B_\varepsilon(\bar{x}) \cap \text{bd } \Omega$  such that  $\widehat{N}_\varepsilon(x; \Omega) \setminus M\mathcal{B}^* \neq \emptyset$ .*

(ii) *If the approximate extremal principle holds for  $\{\Omega, \{\bar{x}\}, \bar{x}\}$ , then for every  $\varepsilon > 0$  there is  $x \in B_\varepsilon(\bar{x}) \cap \text{bd } \Omega$  such that  $\widehat{N}(x; \Omega) \neq \{0\}$ .*

*Therefore, the validity of the approximate extremal principle (the  $\varepsilon$ -extremal principle) in  $X$  implies, respectively, the density of the set*

$$\left\{ x \in \text{bd } \Omega \mid \widehat{N}(x; \Omega) \neq \{0\} \right\} \quad (2.8)$$

*for every proper closed subset  $\Omega \subset X$ , and the set*

$$\left\{ x \in \text{bd } \Omega \mid \widehat{N}_\varepsilon(x; \Omega) \setminus M\mathcal{B}^* \neq \emptyset \right\} \quad (2.9)$$

*for every proper closed subset  $\Omega \subset X$ , every  $\varepsilon > 0$ , and every  $M > \varepsilon$ .*

**Proof.** Assertion (i) for  $0 < M < 1/2$  follows immediately from Definition 2.5(i) with  $n = 2$ ,  $\Omega_1 = \Omega$ , and  $\Omega_2 = \{\bar{x}\}$ . Let us prove it for any  $M > \varepsilon$ . Fix arbitrary  $\varepsilon > 0$  and  $M \geq 1/2$  and employ the relations of the  $\varepsilon$ -extremal principle to  $\{\Omega, \{\bar{x}\}, \bar{x}\}$  with  $\tilde{\varepsilon} := \varepsilon/(2M+1)$ . We find  $x \in \Omega$  and  $\tilde{x}^* \in X^*$  satisfying

$$\|x - \bar{x}\| \leq \tilde{\varepsilon} < \varepsilon, \quad \tilde{x}^* \in \widehat{N}_{\tilde{\varepsilon}}(x; \Omega), \quad \text{and} \quad \|\tilde{x}^*\| = 1/2,$$

which implies that  $x \in \text{bd } \Omega$ . Then putting  $x^* := (2M+1)\tilde{x}^*$  and using the definition of  $\varepsilon$ -normals (1.2), we get

$$\limsup_{u \xrightarrow{\Omega} x} \frac{\langle x^*, u - x \rangle}{\|u - x\|} = (2M+1) \limsup_{u \xrightarrow{\Omega} x} \frac{\langle \tilde{x}^*, u - x \rangle}{\|u - x\|} \leq (2M+1)\tilde{\varepsilon} = \varepsilon,$$

i.e.,  $x^* \in \widehat{N}_{\varepsilon}(x; \Omega)$  with  $\|x^*\| = (2M+1)/2 > M$ . This gives (i).

To prove (ii), we use the approximate extremal principle for  $\{\Omega, \{\bar{x}\}, \bar{x}\}$  with  $\varepsilon \in (0, 1/2)$ . In this way we find  $x \in B_{\varepsilon}(x) \cap \Omega$  and  $x^* \in \widehat{N}(x; \Omega) + \varepsilon B^*$  with  $\|x^*\| = 1/2$ . The latter yields  $x \in \text{bd } \Omega$  and  $\widehat{N}(x; \Omega) \neq \{0\}$ .  $\triangle$

If  $\Omega$  is *convex*, then (2.8) describes the set of *support points* to  $\Omega$ . Hence the approximate extremal principle in a Banach space  $X$  implies the density of support points to every closed convex subset of  $X$ , which is the contents of the celebrated Bishop-Phelps theorem (see Theorem 3.18 in Phelps [1073]).

A natural question arises about the reverse implications in Proposition 2.6, i.e., about the possibility to derive relations of the approximate extremal principle (resp. the  $\varepsilon$ -extremal principle) from the density of sets (2.8) and (2.9) for *every* proper closed subset of  $X$ . To explore this way, let us fix an extremal system  $\{\Omega_1, \Omega_2, \bar{x}\}$  and observe that the local extremality of  $\bar{x} \in \Omega_1 \cap \Omega_2$  implies that  $0 \in \text{bd } (\Omega_1 - \Omega_2)$ . Hence one can apply the mentioned density results to the set  $\Omega_1 - \Omega_2$  around the origin if  $\Omega_1 - \Omega_2$  is *assumed to be closed*. For simplicity let us consider the case of (2.8) and find  $x_i \in \Omega_i$ ,  $i = 1, 2$ , such that

$$\widehat{N}(x_1 - x_2; \Omega_1 - \Omega_2) \neq \{0\} \quad \text{and} \quad \|x_1 - x_2\| \leq \varepsilon.$$

Taking  $x^* \in \widehat{N}(x_1 - x_2; \Omega_1 - \Omega_2)$  with  $\|x^*\| = 1/2$ , we have from (1.2) that

$$\limsup_{u \xrightarrow{\Omega_1 - \Omega_2} x_1 - x_2} \frac{\langle x^*, u - (x_1 - x_2) \rangle}{\|u - (x_1 - x_2)\|} \leq 0.$$

Now putting  $u = v - x_2$ ,  $v \in \Omega_1$  and then  $u = x_1 - v$ ,  $v \in \Omega_2$ , one gets  $x^* \in \widehat{N}(x_1; \Omega_1)$  and  $-x^* \in \widehat{N}(x_2; \Omega_2)$ . In this way we arrive at all the relations of the approximate extremal principle *except* that  $x_i \in \bar{x} + \varepsilon B^*$ ,  $i = 1, 2$ . Thus we cannot obtain the reverse statements in Proposition 2.6 using the reduction of local extremal points to the boundary of  $\Omega_1 - \Omega_2$ . Moreover, the above arguments actually provide *characterizations* of the supporting properties  $\widehat{N}_{\varepsilon}(x; \Omega) \setminus MB^* \neq \emptyset$  and  $\widehat{N}(x; \Omega) \neq \{0\}$  in terms of relations (2.2)–(2.4), which *don't involve* extremal points and their small perturbations.

**Proposition 2.7 (characterizations of supporting properties).** *Given a Banach space  $X$  and numbers  $\varepsilon \geq 0$  and  $M \geq \varepsilon$ , the following properties are equivalent:*

(a) *For every proper closed set  $\Omega \subset X$  there exists  $x \in \text{bd } \Omega$  satisfying  $\widehat{N}_\varepsilon(x; \Omega) \setminus MB^* \neq \emptyset$ , which corresponds to  $\widehat{N}(x; \Omega) \neq \{0\}$  if  $\varepsilon = 0$ .*

(b) *Let  $\Omega_1$  and  $\Omega_2$  be arbitrary subsets of  $X$  such that  $\Omega_1 - \Omega_2$  is proper and closed around the origin. Then there are  $x_1 \in \Omega_1$  and  $x_2 \in \Omega_2$  satisfying*

$$0 \in (\widehat{N}_\varepsilon(x_1; \Omega_1) \setminus MB^*) + \widehat{N}_\varepsilon(x_2; \Omega_2) .$$

**Proof.** To establish (a) $\Rightarrow$ (b), we take  $\Omega := \Omega_1 - \Omega_2$  in (a) and use the above arguments for  $x_1 - x_2 \in \Omega_1 - \Omega_2$  and  $x^* \in \widehat{N}_\varepsilon(x_1 - x_2; \Omega_1 - \Omega_2)$  with  $\|x^*\| > M > \varepsilon \geq 0$ . Implication (b) $\Rightarrow$ (a) is proved similarly to Proposition 2.6 putting  $\Omega_1 := \Omega$  and  $\Omega_2 := \{\bar{x}\}$ , where  $\bar{x}$  is a fixed boundary point of  $\Omega$ .  $\triangle$

### 2.1.3 Extremal Principle in Finite Dimensions

In this subsection we give a direct proof of the exact extremal principle in finite-dimensional spaces. The proof is based on the *method of metric approximations*, which provides an efficient approximation of extremal set systems by families of *smooth* problems of *unconstrained optimization*. Without loss of generality we use the Euclidean norm on  $X$ .

**Theorem 2.8 (exact extremal principle in finite dimensions).** *The exact extremal principle holds in any space  $X$  with  $\dim X < \infty$ .*

**Proof.** Let  $\bar{x}$  be a local extremal point of the set system  $\{\Omega_1, \dots, \Omega_n\}$ , where all the sets  $\Omega_i$  are closed around  $\bar{x}$ . Take sequences  $\{a_{ik}\}$  and a neighborhood  $U$  from Definition 2.1 and assume without loss of generality that  $U = X$ . For each  $k = 1, 2, \dots$  we consider the following problem of unconstrained minimization:

$$\text{minimize } d_k(x) := \left[ \sum_{i=1}^n \text{dist}^2(x + a_{ik}; \Omega_i) \right]^{1/2} + \|x - \bar{x}\|^2, \quad x \in X. \quad (2.10)$$

Since the function  $d_k$  is continuous and its level sets are bounded, there is an optimal solution  $x_k$  to (2.10) by the classical Weierstrass theorem. Due to the local extremality of  $\bar{x}$  one has

$$\alpha_k := \left[ \sum_{i=1}^n \text{dist}^2(x_k + a_{ik}; \Omega_i) \right]^{1/2} > 0 .$$

Taking into account that  $x_k$  is an optimal solution to (2.10), we get



$$d_k(x_k) = \alpha_k + \|x_k - \bar{x}\|^2 \leq \left[ \sum_{i=1}^n \|a_{ik}\|^2 \right]^{1/2} \downarrow 0 ,$$

which implies that  $x_k \rightarrow \bar{x}$  and  $\alpha_k \downarrow 0$  as  $k \rightarrow \infty$ .

Now let us arbitrarily pick  $w_{ik} \in \Pi(x_k + a_{ik}; \Omega_i)$  for  $i = 1, \dots, n$  (the *best approximations* to  $x_k + a_{ik}$  in the closed set  $\Omega_i$ ) and consider the problem:

$$\text{minimize } \rho_k(x) := \left[ \sum_{i=1}^n \|x + a_{ik} - w_{ik}\|^2 \right]^{1/2} + \|x - \bar{x}\|^2 \quad (2.11)$$

that obviously has the same optimal solution  $x_k$  as (2.10). Since  $\alpha_k > 0$  and the norm  $\|\cdot\|$  is Euclidean,  $\rho_k(x)$  is continuously differentiable around  $x_k$ . Thus (2.11) is a *smooth* problem of unconstrained minimization. Employing the classical Fermat rule in (2.11), we get

$$\nabla \rho_k(x_k) = \sum_{i=1}^n x_{ik}^* + 2(x_k - \bar{x}) = 0 , \quad (2.12)$$

where  $x_{ik}^* = (x_k + a_{ik} - w_{ik})/\alpha_k$ ,  $i = 1, \dots, n$ , with

$$\|x_{1k}^*\|^2 + \dots + \|x_{nk}^*\|^2 = 1 .$$

Taking into account the compactness of the unit sphere in finite dimensions, we find vectors  $x_i^* \in X = X^*$ ,  $i = 1, \dots, n$ , satisfying the normalization condition in (2.3) and such that  $x_{ik}^* \rightarrow x_i^*$  as  $k \rightarrow \infty$ . Passing to the limit in (2.12), one gets the first condition in (2.3) as well. It follows from representation (1.9) of basic normals in Theorem 1.6 that  $x_i^* \in N(\bar{x}; \Omega_i)$  for all  $i = 1, \dots, n$ . This completes the proof of the exact extremal principle in finite-dimensional spaces.  $\triangle$

**Corollary 2.9 (nontriviality of basic normals in finite dimensions).**

*Let  $\dim X < \infty$ . Then the nontriviality property (2.7) holds for basic normals to every proper closed set  $\Omega \subset X$ .*

**Proof.** Follows from the extremal principle as discussed above. It can also be proved directly by using the definition of boundary points and representation (1.9) in Theorem 1.6.  $\triangle$

The proof of the exact extremal principle given in Theorem 2.8 is essentially based on the *geometry of finite-dimensional spaces*. Namely, it uses the *compactness* of the closed unit ball and the unit sphere as well as *variational properties* of the *Euclidean norm* that have been also exploited above for representation (1.9) of the basic normal cone. An important feature of finite-dimensional spaces is that they always admit a *smooth renorm* (by the Euclidean norm) differentiable away from the origin.

In the next section we justify, based on *variational arguments*, all the three versions of the extremal principle formulated above for a broad class of infinite-dimensional spaces that possess remarkable geometric properties *not related* to the Euclidean norm.

## 2.2 Extremal Principle in Asplund Spaces

The results of this section play a *crucial role* for the whole subsequent material of the book. We start with a direct variational proof of the approximate extremal principle in spaces admitting a *Fréchet smooth renorm*, which form a special subclass of Asplund spaces. Then we develop the method of *separable reduction* for Fréchet-like normals and subgradients that allows us to reduce certain problems involving such constructions in *nonseparable* Banach spaces to separable ones. This method is particularly helpful for the class of *Asplund spaces*, where every separable subspace admits a Fréchet smooth renorm. In such a way we prove the *extremal principle* in Asplund spaces (in both approximate and exact forms) and then establish *variational characterizations* of this class of Banach spaces.

### 2.2.1 Approximate Extremal Principle in Smooth Banach Spaces

In this subsection we pay the main attention to the proof of the approximate extremal principle in Banach spaces that admit *Fréchet smooth renorming*, i.e., an equivalent norm Fréchet differentiable at any nonzero point. It is well known that this class includes every reflexive Banach space; see, e.g., Diestel [332]. Since the prenormal cone  $\hat{N}$  is invariant with respect to equivalent norms on  $X$ , we don't restrict the generality by assuming that  $\|\cdot\|$  is such a smooth norm on  $X$ .

**Theorem 2.10 (approximate extremal principle in Fréchet smooth spaces).** *The approximate extremal principle holds in any space  $X$  admitting a Fréchet smooth renorm.*

**Proof.** We first prove the theorem for the case of two sets and then obtain the general statement by induction. Let  $\bar{x} \in \Omega_1 \cap \Omega_2$  be a local extremal point of some sets  $\Omega_i$  closed around  $\bar{x}$ . We have a neighborhood  $U$  of  $\bar{x}$  such that for any  $\varepsilon > 0$  there is  $a \in X$  with  $\|a\| \leq \varepsilon^3/2$  and  $(\Omega_1 + a) \cap \Omega_2 \cap U = \emptyset$ . Assume for simplicity that  $U = X$  and also that  $\varepsilon < 1/2$ . Then considering the function

$$\varphi(z) := \|x_1 - x_2 + a\| \quad \text{for } z = (x_1, x_2) \in X \times X,$$

we conclude that  $\varphi(z) > 0$  on  $\Omega_1 \times \Omega_2$ , and hence  $\varphi$  is Fréchet differentiable at any point  $z \in \Omega_1 \times \Omega_2$ . In what follows we use the product norm  $\|z\| := (\|x_1\|^2 + \|x_2\|^2)^{1/2}$  that is obviously Fréchet differentiable away from the origin

in  $X \times X$ . Observe the link between the above function  $\varphi$  and the distance function (2.10) used in the proof of the extremal principle in finite dimensions. In contrast to the finite-dimensional proof of Theorem 2.8, now we cannot use the compactness of the unit ball and the Weierstrass existence theorem, which are replaced below by variational arguments based on the completeness of  $X$  and then on the smoothness of the norm.

To proceed, we take  $z_0 := (\bar{x}, \bar{x})$  and form the set

$$W(z_0) := \{z \in \Omega_1 \times \Omega_2 \mid \varphi(z) + \varepsilon \|z - z_0\|^2/2 \leq \varphi(z_0)\}$$

that is nonempty and closed. Moreover, for each  $z \in W(z_0)$  one has

$$\|x_1 - \bar{x}\|^2 + \|x_2 - \bar{x}\|^2 \leq 2\varphi(z_0)/\varepsilon = 2\|a\|/\varepsilon \leq \varepsilon^2,$$

which implies that  $W(z_0) \subset B_\varepsilon(\bar{x}) \times B_\varepsilon(\bar{x})$ . Next let us inductively define sequences of vectors  $z_k \in \Omega_1 \times \Omega_2$  and nonempty closed sets  $W(z_k)$ ,  $k \in \mathbb{N}$ , as follows. Given  $z_k$  and  $W(z_k)$ ,  $k = 0, 1, \dots$ , we select  $z_{k+1} \in W(z_k)$  satisfying

$$\varphi(z_{k+1}) + \varepsilon \sum_{j=0}^k \frac{\|z_{k+1} - z_j\|^2}{2^{j+1}} < \inf_{z \in W(z_k)} \left\{ \varphi(z) + \varepsilon \sum_{j=0}^k \frac{\|z - z_j\|^2}{2^{j+1}} \right\} + \frac{\varepsilon^3}{2^{3k+2}}.$$

Then we form the set

$$\begin{aligned} W(z_{k+1}) &:= \left\{ z \in \Omega_1 \times \Omega_2 \mid \varphi(z) + \varepsilon \sum_{j=0}^{k+1} \frac{\|z - z_j\|^2}{2^{j+1}} \right. \\ &\quad \left. \leq \varphi(z_{k+1}) + \varepsilon \sum_{j=0}^k \frac{\|z_{k+1} - z_j\|^2}{2^{j+1}} \right\}. \end{aligned}$$

It is easy to check that  $\{W(z_k)\}$  is a *nested* sequence of nonempty closed subsets of  $\Omega_1 \times \Omega_2$ . Let us show that  $\text{diam } W(z_k) := \sup \{\|z - w\| \mid z, w \in W(z_k)\} \rightarrow 0$  as  $k \rightarrow \infty$ . Indeed, for each  $z \in W(z_{k+1})$  and  $k \in \mathbb{N}$  we have

$$\begin{aligned} \frac{\varepsilon \|z - z_{k+1}\|^2}{2^{k+2}} &\leq \varphi(z_{k+1}) + \varepsilon \sum_{j=0}^k \frac{\|z_{k+1} - z_j\|^2}{2^{j+1}} - \left( \varphi(z) + \varepsilon \sum_{j=0}^k \frac{\|z - z_j\|^2}{2^{j+1}} \right) \\ &\leq \varphi(z_{k+1}) + \varepsilon \sum_{j=0}^k \frac{\|z_{k+1} - z_j\|^2}{2^{j+1}} - \inf_{z \in W(z_k)} \left\{ \varphi(z) + \varepsilon \sum_{j=0}^k \frac{\|z - z_j\|^2}{2^{j+1}} \right\} < \frac{\varepsilon^3}{2^{3k+2}}, \end{aligned}$$

which implies that  $\text{diam } W(z_k) \leq \varepsilon/2^{k-1} \rightarrow 0$ . Thus (due to the completeness of  $X$ )  $\cap_{k=0}^\infty W(z_k) = \{\bar{z}\}$  with  $z_k \rightarrow \bar{z} = (\bar{x}_1, \bar{x}_2) \in \Omega_1 \times \Omega_2$  as  $k \rightarrow \infty$ . By  $\bar{z} \in W(z_0)$  one has  $\bar{z} \in B_\varepsilon(\bar{x}) \times B_\varepsilon(\bar{x})$ . Let us show that  $\bar{z}$  is a minimum point of the function

$$\phi(z) := \varphi(z) + \varepsilon \sum_{j=0}^\infty \frac{\|z - z_j\|^2}{2^{j+1}}$$

over the set  $\Omega_1 \times \Omega_2$ . Indeed, taking  $\bar{z} \neq z \in \Omega_1 \times \Omega_2$  and using the construction of  $W(z_k)$ , we find  $k \in \mathbb{N}$  such that

$$\varphi(z) + \varepsilon \sum_{j=0}^k \frac{\|z - z_j\|^2}{2^{j+1}} > \varphi(z_k) + \varepsilon \sum_{j=0}^{k-1} \frac{\|z_k - z_j\|^2}{2^{j+1}}. \quad (2.13)$$

This implies that  $\bar{z}$  is a minimum point of  $\phi$  over  $\Omega_1 \times \Omega_2$ , since the sequence on the right-hand side of (2.13) is nonincreasing as  $k \rightarrow \infty$ . Therefore the function  $\psi(z) := \phi(z) + \delta(z; \Omega_1 \times \Omega_2)$  achieves at  $\bar{z}$  its minimum over  $X \times X$ . Thus  $0 \in \hat{\partial}\psi(\bar{z})$  by the generalized Fermat rule of Proposition 1.114. Note that  $\phi$  is Fréchet differentiable at  $\bar{z}$  due to  $\varphi(\bar{z}) \neq 0$  and the smoothness of  $\|\cdot\|^2$ . Now applying the sum rule of Proposition 1.107(i) and then (1.50) as  $\varepsilon = 0$  and the product formula of Proposition 1.2, we get

$$-\nabla\phi(\bar{z}) \in \hat{N}(\bar{z}; \Omega_1 \times \Omega_2) = \hat{N}(\bar{x}_1; \Omega_1) \times \hat{N}(\bar{x}_2; \Omega_2).$$

It follows from the construction of  $\phi$  that  $\nabla\phi(\bar{z}) = (u_1^*, u_2^*) \in X^* \times X^*$ , where

$$u_1^* = x^* + \varepsilon \sum_{j=0}^{\infty} w_{1j}^* \frac{\|\bar{x}_1 - x_{1j}\|}{2^j}, \quad u_2^* = -x^* + \varepsilon \sum_{j=0}^{\infty} w_{2j}^* \frac{\|\bar{x}_2 - x_{2j}\|}{2^j}$$

with  $(x_{1j}, x_{2j}) = z_j$ ,  $x^* = \nabla(\|\cdot\|)(\bar{x}_1 - \bar{x}_2 + a)$ , and

$$w_{ij}^* = \begin{cases} \nabla(\|\cdot\|)(\bar{x}_i - x_{ij}) & \text{if } \bar{x}_i - x_{ij} \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

for  $j = 0, 1, \dots$  and  $i = 1, 2$ . One clearly has  $\sum_{j=0}^{\infty} \|w_{ij}^*\| \cdot \|\bar{x}_i - x_{ij}\|/2^j \leq 1$ ,  $i = 1, 2$ , and  $\|x^*\| = 1$ . Thus putting  $x_i := \bar{x}_i$  and  $x_i^* := (-1)^i x^*/2$  for  $i = 1, 2$ , we arrive at relations (2.3) and (2.4) of the approximate extremal principle in the case of two sets.

Now let us consider the general case of  $n$  sets  $\{\Omega_1, \dots, \Omega_n\}$  in  $X$  and prove the approximate extremal principle by induction when  $n > 2$ . It is easy to see that if  $\bar{x}$  is a local extremal point of  $\{\Omega_1, \dots, \Omega_n\}$ , then the point  $\bar{z} = (\bar{x}, \dots, \bar{x}) \in X^{n-1}$  is locally extremal for the system of two sets

$$A_1 := \Omega_1 \times \dots \times \Omega_{n-1} \quad \text{and} \quad A_2 := \{(x, \dots, x) \in X^{n-1} \mid x \in \Omega_n\},$$

which are closed around  $\bar{z}$  if all  $\Omega_i$  are assumed to be closed around  $\bar{x}$ . It is obvious that  $X^{n-1}$  admits a Fréchet smooth renorm if  $X$  does. Hence we can employ the previous consideration with  $n = 2$  and get the approximate extremal principle for  $\{A_1, A_2, \bar{z}\}$ . In this way, taking into account Proposition 1.2 and the representation

$$\hat{N}(\bar{z}; A_2) = \{(x_1^*, \dots, x_{n-1}^*) \in (X^*)^{n-1} \mid x_1^* + \dots + x_{n-1}^* \in \hat{N}(\bar{x}; \Omega_n)\},$$

we finish the proof of the theorem.  $\triangle$

**Remark 2.11 (bornologically smooth spaces).** The arguments used in the proof of Theorem 2.10 for  $n = 2$  are now typical in the area of *variational principles*; cf. Li and Shi [785] and discussions in the next section. In particular, they can be modified to prove the smooth variational principle of Borwein and Preiss [154] in spaces admitting a smooth renorm with respect to any given bornology on  $X$ . Recall that a *bornology*  $\beta$  on  $X$  is a family of bounded and centrally symmetric subsets of  $X$  whose union is  $X$ , which is closed under multiplication by positive numbers and such that the union of any two members of  $\beta$  is contained in some member of  $\beta$ . The *Fréchet bornology* considered above is the strongest one, where  $\beta$  consists of all bounded symmetric subsets of  $X$ . The weakest one is the *Gâteaux bornology*, where  $\beta$  consists of all finite subsets of  $X$ . It is well known that every separable Banach space admits a Gâteaux smooth renorm. There are useful bornologies in-between; particularly the *Hadamard bornology*, where  $\beta$  consists of all compact symmetric subsets of  $X$ .

One can check that the way of proving Theorem 2.10 allows us to justify the approximate extremal principle (under a suitable modification of generalized normals to nonconvex sets) in Banach spaces admitting a *smooth renorm of any kind*. Actually the corresponding versions of the approximate extremal principle and the smooth variational principle are *equivalent* in Banach spaces with smooth renorms; see Borwein, Mordukhovich and Shao [151] for more details. It will be shown in Section 2.3 that a smoothness of the space in question is not only sufficient and but also *necessary* for the validity of smooth variational principles. On the other hand, the version of the extremal principle in Definition 2.5 will be justified in arbitrary Asplund spaces, which may not admit even a Gâteaux smooth renorm. This is due to the possibility of *separable reduction* for Fréchet-like normals and subgradients considered next.

### 2.2.2 Separable Reduction

In this subsection we develop the method of separable reduction that allows us to reduce certain problems involving Fréchet-like constructions from an arbitrary Banach space to the case of *separable subspaces*. The main goal is to obtain separable reduction results valuable for applications to the extremal principle in the approximate form of Definition 2.5(ii). A suitable assertion for this purpose can be formulated as follows.

*Given proper functions  $f_i: X \rightarrow \overline{\mathbb{R}}$ ,  $i = 1, \dots, N$ , a separable subspace  $Y_0$  of  $X$ , and a number  $M > 0$ , there is a closed separable subspace  $Y$  of  $X$  such that  $Y_0 \subset Y$  and*

$$0 \in (\widehat{\partial} f_1(x_1) \setminus M B^*) + \widehat{\partial} f_2(x_2) + \dots + \widehat{\partial} f_N(x_N) \quad (2.14)$$

*whenever  $x_1, x_2, \dots, x_N \in Y$  and*

$$0 \in (\widehat{\partial} f_{1|Y}(x_1) \setminus M\mathcal{B}^*) + \widehat{\partial} f_{2|Y}(x_2) + \dots + \widehat{\partial} f_{N|Y}(x_N), \quad (2.15)$$

where  $f_{|Y}$  denotes the restriction of  $f$  to  $Y$  and where  $\mathcal{B}^* = \mathcal{B}_{X^*}$ .

This result, being applied to the indicator functions  $f_i(x) := \delta(x; \Omega_i)$ ,  $i = 1, \dots, n$ , with  $f_{n+1}(x) := \varepsilon\|x\|$ , ensures the desired separable reduction of the approximate extremal principle for  $n$  sets from a nonseparable space  $X$  to its separable subspace  $Y$ , provided that the initial subspace  $Y_0$  is properly selected; see below. Note that it is *crucial* to have  $M > 0$  in (2.14) and (2.15) independently from the other data; otherwise we don't get the *nontriviality* condition in the extremal principle.

To justify the desired separable reduction, we have to overcome essential technical difficulties in constructing a separable subspace  $Y_0 \subset Y \subset X$  for the given data. This requires working only with elements of the primal Banach space  $X$ . However, formulations of the extremal principle and the assertion needed for its separable reduction involve elements of the dual space  $X^*$ . Thus an important part of the separable reduction procedure is to translate the required assertion into the language of the *space  $X$  only*. We'll do it first for convex functions, based on the fundamental duality in convex analysis, and then apply to general extended-real-valued functions using some *convexification* via *infimal convolution*, which is possible due to the very definition of Fréchet subgradients.

**Lemma 2.12 (primal characterization of convex subgradients).** *Let  $\varphi: X \rightarrow \overline{\mathbb{R}}$  be a proper convex function with  $0 \in \text{dom } \varphi$ . Then for any given  $M > 0$  one has*

$$\partial\varphi(0) \setminus M\mathcal{B}^* \neq \emptyset \quad (2.16)$$

*if and only if there are  $c \geq 0$ ,  $\gamma > 0$ , and a nonempty open set  $U \subset X$  such that the following properties hold:*

(a)  $\varphi(h) \geq \varphi(0) - c\|h\|$  for all  $h \in X$ ;

(b)  $\varphi(th) \geq \varphi(0) + (M + \gamma)t\|h\|$  whenever  $h \in U$  and  $t \in [0, 1]$ .

*In this case for every  $0 \neq h \in U$  there is  $x^* \in \partial\varphi(0)$  with  $\langle x^*, h \rangle > M\|h\|$ .*

**Proof.** To prove the necessity, we pick any  $x^* \in \partial\varphi(0) \setminus M\mathcal{B}^*$  and observe that (a) holds with  $c = \|x^*\|$ . Then choose  $\gamma > 0$  with  $\|x^*\| > M + \gamma$  and find a nonempty open set  $U \subset X$  such that  $\langle x^*, h \rangle > (M + \gamma)\|h\|$  for every  $h \in U$ . This implies (b).

Let us prove the sufficiency, which includes the last statement of the lemma. Take  $(c, \gamma, U)$  satisfying (a) and (b) and then fix  $0 \neq h \in U$ . By (b) we find nonempty open convex sets  $U_0 \subset U$  and  $U_1 \subset \mathbb{R}$  such that  $0 \notin U_0$ ,  $h \in U_0$ ,  $0 \notin U_1$ , and

$$M < \tau/\|u\| < M + \gamma \quad \text{whenever } (u, \tau) \in U_0 \times U_1.$$

Since  $\varphi$  is convex, we get from (b) that  $\varphi'_+(0)(u) \geq (M + \gamma)\|u\|$  whenever  $u \in U_0$ . Consider the nonempty convex sets

$$C_1 := \{(u, t) \in X \times \mathbb{R} \mid \varphi(u) \leq t\}, \quad C_2 := \bigcup_{\lambda > 0} \lambda(U_0 \times U_1)$$

and observe that  $C_1 \cap C_2 = \emptyset$ . Indeed, if  $\lambda(u, \tau) \in C_1 \cap C_2$  for some  $\lambda > 0$ , then one has

$$\lambda\tau \geq \varphi(\lambda u) \geq \varphi'_+(0)(\lambda u) = \lambda\varphi'_+(0)u \geq (M + \gamma)\lambda\|u\| > \lambda\tau$$

due to the choice of  $\tau$ , a contradiction. Since  $C_2$  is open, we apply the classical separation theorem and find  $(0, 0) \neq (\tilde{x}^*, \tilde{v}) \in (X \times \mathbb{R})^* = X^* \times \mathbb{R}$  such that

$$l := \inf \langle (\tilde{x}^*, \tilde{v}), C_1 \rangle \geq \sup \langle (\tilde{x}^*, \tilde{v}), C_2 \rangle =: r.$$

Note that  $l \leq 0$  due to  $(0, 0) \in C_1$  and that  $r \geq 0$  due to the structure of  $C_2$ . Thus  $l = r = 0$ , and we have

$$\begin{aligned} & \inf \{ \langle \tilde{x}^*, u \rangle + \tilde{v}t \mid (u, t) \in X \times \mathbb{R}, \varphi(u) \leq \varphi(0) + t \} \\ &= \sup \{ \lambda \langle \tilde{x}^*, u \rangle + \lambda\tau\tilde{v} \mid (u, \tau) \in U_0 \times U_1, \lambda > 0 \} = 0. \end{aligned} \quad (2.17)$$

Since  $\tilde{v}t = \langle \tilde{x}^*, 0 \rangle + \tilde{v}t \geq 0$  for all  $t \geq 0$ , we get  $\tilde{v} \geq 0$ . To proceed, we first assume that  $\tilde{v} > 0$ . Then putting  $t = \varphi(u)$  in (2.17), we have  $\langle -\tilde{x}^*/\tilde{v}, u \rangle \leq \varphi(u) = \varphi(u) - \varphi(0)$  if  $u \in \text{dom } \varphi$ . This also obviously holds if  $\varphi(u) = \infty$ , and so we conclude that  $-\tilde{x}^*/\tilde{v} \in \partial\varphi(0)$ .

On the other hand, it follows from (2.17) for  $\tau \in U_1$  and  $u = h$  that  $\langle \tilde{x}^*, h \rangle + \tau\tilde{v} \leq 0$ , and hence

$$\| -\tilde{x}^*/\tilde{v} \| \geq \langle -\tilde{x}^*/\tilde{v}, h/\|h\| \rangle \geq \tau/\|h\| > M$$

due to the choice of  $\tau$ . Thus we obtain

$$\langle -\tilde{x}^*/\tilde{v}, h \rangle > M\|h\| \quad \text{and} \quad -\tilde{x}^*/\tilde{v} \in \partial\varphi(0) \setminus M\mathcal{B}^*,$$

which justifies (2.16) in the case of  $\tilde{v} > 0$ . We haven't used (a) so far.

Next let us consider the remaining case of  $\tilde{v} = 0$  in (2.17) and justify (2.16) using (a). In this case we necessarily have  $\tilde{x}^* \neq 0$  and get from (2.17) that  $\langle \tilde{x}^*, u \rangle \geq 0$  for all  $u \in \text{dom } \varphi$  and  $\langle \tilde{x}^*, u \rangle \leq 0$  for all  $u \in U_0$ . Since  $U_0$  is a neighborhood of  $h$ , the latter yields  $\langle \tilde{x}^*, h \rangle < 0$ . Form the closed convex set

$$C_3 := \{(u, t) \in X \times \mathbb{R} \mid t < -c\|u\|\}$$

and observe that  $C_1 \cap C_3 = \emptyset$  due to (a). Employing again the separation theorem, we find  $(0, 0) \neq (\hat{x}^*, \hat{v}) \in X^* \times \mathbb{R}$  such that

$$l := \inf \langle (\hat{x}^*, \hat{v}), C_1 \rangle \geq \sup \langle (\hat{x}^*, \hat{v}), C_3 \rangle =: r.$$

It is easy to check that  $l = r = 0$ , and thus

$$\begin{aligned} & \inf \{ \langle \hat{x}^*, u \rangle + \hat{v}t \mid (u, t) \in X \times \mathbb{R}, \varphi(u) \leq \varphi(0) + t \} \\ &= \sup \{ \langle \hat{x}^*, u \rangle + \hat{v}t \mid (u, t) \in X \times \mathbb{R}, t < -c\|u\| \} = 0, \end{aligned} \quad (2.18)$$

which implies that  $\hat{v} \geq 0$ . In fact we have  $\hat{v} > 0$ , since otherwise (2.18) yields  $\langle \hat{x}^*, u \rangle \leq 0$  whenever  $u \in X$ , which contradicts the nontriviality of  $(\hat{x}^*, \hat{v})$ . Thus (2.18) gives  $-\hat{x}^*/\hat{v} \in \partial\varphi(0)$  similarly to the case of (2.17). Now put

$$x^* := -\hat{x}^*/\hat{v} - K\tilde{x}^* \quad \text{with } K > \max \left\{ 0, -\frac{M\|h\| + \langle \hat{x}^*/\hat{v}, h \rangle}{\langle \tilde{x}^*, h \rangle} \right\} \quad (2.19)$$

and observe that, by the definition of  $\partial\varphi(0)$  and the condition  $\langle \tilde{x}^*, u \rangle \geq 0$  for all  $u \in \text{dom } \varphi$ , we have

$$\varphi(u) - \varphi(0) \geq \langle -\hat{x}^*/\hat{v}, u \rangle \geq \langle x^*, u \rangle \quad \text{if } u \in \text{dom } \varphi;$$

so  $x^* \in \partial\varphi(0)$ . Moreover, using (2.19) and  $\langle \tilde{x}^*, h \rangle < 0$ , we conclude that

$$\langle x^*, h \rangle = \langle -\hat{x}^*/\hat{v}, h \rangle - K\langle \tilde{x}^*, h \rangle > M\|h\|,$$

which yields  $\|x^*\| > M$  and hence (2.16).  $\triangle$

The next lemma provides a primal characterization of subdifferential sums for convex functions with a nontriviality condition crucial for subsequent applications to the extremal principle.

**Lemma 2.13 (primal characterization of subdifferential sums for convex functions).** *Let  $\varphi_j: X \rightarrow \overline{\mathbb{R}}$ ,  $j = 1, \dots, N$ , be proper convex functions with  $0 \in \text{dom } \varphi_1 \cap \dots \cap \text{dom } \varphi_N$  and  $N > 1$ . Given any  $M > 0$ , one has*

$$0 \in (\partial\varphi_1(0) \setminus M\mathcal{B}^*) + \partial\varphi_2(0) + \dots + \partial\varphi_N(0) \quad (2.20)$$

*if and only if there are  $c \geq 0$ ,  $\gamma > 0$  and a nonempty open set  $U \subset X$  such that the following hold:*

(a)  $\sum_{j=1}^N \varphi_j(h_j) \geq \sum_{j=1}^N \varphi_j(0) - c \max \{ \|h_j - h_1\| \mid j = 2, \dots, N \}$  for all  $h_1, \dots, h_N \in X$ ;

(b)  $\sum_{j=1}^N \varphi_j(th_j) \geq \sum_{j=1}^N \varphi_j(0) + (M + \gamma)t \max \{ \|h_j - h_1\| \mid j = 2, \dots, N \}$  for all  $h_1, \dots, h_N \in X$  with  $h_j - h_1 \in U$ ,  $j = 2, \dots, N$ , and for all  $t \in [0, 1]$ .

**Proof.** Assume that (2.20) holds and find  $x_j^* \in \partial\varphi_j(0)$ ,  $j = 1, \dots, N$ , such that  $\|x_1^*\| > M$  and  $x_1^* + \dots + x_N^* = 0$ . Then

$$\begin{aligned} & \sum_{j=1}^N \varphi_j(h_j) - \sum_{j=1}^N \varphi_j(0) \geq \sum_{j=1}^N \langle x_j^*, h_j \rangle = \sum_{j=2}^N \langle x_j^*, h_j - h_1 \rangle \\ & \geq - \sum_{j=2}^N \|x_j^*\| \max \{ \|h_j - h_1\| \mid j = 2, \dots, N \} \end{aligned}$$



for all  $h_1, \dots, h_N \in X$ , which gives (a) with  $c := \sum_{j=2}^N \|x_j^*\|$ . To justify (b), we take  $\gamma > 0$  and an open set  $\emptyset \neq U \subset X$  such that

$$\sum_{j=2}^N \langle x_j^*, h \rangle = -\langle x_1^*, h \rangle > (M + \gamma) \|h\| \quad \text{for all } h \in U.$$

By diminishing  $U$  if necessary, we may assume that

$$\sum_{j=2}^N \langle x_j^*, h_j \rangle > (M + \gamma) \max \{ \|h_j\| \mid j = 2, \dots, N \}$$

whenever  $h_2, \dots, h_N \in U^{N-1}$ . Then

$$\begin{aligned} \varphi_1(th_1) + \sum_{j=2}^N \varphi_j(th_j) - \sum_{j=1}^N \varphi_j(0) &\geq t \sum_{j=2}^N \langle x_j^*, h_j - h_1 \rangle \\ &\geq (M + \gamma)t \max \{ \|h_j - h_1\| \mid j = 2, \dots, N \} \end{aligned}$$

whenever  $h_1, \dots, h_N \in X$  with  $h_j - h_1 \in U$ ,  $j = 2, \dots, N$ , and  $t \in [0, 1]$ . This gives (b) and proves the necessity in the lemma.

To prove the sufficiency, we assume that  $c$ ,  $\gamma$ , and  $U$  are such that (a) and (b) hold. Define the *inf-convolution*

$$\varphi(h_2, \dots, h_N) := \inf \left\{ \varphi_1(x) + \sum_{j=2}^N \varphi_j(x + h_j) \mid x \in X \right\}$$

for  $(h_2, \dots, h_N) \in X^{N-1}$  and observe that  $\varphi$  is a proper convex function on  $X^{N-1}$  with  $0 \in \text{dom } \varphi$ . It is easy to check that properties (a) and (b) of this lemma implies that  $\varphi$  satisfies properties (a) and (b) of Lemma 2.12 on the product space  $X^{N-1}$  with the norm  $\|(h_2, \dots, h_N)\| := \max \{ \|h_j\| \mid j = 2, \dots, N \}$ . Thus for fixed  $0 \neq h \in U$  we find  $z^* := (x_2^*, \dots, x_N^*) \in (X^{N-1})^*$  such that  $z^* \in \partial\varphi(0, \dots, 0)$  and  $\langle z^*, (h, \dots, h) \rangle > M \max \{ \|h\|, \dots, \|h\| \}$ , i.e.,

$$\left\langle \sum_{j=2}^N x_j^*, h \right\rangle > M \|h\|. \quad (2.21)$$

Since  $z^* \in \partial\varphi(0)$ , the definition of  $\varphi$  gives

$$\varphi_1(x) + \sum_{j=2}^N \varphi_j(x + h_j) \geq \sum_{j=1}^N \varphi_j(0) + \langle z^*, (h_2, \dots, h_N) \rangle = \sum_{j=1}^N \varphi_j(0) + \sum_{j=2}^N \langle x_j^*, h_j \rangle$$

for all  $x \in X$  and all  $(h_2, \dots, h_N) \in X^{N-1}$ . If we fix here one  $j$  and put  $h_i = x = 0$  for all  $i \neq j$ , we get  $x_j^* \in \partial\varphi_j(0)$ ,  $j = 2, \dots, N$ . If we put  $h_j = -x$ ,  $j = 2, \dots, N$ , we get  $x^* := -(x_2^* + \dots + x_N^*) \in \partial\varphi_1(0)$ . Hence

$$0 \in \partial\varphi_1(0) + \dots + \partial\varphi_N(0) \quad \text{and} \quad x^* \in \partial\varphi_1(0) \setminus M\mathcal{B}_{X^*}$$

due to (2.21), which completes the proof of the lemma.  $\triangle$

Now let us consider a general proper function  $f: X \rightarrow \overline{\mathbb{R}}$ , a point  $x \in \text{dom } f$  and associated with them two *convex* functions of the inf-convolution type. First, given positive numbers  $\delta$  and  $\epsilon$ , we define  $\varphi_{f,x,\delta,\epsilon}: X \rightarrow [-\infty, \infty]$  by

$$\begin{aligned} \varphi_{f,x,\delta,\epsilon}(h) := \inf \left\{ \sum_{i=1}^m \alpha_i \left[ f(x + h_i) + \epsilon \|h_i\| \right] \mid m \in \mathbb{N}, h_i \in X, \right. \\ \left. \|h_i\| < \delta, \alpha_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m \alpha_i = 1, \sum_{i=1}^m \alpha_i h_i = h \right\} \end{aligned} \quad (2.22)$$

if  $\|h\| < \delta$  and  $\varphi_{f,x,\delta,\epsilon}(h) := \infty$  otherwise. Then, given a sequence  $\Delta := (\delta_i)_{i=1}^\infty$  with  $\delta_1 > \delta_2 > \dots > 0$  and  $\delta_i \downarrow 0$ , we define  $\varphi_{f,x,\Delta}: X \rightarrow \overline{\mathbb{R}}$  by

$$\begin{aligned} \varphi_{f,x,\Delta}(h) := \inf \left\{ \sum_{i=1}^m \alpha_i \varphi_{f,x,\delta_i,1/i}(h_i) \mid m \in \mathbb{N}, h_i \in X, \right. \\ \left. \alpha_i \geq 0, i = 1, \dots, m, \sum_{i=1}^m \alpha_i = 1, \sum_{i=1}^m \alpha_i h_i = h \right\}, \end{aligned} \quad (2.23)$$

where each  $\varphi_{f,x,\delta_i,1/i}$ ,  $i \in \mathbb{N}$ , is constructed in (2.22). It follows from the definitions that both functions (2.22) and (2.23) are convex and not greater than  $f(x)$  at  $h = 0$ . Moreover, the Fréchet subdifferential of  $f$  at  $x$  is closely related to the subdifferential of  $\varphi_{f,x,\Delta}$  at zero. One can easily check that if  $\widehat{\partial}f(x) \neq \emptyset$ , then  $\varphi_{f,x,\Delta}(0) = f(x)$  and  $\widehat{\partial}f(x) \supset \partial\varphi_{f,x,\Delta}(0) \neq \emptyset$  for some  $\Delta$ . On the other hand, if  $\partial\varphi_{f,x,\Delta}(0) \neq \emptyset$  for some  $\Delta$  and  $\varphi_{f,x,\Delta}(0) = f(x)$ , then  $\partial\varphi_{f,x,\Delta}(0) \subset \widehat{\partial}f(x)$  as well.

The following corollary of Lemma 2.13 provides an equivalent translation of the basic assertion (2.14) into the language of the primal space  $X$ .

**Corollary 2.14 (primal characterization for sums of Fréchet subdifferentials).** *Let  $f_j: X \rightarrow \overline{\mathbb{R}}$  be arbitrary proper functions, let  $x_j \in \text{dom } f_j$  as  $= 1, \dots, N$  and  $N > 1$ . Then for any given  $M > 0$  one has (2.14) if and only if there are  $c \geq 0$ ,  $\gamma > 0$ , a sequence  $\Delta = (\delta_i)_{i=1}^\infty \subset (0, \infty)$  with  $\delta_i \downarrow 0$ , and a nonempty open set  $U \subset X$  such that the following hold:*

- (a)  $\sum_{j=1}^N \varphi_{f_j,x_j,\Delta}(h_j) \geq \sum_{j=1}^N f_j(x_j) - c \max \{ \|h_j - h_1\| \mid j = 2, \dots, N \}$  for all  $h_1, \dots, h_N \in X$ ;
- (b)  $\sum_{j=1}^N \varphi_{f_j,x_j,\Delta}(th_j) \geq \sum_{j=1}^N f_j(x_j) + (M + \gamma)t \max \{ \|h_j - h_1\| \mid j = 2, \dots, N \}$  for all  $h_1, \dots, h_N \in X$  with  $h_j - h_1 \in U$ ,  $j = 2, \dots, N$ , and for all numbers  $t \in [0, 1]$ .

**Proof.** If (2.14) holds, then  $\widehat{\partial}f_j(x_j) \neq \emptyset$ , and hence  $\varphi_{f,x,\Delta}(0) = f_j(x_j)$ ,  $j = 1, \dots, N$ , for some sequence  $\Delta$ . Then conditions (a) and (b) of the corollary immediately follow from the corresponding conditions of Lemma 2.13. In the other direction, if conditions (a) and (b) of the corollary hold, then  $\varphi_{f_j,x_j,\Delta}(0) = f_j(x_j)$  by (a), and so (2.14) follows from the sufficiency in Lemma 2.13 for the convex functions  $\varphi_j = \varphi_{f_j,x_j,\Delta}$ ,  $j = 1, \dots, N$ , and the mentioned relationships between  $\widehat{\partial}f(x)$  and  $\partial\varphi_{f,x,\Delta}(0)$ .  $\triangle$

Next we establish the basic separable reduction result for assertion (2.14) that lies at the ground of the whole separable reduction technique for the extremal principle.

**Theorem 2.15 (basic separable reduction).** *Let  $f_1, \dots, f_N: X \rightarrow \overline{\mathbb{R}}$ ,  $N > 1$ , be proper functions bounded from below, and let  $Y_0$  be a separable subspace of  $X$ . Then there is a closed separable subspace  $Y \subset X$  such that  $Y_0 \subset Y$  and, given any  $M > 0$ , assertion (2.14) holds whenever  $x_1, x_2, \dots, x_N \in Y$  and one has (2.15).*

**Proof.** Our strategy is to build  $Y$  inductively starting with  $Y_0$  and then to derive (2.14) from (2.15) and  $(x_1, \dots, x_N) \in Y^N$  based on the primal characterization of (2.14) in Corollary 2.14.

Let  $\mathcal{A}$  be the countable set of all matrices  $(\alpha_i^j \mid i \in \mathbb{N}, j = 1, \dots, N)$  with rational nonnegative entries such that  $\alpha_i^j > 0$  only for finitely many pairs  $(i, j) \in \mathbb{N} \times \{1, \dots, N\}$  and that  $\sum_{i=1}^{\infty} \alpha_i^j = 1$  for all  $j = 1, \dots, N$ . Let  $\mathcal{B}$  be the countable set of all matrices  $(\beta_{il}^j \mid i, l \in \mathbb{N}, j = 1, \dots, N)$  with rational nonnegative entries such that  $\beta_{il}^j > 0$  only for finitely many triples  $(i, l, j) \in \mathbb{N}^2 \times \{1, \dots, N\}$  and that  $\sum_{l=1}^{\infty} \beta_{il}^j = 1$  for all  $i \in \mathbb{N}$  and  $j = 1, \dots, N$ . Let  $\mathcal{D}$  be the countable set of all sequences  $(\delta_i)_{i=1}^{\infty}$  with rational entries for which  $0 < \delta_1 \geq \delta_2 \geq \dots \geq 0$  and  $\delta_i = 0$  if  $i \in \mathbb{N}$  is sufficiently large. Given  $j = 1, \dots, N$  and  $x \in \text{dom } f_j$ , let  $\eta_j(x) > 0$  be such that  $f_j$  is bounded from below on the ball around  $x$  with radius  $\eta_j(x)$ .

For  $\bar{x} := (x_1, \dots, x_N) \in X^N$ , for  $a := (\alpha_i^j) \in \mathcal{A}$ , for  $b := (\beta_{il}^j) \in \mathcal{B}$ , for  $\bar{r} := (r_2, \dots, r_N) \in (0, \infty)^{N-1}$ , for  $\Delta := (\delta_i) \in \mathcal{D}$  satisfying  $\delta_i > 0$  whenever  $\max\{\alpha_i^1, \dots, \alpha_i^N\} > 0$  and  $\delta_1 < \min\{\eta_1(x_1), \dots, \eta_N(x_N)\}$ , and for  $k \in \mathbb{N}$  we find  $u_{il}^j(\bar{x}, a, b, \bar{r}, \Delta, k) \in X$ ,  $i, l \in \mathbb{N}$ ,  $j = 1, \dots, N$ , such that  $\|u_{il}^j(\bar{x}, a, b, \bar{r}, \Delta, k)\| < \delta_i$  if  $\delta_i > 0$  and  $u_{il}^j(\dots) = 0$  if  $\delta_i = 0$  for all  $i, l \in \mathbb{N}$  and  $j = 1, \dots, N$ , that

$$\left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j u_{il}^j(\bar{x}, a, b, \bar{r}, \Delta, k) - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 u_{il}^1(\dots) \right\| < r_j, \quad j = 2, \dots, N,$$

and that

$$\begin{aligned} & \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + u_{il}^j(\bar{x}, a, b, \bar{r}, \Delta, k)) + \frac{1}{i} \|u_{il}^j(\bar{x}, a, b, \bar{r}, \Delta, k)\| \right] \\ & < \frac{1}{k} + \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i} \|h_{il}^j\| \right] \end{aligned}$$

whenever  $h_{il}^j \in X$ ,  $\|h_{il}^j\| < \delta_i$  if  $\delta_i > 0$  and  $h_{il}^j = 0$  if  $\delta_i = 0$ , and

$$\left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j h_{il}^j - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 h_{il}^1 \right\| < r_j, \quad j = 2, \dots, N.$$

Further, for  $\bar{x}, a, b, \bar{r}, \Delta, k$  as above and for  $h \in X$  with  $\|h\| < \delta_1$  we find  $g_{il}^j(\bar{x}, h, a, b, \bar{r}, \Delta, k) \in X$ ,  $i, l \in \mathbb{N}$ ,  $j = 1, \dots, N$ , such that

$$\|g_{il}^j(\bar{x}, h, a, b, \bar{r}, \Delta, k)\| < \delta_i \quad \text{if } \delta_i > 0 \quad \text{and} \quad g_{il}^j(\dots) = 0 \quad \text{if } \delta_i = 0,$$

$$\left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j g_{il}^j(\bar{x}, h, a, b, \bar{r}, \Delta, k) - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 g_{il}^1(\dots) - h \right\| < r_j$$

if  $j = 2, \dots, N$ , and that

$$\begin{aligned} & \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + g_{il}^j(\bar{x}, h, a, b, \bar{r}, \Delta, k)) + \frac{1}{i} \|g_{il}^j(\bar{x}, h, a, b, \bar{r}, \Delta, k)\| \right] \\ & < \frac{1}{k} + \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i} \|h_{il}^j\| \right] \end{aligned}$$

whenever  $h_{il}^j \in X$ ,  $\|h_{il}^j\| < \delta_i$  if  $\delta_i > 0$  and  $h_{il}^j = 0$  if  $\delta_i = 0$ , and

$$\left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j h_{il}^j - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 h_{il}^1 - h \right\| < r_j, \quad j = 2, \dots, N.$$

Now we are ready to construct the required separable subspace  $Y \subset X$ . By induction we build separable subspaces  $Y_0 \subset Y_1 \subset \dots \subset X$  as follows. If  $Y_n$  was already constructed for some  $n \in \mathbb{N} \cup \{0\}$  ( $Y_0$  is given), take any countable subset  $C_n \subset Y_n$  dense in  $Y_n$ . Then let  $Y_{n+1}$  be the closed linear span of  $Y_n$  and the points

$$u_{il}^j(\bar{x}, a, b, \bar{r}, \Delta, k), \quad g_{il}^j(\bar{x}, h, a, b, \bar{r}, \Delta, k),$$

where  $\bar{x} = (x_1, \dots, x_N) \in C_n$ ,  $h \in C_n$ ,  $\|h\| < \delta_1$ ,  $\bar{r} \in (0, \infty)^{N-1}$  with rational entries,  $\Delta = (\delta_i) \in \mathcal{D}$  with  $\delta_1 < \min\{\eta_1(x_1), \dots, \eta_N(x_N)\}$ ,  $a \in \mathcal{A}$ ,  $b \in \mathcal{B}$ ,  $j = 1, \dots, N$ , and  $i, l, k \in \mathbb{N}$ . Denoting  $Y := \text{cl}[\bigcup\{Y_n \mid n \in \mathbb{N}\}]$  and

$C := \bigcup \{C_n \mid n \in \mathbb{N}\}$ , we see that  $\text{cl } C = Y$  and  $Y$  is a separable subspace of  $X$  containing  $Y_0$ .

Fix any  $M > 0$ . We need to prove that for every given  $\bar{x} = (x_1, \dots, x_N) \in Y^N$  satisfying (2.15) one has (2.14). According to Corollary 2.14 the latter is equivalent to the fulfillment of conditions (a) and (b) therein. Using (2.15), we find  $x_j^* \in \widehat{\partial}(f_j|_Y)(x_j)$ ,  $j = 1, \dots, N$ , such that  $\|x_1^*\| > M$  and  $x_1^* + \dots + x_N^* = 0$ . Due to the definition of Fréchet subgradients there is a sequence of rational numbers  $\delta_1 > \delta_2 > \dots > 0$  with

$$f_j(x_j + h) + \frac{1}{i}\|h\| \geq f_j(x_j) + \langle x_j^*, h \rangle \quad \text{whenever } h \in Y, \quad \|h\| < 2\delta_i, \quad (2.24)$$

$i \in \mathbb{N}$ , and  $j = 1, \dots, N$ . We always take  $\delta_1 < \min \{\eta_1(x_1), \dots, \eta_N(x_N)\}$  and show that conditions (a) and (b) of Corollary 2.14 hold along the chosen sequence  $\Delta = \{\delta_1, \delta_2, \dots\}$ . Since  $\bar{x} \in Y^N$ , for any  $n \in \mathbb{N}$  and  $j = 1, \dots, N$  we find  $x_n^j \in C_n \subset Y$  and rational numbers  $\gamma_n^j$  satisfying

$$\|x_j - x_n^j\| \leq \gamma_n^j \leq 2\|x_j - x_n^j\| \quad \text{and} \quad \|x_j - x_n^j\| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

First we verify condition (a) of Corollary 2.14 with  $c := \sum_{j=2}^N \|x_j^*\|$ . Fix any  $h_1, \dots, h_N \in X$  and assume without loss of generality that  $\|h_j\| < \delta_1$  for all  $j = 1, \dots, N$ . Consider any  $a = (\alpha_i^j) \in \mathcal{A}$ , any  $b = (\beta_{il}^j) \in \mathcal{B}$ , any  $h_{il}^j \in X$  with  $\|h_{il}^j\| < \delta_i$ ,  $i, l \in \mathbb{N}$ ,  $j = 1, \dots, N$ , such that

$$\sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j h_{il}^j = h_j \quad \text{for all } j = 1, \dots, N. \quad (2.25)$$

Find  $i_0 \in \mathbb{N}$  so large that  $\alpha_i^j = 0$  for all  $i \geq i_0$  and  $j = 1, \dots, N$ . Then we put  $h_{il}^j = 0$  whenever  $i \geq i_0$ . Taking any rational numbers  $r_j > \|h_j - h_1\|$ ,  $j = 2, \dots, N$ , we observe that

$$\|h_{il}^j\| + \gamma_n^j < \delta_i, \quad i < i_0, \quad l \in \mathbb{N}, \quad j = 1, \dots, N, \quad (2.26)$$

$$\text{and } \|h_j - h_1\| + \gamma_n^j + \gamma_n^1 < r_j, \quad j = 2, \dots, N$$

for all  $n \in \mathbb{N}$  sufficiently large. Denote  $\bar{x}_n := (x_n^1, \dots, x_n^N)$ ,  $n \in \mathbb{N}$ , and

$$h_{il}^{j,n} := h_{il}^j + x_j - x_n^j, \quad i, l \in \mathbb{N}, \quad j = 1, \dots, N. \quad (2.27)$$

Finally, putting  $\bar{\Delta} := (\delta_1, \delta_2, \dots, \delta_{i_0}, 0, 0, \dots)$  and using the  $u_{il}^j$ -part in the construction of  $Y$ , we get the following chain of inequalities valid for all large numbers  $n \in \mathbb{N}$ :

$$\sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i}\|h_{il}^j\| \right] = \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_n^j + h_{il}^{j,n}) \right]$$

$$\begin{aligned}
& + \frac{1}{i} \|h_{il}^j\| \Big] \geq \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_n^j + h_{il}^{j,n}) + \frac{1}{i} \|h_{il}^{j,n}\| \right] - \frac{1}{i} \sum_{j=1}^N \gamma_n^j \\
& > -\frac{1}{n} - \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_n^j + u_{il}^j(\bar{x}_n, a, b, \bar{r}, \bar{\Delta}, n) \right. \\
& \quad \left. + \frac{1}{i} \|u_{il}^j(\dots)\| \right] \quad \left( \text{as } \|h_{il}^{j,n}\| \leq \|h_{il}^j\| + \gamma_n^j < \delta_i, \text{ if } i \leq i_0, \text{ and} \right. \\
& \quad \left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j h_{il}^{n,j} - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 h_{il}^{1,n} \right\| \leq \|h_j - h_1\| + \gamma_n^j + \gamma_n^1 < r_j \Big) \\
& \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + x_n^j - x_j \right. \\
& \quad \left. + u_{il}^j(\bar{x}_n, a, b, \bar{r}, \bar{\Delta}, n) + \frac{1}{i} \|x_n^j - x_j + u_{il}^j(\dots)\| \right] \\
& \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) + \sum_{j=1}^N \left\langle x_j^*, x_n^j - x_j \right. \\
& \quad \left. + \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j u_{il}^j(\bar{x}_n, a, b, \bar{r}, \bar{\Delta}, n) \right\rangle \\
& \quad \left( \text{as } x_n^j - x_j + u_{il}^j(\dots) \in Y \text{ and } \|x_n^j - x_j + u_{il}^j(\dots)\| < \gamma_n^j + \delta_i < 2\delta_i \right) \\
& = -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) + \sum_{j=1}^N \langle x_j^*, x_n^j - x_j \rangle \\
& \quad + \sum_{j=2}^N \left\langle x_j^*, \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j u_{il}^j(\bar{x}_n, a, b, \bar{r}, \bar{\Delta}, n) - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 u_{il}^1(\dots) \right\rangle \\
& \quad \left( \text{as } x_1^* + x_2^* + \dots + x_N^* = 0 \right) \\
& \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) - \sum_{j=1}^N \|x_j^*\| \gamma_n^j \\
& \quad - \sum_{j=2}^N \|x_j^*\| \left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j u_{il}^j(\bar{x}_n, a, b, \bar{r}, \bar{\Delta}, n) - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 u_{il}^1(\dots) \right\| \\
& \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) - \sum_{j=1}^N \|x_j^*\| \gamma_n^j - \sum_{j=2}^N \|x_j^*\| r_j .
\end{aligned}$$

Letting  $n \rightarrow \infty$ , we get the estimate

$$\sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i} \|h_{il}^j\| \right] \geq \sum_{j=1}^N f_j(x_j) - \sum_{j=2}^N \|x_j^*\| r_j .$$

Then letting  $r_j \rightarrow \tilde{r}_j := \|h_j - h_1\|$  for  $j = 2, \dots, N$ , we arrive at

$$\sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i} \|h_{il}^j\| \right] \geq \sum_{j=1}^N f_j(x_j) - c \max \{ \tilde{r}_j \mid j = 2, \dots, N \},$$

which ensures condition (a) of Corollary 2.14 with  $c := \sum_{j=2}^N \|x_j^*\|$  due to the definition of  $\varphi_{f_j, x_j, \Delta}$  in (2.23) along the sequence  $\Delta$  selected in (2.24).

To complete the proof of the theorem, it remains to verify condition (b) in Corollary 2.14 along the sequence  $\Delta$ , some number  $\gamma > 0$ , and an open set  $U \subset X$ . Since  $\|x_1^*\| > M$ , we find  $y \in Y$  with  $\|y\| \leq \delta_1$  and  $\gamma \in (0, 1)$  so that

$$-\langle x_1^*, y \rangle > (M + 3\gamma)\|y\|. \quad (2.28)$$

Choose a number  $\zeta$  satisfying

$$0 < \zeta < \min \left\{ \delta_1 - \|y\|, \gamma \|y\| \left( \sum_{j=1}^N \|x_j^*\| \right)^{-1}, \gamma \|y\| \left[ 2(M + \gamma) \right]^{-1} \right\} \quad (2.29)$$

and put  $U := \{h \in X \mid \|h - y\| < \zeta\}$ . Now fix any  $t \in (0, 1]$  and any  $h_1, \dots, h_N \in X$  with  $h_j - h_1 \in U$ ; then  $\|h_j - h_1\| < \delta$ ,  $j = 2, \dots, N$ . We may assume without loss of generality that  $\|th_j\| \leq \delta_1$  for all  $j = 1, \dots, N$ . Since  $\|h_j - h_1 - y\| < \zeta$ , there is a rational number  $\eta$  with  $\|th_j - th_1 - ty\| < \eta < t\zeta$  for all  $j = 2, \dots, N$ . This allows us to find  $h_0 \in C$  such that

$$\|th_j - th_1 - h_0\| < \eta, \quad j = 2, \dots, N, \quad \text{and} \quad \|h_0 - ty\| < t\zeta. \quad (2.30)$$

As in the proof of the first part of the theorem, we pick any  $a = (\alpha_i^j) \in \mathcal{A}$ , any  $b = (\beta_{il}^j) \in \mathcal{B}$ , and any  $h_{il}^j \in X$ , with  $\|h_{il}^j\| < \delta_i$ ,  $i, l \in \mathbb{N}$ ,  $j = 1, \dots, N$ , and such that (2.25) holds. Find  $i_0 \in \mathbb{N}$  so large that  $\alpha_i^j = 0$  whenever  $i \geq i_0$  and  $j = 1, \dots, N$ . We may choose  $h_{il}^j = 0$  whenever  $i \geq i_0$ . Thus we have (2.26) for all large  $n \in \mathbb{N}$ . Take  $\overline{\Delta} = (\delta_1, \delta_2, \dots, \delta_{i_0}, 0, 0, \dots)$ , define  $\overline{x}_n$  and  $h_{il}^{j,n}$  as in (2.27), and put  $\overline{r}_n := (\eta + \gamma_n^2 + \gamma_n^1, \dots, \eta + \gamma_n^N + \gamma_n^1)$ . Now using the  $g_{il}^j$ -part in the construction of  $Y$ , we perform the following chain of inequalities for all  $n \in \mathbb{N}$  sufficiently large:

$$\begin{aligned} & \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i} \|h_{il}^j\| \right] \\ & \geq \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_n^j + h_{il}^{j,n}) + \frac{1}{i} \|h_{il}^{j,n}\| \right] - \frac{1}{i} \sum_{j=1}^N \gamma_n^j \\ & > -\frac{1}{n} - \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_n^j + g_{il}^j(\overline{x}_n, h_0, a, b, \overline{r}_n, \overline{\Delta}, n)) \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{i} \|g_{il}^j(\dots)\| \Bigg] \left( \text{as } \|h_{il}^{j,n}\| \leq \|h_{il}^j\| + \gamma_n^j < \delta_i, \ i \leq i_0, \text{ and } \left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j h_{il}^{n,j} \right. \right. \\
& \left. \left. - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 h_{il}^{1,n} - h_0 \right\| \leq \|th_j - th_1 - h_0\| + \gamma_n^j + \gamma_n^1 < \eta + \gamma_n^j + \gamma_n^1 \right) \\
& \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + x_n^j - x_j \right. \\
& \left. + g_{il}^j(\bar{x}_n, h_0, a, b, \bar{r}_n, \bar{\Delta}, n)) + \frac{1}{i} \|x_n^j - x_j + g_{il}^j(\dots)\| \right] \\
& \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) \\
& + \sum_{j=1}^N \left\langle x_j^*, x_n^j - x_j + \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j g_{il}^j(\bar{x}_n, h_0, a, b, \bar{r}_n, \bar{\Delta}, n) \right\rangle \\
& \left( \text{as } x_n^j - x_j + g_{il}^j(\dots) \in Y \text{ and } \|x_n^j - x_j + g_{il}^j(\dots)\| < \gamma_n^j + \delta_i < 2\delta_i \right) \\
& = -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) + \sum_{j=1}^N \langle x_j^*, x_n^j - x_j \rangle - \langle x_1^*, h_0 \rangle \\
& + \sum_{j=2}^N \left\langle x_j^*, \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j g_{il}^j(\bar{x}_n, h_0, a, b, \bar{r}_n, \bar{\Delta}, n) \right. \\
& \left. - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 g_{il}^1(\dots) - h_0 \right\rangle \left( \text{as } x_1^* + x_2^* + \dots + x_N^* = 0 \right) \\
& \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) - \sum_{j=1}^N \|x_j^*\| \gamma_n^j - \langle x_1^*, h_0 \rangle \\
& - \sum_{j=2}^N \|x_j^*\| \left\| \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j g_{il}^j(\bar{x}_n, h_0, a, b, \bar{r}_n, \bar{\Delta}, n) \right. \\
& \left. - \sum_{i=1}^{\infty} \alpha_i^1 \sum_{l=1}^{\infty} \beta_{il}^1 g_{il}^1(\dots) - h_0 \right\| \geq -\frac{1}{n} - 2 \sum_{j=1}^N \gamma_n^j + \sum_{j=1}^N f_j(x_j) - \sum_{j=1}^N \|x_j^*\| \gamma_n^j \\
& - \langle x_1^*, h_0 \rangle - \sum_{j=2}^N \|x_j^*\| (\eta + \gamma_n^j + \gamma_n^1) .
\end{aligned}$$

Letting  $n \rightarrow \infty$ , we get

$$\sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i} \|h_{il}^j\| \right] \geq \sum_{j=1}^N f_j(x_j) - \langle x_1^*, h_0 \rangle - \sum_{j=2}^N \|x_j^*\| \eta .$$



Now using (2.28)–(2.30), we finally have

$$\begin{aligned}
& \sum_{j=1}^N \sum_{i=1}^{\infty} \alpha_i^j \sum_{l=1}^{\infty} \beta_{il}^j \left[ f_j(x_j + h_{il}^j) + \frac{1}{i} \|h_{il}^j\| \right] - \sum_{j=1}^N f_j(x_j) \\
& \geq -\langle x_1^*, h_0 \rangle - \sum_{j=2}^N \|x_j^*\| t \zeta \geq -\langle x_1^*, ty \rangle - \|x_1^*\| \cdot \|ty - h_0\| - \sum_{j=2}^N \|x_j^*\| t \zeta \\
& > (M + 3\gamma) \|ty\| - \sum_{j=1}^N \|x_j^*\| t \zeta > (M + 2\gamma) \|ty\| > (M + \gamma) (\|h_0\| - t \zeta) \\
& + \gamma t \|y\| > (M + \gamma) (t \|h_j - h_1\| - 2t \zeta) + \gamma t \|y\| > (M + \gamma) t \|h_j - h_1\|
\end{aligned}$$

for all  $j = 2, \dots, N$  and  $t \in [0, 1]$ . Due to the definition of  $\varphi_{f_j, x_j, \Delta}$  in (2.23) we get condition (b) in Corollary 2.14 and end the proof of the theorem.  $\triangle$

Note that the boundedness from below assumption on the functions  $f_1, \dots, f_N$  in Theorem 2.15 can be dropped by an additional separable reduction. As a consequence of Theorem 2.15, we arrive at the following result needed for the separable reduction of the extremal principle.

**Corollary 2.16 (separable reduction for the extremal principle).** *Let  $Y_0$  be a separable subspace of a (nonseparable) Banach space  $X$ , and let  $\varepsilon > 0$ . Given nonempty subsets  $\Omega_1, \dots, \Omega_n$  of  $X$ ,  $n \geq 2$ , there is a closed separable subspace  $Y \subset X$  such that  $Y_0 \subset Y$  and, for any fixed  $M > 0$ , one has*

$$0 \in (\widehat{N}(x_1; \Omega_1) \setminus MIB_{X^*}) + \widehat{N}(x_2; \Omega_2) + \dots + \widehat{N}(x_n; \Omega_n) + \varepsilon IB_{X^*} \quad (2.31)$$

whenever  $x_1, x_2, \dots, x_n \in Y$  and

$$0 \in (\widehat{N}(x_1; \Omega_1 \cap Y) \setminus MIB_{X^*}) + \widehat{N}(x_2; \Omega_2 \cap Y) + \dots + \widehat{N}(x_n; \Omega_n \cap Y) + \varepsilon IB_{Y^*}.$$

**Proof.** This follows from Theorem 2.15 applied to  $n + 1$  functions

$$f_i(x) := \delta(x; \Omega_i), \quad i = 1, \dots, n, \quad \text{and} \quad f_{n+1}(x) := \varepsilon \|x\|$$

with  $x_1, \dots, x_n \in Y$  and  $x_{n+1} = 0$ .  $\triangle$

### 2.2.3 Extremal Characterizations of Asplund Spaces

In this subsection we consider a general class of Banach spaces, called Asplund spaces, which plays a prominent role in the subsequent variational analysis. We show, based on separable reduction, that the approximate extremal principle unconditionally holds in Asplund spaces, is equivalent to the version of the extremal principle in terms of  $\varepsilon$ -normals, and provides a *characterization* of this class of Banach spaces. Furthermore, we justify the validity of the exact

extremal principle in Asplund spaces under the sequential normal compactness condition imposed on all but one of the sets involved in the extremal system. We also obtain related characterizations of Asplund spaces in terms of supporting properties of Fréchet normals and  $\varepsilon$ -normals at boundary points of closed sets.

**Definition 2.17 (Asplund spaces).** *A Banach space  $X$  is ASPLUND, or it has the ASPLUND PROPERTY, if every convex continuous function  $\varphi: U \rightarrow \mathbb{R}$  defined on an open convex subset  $U$  of  $X$  is Fréchet differentiable on a dense subset of  $U$ .*

Note that Definition 2.17 is equivalent to the standard definition of Asplund spaces, which requires the *generic* Fréchet differentiability of  $\varphi$  on  $U$ , i.e., its Fréchet differentiability on a dense  $G_\delta$  subset of  $U$ . This follows from the well-known fact that the collection of points where a convex continuous function is Fréchet differentiable is automatically a  $G_\delta$  set. For simplicity we always put  $U = X$  in Definition 2.17 that doesn't restrict the generality.

The class of Asplund spaces is well investigated in the geometric theory of Banach spaces. We refer the reader to the books of Deville, Godefroy and Zizler [331], Fabian [416], Phelps [1073], and to the survey paper of Yost [1348] for various characterizations, classifications, properties, and examples of Asplund spaces. Note that this class includes all Banach spaces having Fréchet smooth bump functions (in particular, spaces with Fréchet smooth renorms, hence every reflexive space); spaces with separable duals; spaces of continuous functions  $\mathcal{C}(K)$  on a scattered compact Hausdorff space  $K$  (i.e., such that every subset of  $K$  has an isolated point); the classical space of sequences  $c_0$  with the supremum norm and its generalization  $c_0(\Gamma)$  to an arbitrary set  $\Gamma$ , etc. Although Asplund spaces are generally related to the Fréchet type of differentiability and subdifferentiability, they may fail to have even an equivalent norm Gâteaux differentiable off the origin.

Asplund spaces possess many useful properties some of them are employed in what follows. Let us mention that every closed subspace of an Asplund space is Asplund itself; moreover, *every separable Asplund space admits a Fréchet differentiable renorm*, which is especially important for the method of separable reduction. It is also important that the class of Asplund spaces is stable under Cartesian products and linear isomorphisms. A crucial topological property of duals to Asplund spaces is that the dual unit ball  $B^*$  is weak\* sequentially compact.

There is a number of nice geometric characterizations of Asplund spaces. One of the most striking characterizations is that  *$X$  is Asplund if and only if every separable closed subspace of  $X$  has a separable dual*. In the sequel we often use another characterization of Banach spaces not having the Asplund property: they admit a “rough” equivalent norm *nowhere Fréchet differentiable*. The exact formulation is as follows.

**Proposition 2.18 (Banach spaces with no Asplund property).** *Let  $X$  be a Banach space with the norm  $\|\cdot\|$ . Then  $X$  is not Asplund if and only if there exist a number  $\vartheta > 0$  and an equivalent norm  $|\cdot|$  on  $X$  satisfying  $|\cdot| \leq \|\cdot\|$  and*

$$\limsup_{h \rightarrow 0} \left[ \frac{|x+h| + |x-h| - 2|x|}{\|h\|} \right] > \vartheta \text{ for all } x \in X. \quad (2.32)$$

**Proof.** It is not difficult to show (cf. Proposition 1.23 in Phelps [1073]) that condition (2.32) implies that the convex function  $\varphi(x) = |x|$  is nowhere Fréchet differentiable on  $X$ . Thus (2.32) doesn't hold if  $X$  is Asplund.

To prove the converse statement, we recall that a *weak\* slice* of  $\Lambda \subset X^*$  is a set of the form

$$S(x, \Lambda, \alpha) := \{x^* \in \Lambda \mid \langle x^*, x \rangle > \sigma_\Lambda(x) - \alpha\},$$

where  $x \in X$ ,  $\alpha > 0$ , and  $\sigma_\Lambda(x) := \sup \{\langle x^*, x \rangle \mid x^* \in \Lambda\}$ . Assuming that  $X$  is not Asplund and applying Theorem 2.32 from Phelps [1073], we find a convex symmetric subset  $\Lambda \subset \mathcal{B}^*$  with nonempty interior in  $X^*$  and a number  $\vartheta > 0$  such that  $\Lambda$  doesn't admit a weak\* slice of diameter less than  $2\vartheta$ . Observe that  $|x| := \sigma_\Lambda(x)$  defines an equivalent norm on  $X$  with  $|\cdot| \leq \|\cdot\|$ . For any fixed  $0 \neq x \in X$  we take an arbitrary small  $t > 0$  and select  $x_1^*, x_2^* \in S(x, \Lambda, t\vartheta/2)$  such that  $\|x_1^* - x_2^*\| > 2\vartheta$ . Then we find  $h \in X$ ,  $\|h\| = 1$ , with  $\langle x_1^* - x_2^*, h \rangle > 2\vartheta$ . This yields the estimates

$$\begin{aligned} \left[ \frac{|x+th| + |x-th| - 2|x|}{\|th\|} \right] &\geq \left[ \frac{\langle x_1^*, x+th \rangle + \langle x_2^*, x-th \rangle - 2|x|}{t} \right] \\ &> \frac{1}{t} \left[ |x| - \frac{t\vartheta}{2} + |x| - \frac{t\vartheta}{2} - 2|x| \right] + \langle x_1^* - x_2^*, h \rangle > -\vartheta + 2\vartheta = \vartheta \end{aligned}$$

and implies the required inequality (2.32).  $\triangle$

Based on Proposition 2.18, we now construct an important example showing that in any non-Asplund space there are simple sets with pathological behavior of normals to every boundary point.

**Example 2.19 (degeneracy of normals in non-Asplund spaces).** *Let  $X$  be a Banach space with no Asplund property. Then there exists a closed epi-Lipschitzian set  $\Omega \subset X$  for which the following hold:*

(a) *There is  $K > 1$  such that*

$$\|x^*\| \leq K\varepsilon \text{ for all } x^* \in \widehat{N}_\varepsilon(x; \Omega), \text{ all } x \in \text{bd } \Omega, \text{ and all } \varepsilon > 0.$$

(b)  *$\Omega$  is normally regular at every boundary point with*

$$N(\bar{x}; \Omega) = \widehat{N}(\bar{x}; \Omega) = \{0\} \text{ for all } \bar{x} \in \text{bd } \Omega.$$

**Proof.** Take an arbitrary non-Asplund space  $X$  and represent it in the form  $X = Z \times \mathbb{R}$  with the norm  $\|(z, \alpha)\| := \|z\| + |\alpha|$  for  $(z, \alpha) \in X$ . Then  $Z$  is non-Asplund as well, since the opposite implies the Asplund property of  $X$ . By Proposition 2.18 we find a number  $\vartheta > 0$  and a norm  $|\cdot|$  on  $Z$ , which is equivalent to the original norm  $\|\cdot\|$ , so that  $|\cdot| \leq \|\cdot\|$  and one has (2.32) with  $X = Z$  and  $x = z$ . Based on the norm  $|\cdot|$ , we construct a set  $\Omega \subset X$  in the epigraphical form

$$\Omega := \{(z, \alpha) \in X \mid \alpha \geq \varphi(z)\} \quad \text{with } \varphi := -|\cdot| \quad \text{and } \text{bd } \Omega = \text{gph } \varphi. \quad (2.33)$$

Since  $\varphi$  in (2.33) is Lipschitz continuous on  $X$ , the set  $\Omega$  is epi-Lipschitzian at every boundary point. To justify (a), we need to find a constant  $K > 1$  providing the estimate

$$\|(z^*, \lambda)\| \leq K\varepsilon \quad \text{if } (z^*, \lambda) \in \widehat{N}_\varepsilon((z, \varphi(z)); \Omega), \quad z \in Z, \quad \varepsilon > 0, \quad (2.34)$$

where  $\|(z^*, \lambda)\| := \max\{\|z^*\|, |\lambda|\}$  is the dual norm to  $\|(z, \alpha)\| = \|z\| + |\alpha|$ . Fix arbitrary  $\bar{z} \in Z$  and  $(z^*, \lambda) \in \widehat{N}_\varepsilon((\bar{z}, \varphi(\bar{z})); \Omega)$ . It follows directly from the definition of  $\widehat{N}_\varepsilon$  that

$$\langle z^*, z - \bar{z} \rangle + \lambda(\alpha - \varphi(\bar{z})) \leq 2\varepsilon(\|z - \bar{z}\| + |\alpha - \varphi(\bar{z})|)$$

for all  $(z, \alpha) \in \text{epi } \varphi$  around  $(\bar{z}, \varphi(\bar{z}))$ . Putting here  $z = \bar{z}$ , one gets  $\lambda \leq 2\varepsilon$ . Since  $|\cdot| \leq \|\cdot\|$  and  $|\varphi(z) - \varphi(\bar{z})| \leq |z - \bar{z}|$ , we conclude that

$$\langle z^*, z - \bar{z} \rangle + \lambda(\varphi(z) - \varphi(\bar{z})) \leq 4\varepsilon\|z - \bar{z}\|$$

and further that

$$\langle z^*, z - \bar{z} \rangle \leq (4\varepsilon + |\lambda|)\|z - \bar{z}\|$$

for all  $z$  around  $\bar{z}$ . The latter gives

$$\|z^*\| \leq 4\varepsilon + |\lambda| \quad \text{for any } (z^*, \lambda) \in \widehat{N}_\varepsilon((\bar{z}, \varphi(\bar{z})); \Omega). \quad (2.35)$$

Let us show that (2.35) ensures (2.34) with  $K := \max\{6, 4 + 8/\vartheta\}$ . Indeed, for  $\lambda \geq 0$  we get from (2.35) that  $\|(z^*, \lambda)\| \leq 6\varepsilon$  and arrive at (2.34) with  $K = 6$ . For  $\lambda < 0$  we have from the above definition of  $\widehat{N}_\varepsilon((\bar{z}, \varphi(\bar{z})); \Omega)$  with  $\varphi = -|\cdot|$  that

$$|z| - |\bar{z}| - \left\langle \frac{z^*}{\lambda}, z - \bar{z} \right\rangle \leq -\frac{4\varepsilon}{\lambda}\|z - \bar{z}\|$$

for all  $z$  around  $\bar{z}$ . Putting there  $2\bar{z} - z$  instead of  $z$ , we get

$$|2\bar{z} - z| - |\bar{z}| + \left\langle \frac{z^*}{\lambda}, z - \bar{z} \right\rangle \leq -\frac{4\varepsilon}{\lambda}\|z - \bar{z}\|.$$

Adding the two previous inequalities together, we arrive at

$$|\bar{z} + (z - \bar{z})| + |\bar{z} - (z - \bar{z})| - 2|\bar{z}| \leq -\frac{8\varepsilon}{\lambda} \|z - \bar{z}\|.$$

The latter implies, according to Proposition 2.18 with  $x = \bar{z}$  and  $h = z - \bar{z}$ , that  $|\lambda| < 8\varepsilon/\vartheta$ , where  $\vartheta$  is the fixed positive number from (2.32). Thus (2.35) gives  $\|z^*\| \leq 4\varepsilon + (8\varepsilon/\vartheta)$  for  $\lambda < 0$ , and we arrive at (2.34) with  $K = 4 + 8/\vartheta$ , which justifies (a).

Property (b) follows from (a) due to Definitions 1.1 and 1.4 by passing to the limit as  $\varepsilon \downarrow 0$  and  $x \rightarrow \bar{x}$ .  $\triangle$

Now we are ready to establish the main result of this section ensuring that the first two versions of the extremal principle in Definition 2.5, being applied to *every* extremal system in a Banach space  $X$ , are equivalent to the Asplund property of  $X$ .

**Theorem 2.20 (extremal characterizations of Asplund spaces).** *Let  $X$  be a Banach space. The following are equivalent:*

- (a)  *$X$  is Asplund.*
- (b) *The approximate extremal principle holds in  $X$ .*
- (c) *The  $\varepsilon$ -extremal principle holds in  $X$ .*

**Proof.** First we prove (a) $\Rightarrow$ (b). Let  $X$  be an Asplund space, and let  $\bar{x}$  be a local extremal point of some sets  $\Omega_1, \dots, \Omega_n$  closed around  $\bar{x}$ . By Definition 2.1 we take sequences  $\{a_{ik}\} \subset X, i = 1, \dots, n$ , and then consider a separable subspace  $Y_0$  of  $X$  defined as

$$Y_0 := \text{span} \{ \bar{x}, a_{ik} \mid i = 1, \dots, n, k \in \mathbb{N} \}.$$

Applying the separable reduction result of Corollary 2.16, for every fixed  $\varepsilon > 0$  we find a closed separable subspace  $Y_0 \subset Y \subset X$  that ensures the fulfillment of (2.31) under the conditions imposed in the corollary. Observe that

$$\{ \Omega_1 \cap Y, \dots, \Omega_n \cap Y, \bar{x} \} \quad (2.36)$$

is an extremal system in the space  $Y$ . Indeed,  $\bar{x}$  is obviously a common point of the sets  $\Omega_i \cap Y, i = 1, \dots, n$ , since  $\bar{x} \in Y_0 \subset Y$ . On the other hand, these sets shifted by the corresponding sequences  $a_{ik}, i = 1, \dots, n$ , don't have any common points in the neighborhood  $U \cap Y$  of  $\bar{x}$  in  $Y$  for all large  $k \in \mathbb{N}$ . Since  $a_{ik} \in Y_0 \subset Y$ , this means that  $\bar{x}$  is a local extremal point of the set system  $\{ \Omega_1 \cap Y, \dots, \Omega_n \cap Y \}$  in the space  $Y$ .

Since  $Y$  is a separable Asplund space, it admits an equivalent Fréchet smooth (re)norm denoted again by  $\|\cdot\|$ . Thus one can apply Theorem 2.10 ensuring the fulfillment of the approximate extremal principle for the extremal system (2.36) in  $Y$ . Without loss of generality we assume that  $\varepsilon < 1/4$  and use relations (2.3) and (2.4) of the extremal principle with  $\varepsilon/n$ . In this way we find  $x_i \in \Omega_i \cap (\bar{x} + (\varepsilon/n)\mathcal{B}_Y)$  and

$$y_i^* \in \widehat{N}(x_i; \Omega_i \cap Y) + (\varepsilon/n)\mathcal{B}_{Y^*}$$

satisfying (2.3) for  $y_i^*$ . Hence  $\|y_i^*\| > 1/2n$  for at least one  $i \in \{1, \dots, n\}$ ; let it hold for  $i = 1$ . Thus we have  $y_i^* = \tilde{y}_i^* + u_i^*$  with  $\tilde{y}_i^* \in \hat{N}(x_i; \Omega_i \cap Y)$  and  $\|u_i^*\| \leq \varepsilon/n$  for  $i = 1, \dots, n$  and with

$$\|\tilde{y}_1^*\| \geq \|y_1^*\| - \frac{\varepsilon}{n} > \frac{1-2\varepsilon}{2n} > \frac{1}{4n} := M > 0.$$

This implies the relation

$$0 \in (\hat{N}(x_1; \Omega_1 \cap Y) \setminus \frac{1}{4n} \mathcal{B}_{X^*}) + \hat{N}(x_2; \Omega_2 \cap Y) + \dots + \hat{N}(x_n; \Omega_n \cap Y) + \varepsilon \mathcal{B}_{Y^*}.$$

Due to Corollary 2.16 we get (2.31) with  $M = 1/4n$ . The latter means that there are  $\tilde{x}_i^* \in \hat{N}(x_i; \Omega_i)$ ,  $i = 1, \dots, n$ , and  $v^* \in X^*$  with  $\|v^*\| \leq \varepsilon$  satisfying  $\|\tilde{x}_1^*\| > 1/4n$  and  $\tilde{x}_1^* + \dots + \tilde{x}_n^* + v^* = 0$ . Now denoting  $x_i^* := \tilde{x}_i^*$  for  $i = 1, \dots, n-1$  and  $x_n^* := \tilde{x}_n^* + v^*$ , we have all the relations in (2.3) and (2.4) except the normalization condition  $\|x_1^*\| + \dots + \|x_n^*\| = 1$ . Since  $\gamma := \|\tilde{x}_1^*\| + \dots + \|\tilde{x}_n^*\| > 1/4n$  independently of  $\varepsilon$ , we can easily obtain the normalization condition for  $x_i^*/\gamma$  by adjusting  $\varepsilon$  in (2.4). This gives (a) $\Rightarrow$ (b).

As mentioned above, (b) $\Rightarrow$ (c) always holds. It remains to justify (c) $\Rightarrow$ (a). Assuming that  $X$  is not Asplund, we have the closed set  $\Omega$  from Example 2.19. Then the  $\varepsilon$ -extremal principle is not valid for  $\{\Omega, \{\bar{x}\}, \bar{x}\}$  with any  $\bar{x} \in \text{bd } \Omega$ , since the opposite contradicts Proposition 2.6(i) with  $M = K\varepsilon > \varepsilon$ .  $\triangle$

As a consequence of the results obtained, we arrive at the following characterizations of Asplund spaces via supporting properties of closed sets expressed in terms of Fréchet normals and  $\varepsilon$ -normals at boundary points.

**Corollary 2.21 (boundary characterizations of Asplund spaces).** *Let  $X$  be a Banach space. The following are equivalent:*

- (a)  *$X$  is Asplund.*
- (b) *For every proper closed subset  $\Omega$  of  $X$  the set of points  $x \in \text{bd } \Omega$  with  $\hat{N}(x; \Omega) \neq \{0\}$  is dense in the boundary of  $\Omega$ .*
- (c) *For every proper closed subset  $\Omega$  of  $X$  there is  $x \in \text{bd } \Omega$  such that  $\hat{N}(x; \Omega) \neq \{0\}$ .*
- (d) *For every proper closed subset  $\Omega$  of  $X$ , every  $\varepsilon > 0$ , and every  $M > \varepsilon$  the set of points  $x \in \text{bd } \Omega$  with  $\hat{N}_\varepsilon(x; \Omega) \setminus M \mathcal{B}^* \neq \emptyset$  is dense in the boundary of  $\Omega$ .*
- (e) *For every proper closed subset  $\Omega$  of  $X$ , every  $\varepsilon > 0$ , and every  $M > \varepsilon$  there is  $x \in \text{bd } \Omega$  such that  $\hat{N}_\varepsilon(x; \Omega) \setminus M \mathcal{B}^* \neq \emptyset$ .*

**Proof.** Implication (a) $\Rightarrow$ (b) follows from Theorem 2.20 and Proposition 2.6(ii). Implications (b) $\Rightarrow$ (c) $\Rightarrow$ (e) and (b) $\Rightarrow$ (d) $\Rightarrow$ (e) are trivial. Implication (e) $\Rightarrow$ (a) follows from Example 2.19; see the end of the proof of Theorem 2.20.  $\triangle$

As follows from the above proof, an arbitrary number  $M > \varepsilon$  in (d) and (e) can be equivalently replaced with  $K\varepsilon$ ,  $K > 1$ . Related characterizations

of Asplund spaces in terms of  $\varepsilon$ -normals can be written in the form: *for every proper closed subset  $\Omega \subset X$  there is  $\lambda > 0$  such that for each  $\varepsilon > 0$  the set*

$$\left\{ x \in \text{bd } \Omega \mid \exists x^* \in \widehat{N}_\varepsilon(x; \Omega) \text{ with } \|x^*\| = \lambda \right\}$$

*is dense in the boundary of  $\Omega$ , or is just nonempty*; see Mordukhovich and B. Wang [960] for the proof and discussions.

We can see from the above results that the supporting properties (b)–(e) in Corollary 2.21 applied to *every* closed subset of  $X$  are equivalent to the “fuzzy” versions of the extremal principle in Theorem 2.20, since each of them characterizes Asplund spaces. This is essentially based on properties of Fréchet normals and  $\varepsilon$ -normals in Asplund spaces: cf. the related discussions in Subsect. 2.1.2. It follows from the proofs that for the equivalencies in Corollary 2.21 one can consider only *epigraphical sets* of type (2.33).

Next let us obtain conditions ensuring the fulfillment of the *exact* extremal principle in Definition 2.5(iii). For this purpose we employ the sequential normal compactness (SNC) property of sets introduced in Subsect. 1.1.3.

**Theorem 2.22 (exact extremal principle in Asplund spaces).**

(i) *Let  $X$  be an Asplund space, and let  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$  be an extremal system in  $X$  such that all  $\Omega_i$  are locally closed around  $\bar{x}$  and all but one of  $\Omega_i$  are sequentially normally compact at  $\bar{x}$ . Then the exact extremal principle holds for  $\{\Omega_1, \dots, \Omega_n, \bar{x}\}$ .*

(ii) *Conversely, let the exact extremal principle hold for every extremal system  $\{\Omega_1, \Omega_2, \bar{x}\}$  in  $X$ , where both sets  $\Omega_i$  are closed and one of them is sequentially normally compact at  $\bar{x}$ . Then  $X$  is Asplund.*

**Proof.** To justify (i), we use the  $\varepsilon$ -extremal principle that holds in any Asplund space by Theorem 2.20. Take a sequence of  $\varepsilon_k \downarrow 0$  as  $k \rightarrow \infty$  and consider the corresponding sequence of  $x_{ik}$  and  $x_{ik}^*$ ,  $i = 1, \dots, n$ , satisfying (2.2) and (2.3) with  $\varepsilon = \varepsilon_k$ . Then  $x_{ik} \rightarrow \bar{x}$  for all  $i = 1, \dots, n$ . Since the sequences  $\{x_{ik}^*\}$  are bounded in  $X^*$  and since bounded sets in duals to Asplund spaces are weak\* sequentially compact, we find  $x_i^* \in X^*$  such that  $x_{ik}^* \xrightarrow{w^*} x_i^*$  for  $i = 1, \dots, n$ . Passing to the limit in (2.2) as  $k \rightarrow \infty$  and using the definition of basic normals, we get (2.5). Also one obviously has  $x_1^* + \dots + x_n^* = 0$ . It remains to show that  $(x_1^*, \dots, x_n^*) \neq 0$  under the SNC assumptions of the theorem. On the contrary, assume that  $x_i^* = 0$  while  $\Omega_i$  are SNC at  $\bar{x}$  for  $i = 1, \dots, n - 1$ . By Definition 1.20 the latter implies that  $\|x_{ik}^*\| \rightarrow 0$  as  $k \rightarrow \infty$  for  $i = 1, \dots, n - 1$ . Hence

$$\|x_{nk}^*\| \leq \|x_{1k}^*\| + \dots + \|x_{n-1k}^*\| \rightarrow 0 \text{ as } k \rightarrow \infty,$$

which contradicts the nontriviality condition  $\|x_{1k}^*\| + \dots + \|x_{nk}^*\| = 1$  for all  $k \in \mathbb{N}$  and ends the proof of (i).

To prove (ii), we assume that  $X$  is not an Asplund space and represent it as  $X = Z \times \mathbb{R}$ , where  $Z$  must be non-Asplund as well. Then consider

$\Omega_1 := \{0\} \times (-\infty, 0] \in Z \times \mathbb{R}$  and  $\Omega_2 := \Omega$  defined in (2.33). One can easily check that  $\bar{x} = (0, 0)$  is a local extremal point of these closed sets in  $X$ . Since  $\Omega_2$  is epi-Lipschitzian at  $\bar{x}$ , it is SNC at this point due to Theorem 1.26. However, the exact extremal principle doesn't hold for  $\{\Omega_1, \Omega_2, \bar{x}\}$ . Indeed, we have  $N((0, 0); \Omega_2) = \{(0, 0)\}$  from property (b) in Example 2.19, while  $N((0, 0); \Omega_1) = Z^* \times [0, \infty)$ . That is,  $N(\bar{x}; \Omega_1) \cap (-N(\bar{x}; \Omega_2)) = \{(0, 0)\}$ , which justifies (ii) and ends the proof of the theorem.  $\triangle$

Let us show that the SNC assumption in Theorem 2.22 is *essential* for the fulfillment of the exact extremal principle in infinite-dimensional spaces.

**Example 2.23 (violation of the exact extremal principle in the absence of SNC).** *Every infinite-dimensional separable Banach space contains an extremal system  $\{\Omega_1, \Omega_2, \bar{x}\}$  that doesn't satisfy the relations of the exact extremal principle.*

**Proof.** Let  $X$  be a separable Banach space, and let  $\{e_k\}_1^\infty$  be unit independent vectors that densely span  $X$ . Consider the sets

$$\Omega_1 := \text{clco} \left\{ \frac{e_n}{2^n}, -\frac{e_n}{2^n} \mid n \in \mathbb{N} \right\},$$

and  $\Omega_2 = \{0\}$ , which are convex and compact in the norm topology of  $X$ . Note that  $\Omega_1$  and  $\Omega_2$  are *not* SNC unless  $X$  is finite-dimensional; see Theorem 1.21. Let us check that  $0 \in \Omega_1 \cap \Omega_2$  is a local extremal point of the set system  $\{\Omega_1, \Omega_2\}$ . Indeed, taking

$$a := \sum_{n=1}^{\infty} \frac{e_n}{n^2} \in X,$$

we observe that for any sequence of  $v_k \downarrow 0$  one has

$$\Omega_1 \cap (v_k a + \Omega_2) = \Omega_1 \cap \{v_k a\} = \emptyset.$$

It follows from the structure of  $\Omega_1$  that  $N(0; \Omega_1) = \{0\}$ , and thus  $\{\Omega_1, \Omega_2, 0\}$  doesn't satisfy the exact extremal principle.  $\triangle$

Next we consider some properties of the basic normal cone  $N(\cdot; \Omega)$  on *boundaries* of closed sets. It immediately follows from Corollary 2.21 that in Asplund spaces the sets of point  $x \in \text{bd } \Omega$  with  $N(x; \Omega) \neq \{0\}$  is *dense* in the boundary of any proper closed subset  $\Omega \subset X$ . Moreover, Example 2.19 shows that even nonemptiness of this set for any  $\Omega$  of type (2.33) implies that  $X$  is Asplund. Theorem 2.22 gives conditions under which this nontriviality property of basic normals holds at *every* boundary point of closed sets.

**Corollary 2.24 (nontriviality of basic normals in Asplund spaces).** *Let  $X$  be an Asplund space, and let  $\Omega$  be a proper closed subset of  $X$ . Then  $N(\bar{x}; \Omega) \neq \{0\}$  at every point  $\bar{x} \in \text{bd } \Omega$  where the set  $\Omega$  is sequentially normally compact.*



**Proof.** Follows from Theorem 2.22 applied to the system  $\{\Omega, \{\bar{x}\}, \bar{x}\}$ .  $\triangle$

Note that the result of Corollary 2.24 gives a new condition for the *supporting hyperplane* property even for *closed convex cones* in Asplund spaces, where the SNC assumption may be strictly weaker than the CEL one; see Remark 1.27 with its references and Example 3.6 in Subsect. 3.1.1.

In conclusion of this section we present a consequence of the results above that characterizes Asplund spaces via the existence of basic subgradients for every locally Lipschitzian function.

**Corollary 2.25 (subdifferentiability of Lipschitzian functions on Asplund spaces).** *Let  $X$  be a Banach space. Then  $\partial\varphi(\bar{x}) \neq \emptyset$  for every function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  locally Lipschitzian around  $\bar{x}$  if and only if  $X$  is Asplund.*

**Proof.** Consider any function  $\varphi$  on an Asplund space  $X$  that is Lipschitz continuous around  $\bar{x}$ . Then  $N((\bar{x}, \varphi(\bar{x})); \text{epi } \varphi) \neq \{(0, 0)\}$  due to Corollary 2.24. By Corollary 1.81 we have  $\partial\varphi(\bar{x}) \neq \emptyset$ . Conversely, if  $X$  is not Asplund, then  $\partial\varphi(x) \equiv \emptyset$  on  $X$  for the Lipschitz continuous function  $\varphi$  in (2.33).  $\triangle$

## 2.3 Relations with Variational Principles

By *variational principles*, in the conventional terminology of variational analysis, one means a group of results stating that for any lower semicontinuous (l.s.c.) and bounded from below function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  and a point  $x_0$  close to its minimum there is an arbitrary small perturbation  $\theta(\cdot)$  such that the resulting function  $\varphi + \theta$  achieves its minimum at some point  $\bar{x}$  near  $x_0$ . A variational principle is said to be *smooth* when the perturbation function may be chosen as smooth in some sense. The first general variational principle was established by Ekeland [396, 397] in complete metric spaces. Among smooth variational principles the most powerful are those by Borwein and Preiss [154] in Banach spaces with smooth renorms and by Deville, Godefroy and Zizler [331] in Banach spaces with smooth bump functions. Variational principles play a prominent role in many aspects of nonlinear analysis, optimization, and numerous applications.

For  $\dim X < \infty$  such principles easily follow from the classical Weierstrass existence theorem and the compactness of the unit ball  $\mathcal{B} \subset X$ . In the case of infinite-dimensional spaces they ensure the existence of optimal solutions to perturbed problems and hence lead, by employing some calculus, to “almost” minimal points of the original function  $\varphi$  that “almost” satisfy necessary optimality conditions in terms of corresponding subgradients of  $\varphi$ . If  $X$  admits a smooth variational principle, such conditions can be obtained in terms of Fréchet subgradients by using the simple rule of Proposition 1.107(i). However, as we’ll see below, smooth variational principles may be applied *only if*  $X$  has some *smoothness* properties, while the required subgradient conditions can be derived from the approximate extremal principle in *any Asplund*

space. In this way we establish relationships between the extremal principle and appropriate versions of variational principles in  $X$  and obtain variational characterizations of Asplund spaces in terms of Fréchet subgradients and  $\varepsilon$ -subgradients of lower semicontinuous functions.

### 2.3.1 Ekeland Variational Principle

Let us start with the fundamental variational principle of Ekeland that turns out to be a *characterization* of complete metric spaces  $(X, d)$ .

**Theorem 2.26 (Ekeland's variational principle).** *Let  $(X, d)$  be a metric space. The following hold:*

(i) *Assume that  $X$  is complete and that  $\varphi: X \rightarrow \overline{\mathbb{R}}$  is a proper l.s.c. function bounded from below. Let  $\varepsilon > 0$  and  $x_0 \in X$  be given such that  $\varphi(x_0) \leq \inf_X \varphi + \varepsilon$ . Then for any  $\lambda > 0$  there is  $\bar{x} \in X$  satisfying*

- (a)  $\varphi(\bar{x}) \leq \varphi(x_0)$ ,
- (b)  $d(\bar{x}, x_0) \leq \lambda$ ,
- (c)  $\varphi(x) + (\varepsilon/\lambda)d(x, \bar{x}) > \varphi(\bar{x})$  for all  $x \neq \bar{x}$ .

(ii) *Conversely,  $X$  is complete if for every Lipschitz continuous function  $\varphi: X \rightarrow \mathbb{R}$  bounded from below and every  $\varepsilon > 0$  there is  $\bar{x} \in X$  satisfying*

(a')  $\varphi(\bar{x}) \leq \inf_X \varphi + \varepsilon$  and property (c) above with  $\lambda = 1$ .

**Proof.** Let us justify (i) observing that it is sufficient to consider the case of  $\varepsilon = \lambda = 1$ . Indeed, the general case in (i) can be easily reduced to this special case applied to the function  $\tilde{\varphi}(x) := \varepsilon^{-1}\varphi(x)$  on the metric space  $(X, \tilde{d})$  with  $\tilde{d}(x, y) := \lambda^{-1}d(x, y)$ . Putting  $\varepsilon = \lambda = 1$  in what follows, we first prove that there always exists  $\bar{x} \in X$  satisfying (c) under the assumptions in (i). Define a mapping  $T: X \rightrightarrows X$  by

$$T(x) := \{u \in X \mid \varphi(u) + d(x, u) \leq \varphi(x)\}.$$

Starting with an arbitrary point  $x_1 \in \text{dom } \varphi$ , we inductively construct a sequence  $\{x_k\}$ ,  $k \in \mathbb{N}$ , as follows. Assume that  $x_k$  is known and select the next iteration  $x_{k+1}$  so that

$$x_{k+1} \in T(x_k) \quad \text{and} \quad \varphi(x_{k+1}) < \inf_{x \in T(x_k)} \varphi(x) + \frac{1}{k}, \quad k \in \mathbb{N}.$$

Observe that all  $T(x_k)$  are nonempty and closed. Moreover,  $T(x_{k+1}) \subset T(x_k)$  due to the triangle inequality. This gives

$$\begin{aligned} d(u, x_{k+1}) &\leq \varphi(x_{k+1}) - \varphi(u) \leq \inf_{x \in T(x_k)} \varphi(x) + \frac{1}{k} - \varphi(u) \\ &\leq \inf_{x \in T(x_{k+1})} \varphi(x) + \frac{1}{k} - \varphi(u) \leq \frac{1}{k} \end{aligned}$$

for all  $u \in T(x_{k+1})$ ,  $k \in \mathbb{N}$ . Therefore

$$\text{diam } T(x_k) := \sup_{x, u \in T(x_k)} d(x, u) \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Due to the completeness of  $X$  we conclude that the sets  $T(x_k)$  shrink to a single point:

$$\bigcap_{k=1}^{\infty} T(x_k) = \{\bar{x}\} \text{ for some } \bar{x} \in \text{dom } \varphi.$$

The latter implies (c) by the construction of  $T(x_k)$ .

Now given  $x_0 \in X$  with  $\varphi(x_0) \leq \inf_X \varphi + 1$ , we consider the space

$$X_0 := \{x \in X \mid \varphi(x) \leq \varphi(x_0) - d(x, x_0)\}$$

with the metric induced by  $d$ . Obviously  $(X_0, d)$  is complete. Applying (c) on this space, we find  $\bar{x} \in X_0$  such that

$$\varphi(x) > \varphi(\bar{x}) - d(x, \bar{x}) \text{ for all } x \in X_0 \setminus \{\bar{x}\}.$$

Let us show that the point  $\bar{x}$  satisfies all the conditions (a)–(c) in (i) with  $\varepsilon = \lambda = 1$ . Indeed, (a) and (b) follow directly from  $\bar{x} \in X_0$ , i.e., from  $\varphi(\bar{x}) + d(\bar{x}, x_0) \leq \varphi(x_0)$  and  $\varphi(x_0) \leq \inf_X \varphi + 1$ . It remains to prove (c) for  $x \in X \setminus X_0$ . Taking  $x \notin X_0$ , one has by the above construction that

$$\begin{aligned} \varphi(x) &> \varphi(x_0) - d(x, x_0) \geq \varphi(\bar{x}) + d(\bar{x}, x_0) - d(x, x_0) \\ &\geq \varphi(\bar{x}) - d(\bar{x}, x), \end{aligned}$$

which ends the proof of (i).

To prove the converse statement (ii), let us consider an arbitrary Cauchy sequence  $\{x_k\}$  in  $X$  and define the function

$$\varphi(x) := \lim_{k \rightarrow \infty} d(x_k, x) \text{ for all } x \in X,$$

where the limit exists due to

$$|d(x_k, x) - d(x_n, x)| \leq d(x_k, x_n) \rightarrow 0 \text{ as } k, n \rightarrow \infty$$

by the triangle inequality. This also gives

$$|d(x_k, x) - d(x_k, u)| \leq d(x, u) \text{ for all } x, u \in X, \quad k \in \mathbb{N},$$

which implies the Lipschitz continuity of  $\varphi$  on  $X$ . Since  $\{x_k\}$  is a Cauchy sequence, for every  $\varepsilon > 0$  we find  $k(\varepsilon) \in \mathbb{N}$  such that  $d(x_k, x_n) \leq \varepsilon$  whenever  $k, n \geq k(\varepsilon)$ . Thus

$$\varphi(x_n) = \lim_{k \rightarrow \infty} d(x_k, x_n) \leq \varepsilon \text{ if } n \geq k(\varepsilon),$$

and hence  $\varphi$  is bounded from below with  $\inf_X \varphi = 0$ . To prove the completeness of  $X$ , we need to find  $\bar{x} \in X$  such that  $\varphi(\bar{x}) = 0$ .

Choose  $\varepsilon \in (0, 1)$  and take  $\bar{x} \in X$  satisfying (a') and (c) with  $\lambda = 1$ . Then  $\varphi(\bar{x}) \leq \varepsilon$  due to (a') and  $\inf_X \varphi = 0$ . Now pick an arbitrary small  $\gamma > 0$  and put  $x = x_n$  in (c) with  $n \in \mathbb{N}$ . From the definition of  $\varphi$  and the fact that  $\{x_k\}$  is a Cauchy sequence, we get  $d(x_n, \bar{x}) \leq \varepsilon + \gamma$  when  $n$  is sufficiently large. This gives  $\varphi(\bar{x}) \leq \varepsilon^2$  by passing to the limit in (c) with  $x = x_n$  as  $n \rightarrow \infty$  and  $\gamma \downarrow 0$ . Repeating this procedure  $m$  times, one has  $\varphi(\bar{x}) \leq \varepsilon^m$  for any  $m \in \mathbb{N}$ . Thus  $\varphi(\bar{x}) = 0$ , which justifies the completeness of  $X$ .  $\triangle$

Condition (c) in Theorem 2.26 means that the perturbed function  $\varphi(x) + (\varepsilon/\lambda)d(x, \bar{x})$  achieves at  $\bar{x}$  its *strict global minimum* over  $X$ . It has many important consequences. Let us present one, which is of special interest for subsequent discussions.

**Corollary 2.27 ( $\varepsilon$ -stationary condition).** *Let  $\varphi: X \rightarrow \mathbb{R}$  be a proper l.s.c. function bounded from below on a Banach space  $X$ . Given  $\varepsilon, \lambda > 0$  and  $x_0 \in X$  with  $\varphi(x_0) \leq \inf_X \varphi + \varepsilon$ , we assume that  $\varphi$  is Fréchet differentiable on a neighborhood  $U$  of  $x_0$  containing  $B_\lambda(x_0)$ . Then there is  $\bar{x} \in X$  with  $\|\bar{x} - x_0\| \leq \lambda$  such that  $\varphi(\bar{x}) \leq \varphi(x_0)$  and  $\|\nabla \varphi(\bar{x})\| \leq \varepsilon/\lambda$ .*

**Proof.** Since  $\bar{x}$  is a minimum point of the sum  $\varphi(x) + \psi(x)$  with  $\psi(x) := (\varepsilon/\lambda)\|x - \bar{x}\|$ , we have  $0 \in \widehat{\partial}(\varphi + \psi)(\bar{x})$  by Proposition 1.114. Now applying Proposition 1.107(i) and taking into account that  $\widehat{\partial}(\|\cdot - \bar{x}\|)(\bar{x}) = \mathbb{B}^*$  for the norm function in Banach spaces, we get all the conclusions of the corollary from Theorem 2.26(i).  $\triangle$

Note that, since the initial  $\varepsilon$ -optimal point  $x_0$  always exists, Corollary 2.27 ensures that every Fréchet differentiable and bounded from below function  $\varphi$  on a Banach space  $X$  admits an  $\varepsilon$ -optimal point  $\bar{x}$  satisfying the  $\varepsilon$ -stationary condition  $\|\nabla \varphi(\bar{x})\| \leq \varepsilon$  for an arbitrary small  $\varepsilon > 0$ . As shown in the original paper of Ekeland [397], this result holds also for Gâteaux differentiable functions, which is a direct consequence of his variational principle.

What happens when  $\varphi$  is nonsmooth? This is considered next.

### 2.3.2 Subdifferential Variational Principles

In this subsection we first obtain a *lower* subdifferential counterpart of the  $\varepsilon$ -stationary result of Corollary 2.27 to the case of arbitrary l.s.c. functions bounded from below. We'll see that such an extension derived by using the extremal principle turns out to be a characterization of Asplund spaces. It actually plays a role of a (local) variational principle in Asplund spaces and has many important consequences, including density results for Fréchet subgradients as well as conventional forms of smooth variational principles under appropriate smoothness assumptions on Banach spaces. Finally,

we derive an *upper* version of the subdifferential variational principle that holds in general Banach spaces and involves *every* upper Fréchet subgradient (provided that they exist) instead of *some* lower subgradient as in the previous lower subdifferential counterpart.

**Theorem 2.28 (lower subdifferential variational principle).** *Let  $X$  be a Banach space. The following are equivalent:*

- (a) *The approximate extremal principle holds in  $X$ .*
- (b) *For every proper l.s.c. function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  bounded from below, every  $\varepsilon > 0$ ,  $\lambda > 0$ , and  $x_0 \in X$  with  $\varphi(x_0) < \inf_X \varphi + \varepsilon$  there are  $\bar{x} \in X$  and  $x^* \in \widehat{\partial}\varphi(\bar{x})$  such that  $\|\bar{x} - x_0\| < \lambda$ ,  $\varphi(\bar{x}) < \inf_X \varphi + \varepsilon$ , and  $\|x^*\| < \varepsilon/\lambda$ .*
- (c)  *$X$  is Asplund.*

**Proof.** Implication (c) $\Rightarrow$ (a) is established in Theorem 2.20. Let us justify the other implications. We begin with (b) $\Rightarrow$ (c) and then derive (a) $\Rightarrow$ (b), which is the main part of the theorem.

(b) $\Rightarrow$ (c). Take an arbitrary convex continuous function  $\varphi: X \rightarrow \mathbb{R}$ . Then  $\widehat{\partial}\varphi(x)$  agrees with the subdifferential of convex analysis and is nonempty at every  $x \in X$ . To establish the Asplund property of  $X$ , it is sufficient to show that there is a dense subset  $S \subset X$  such that  $\widehat{\partial}(-\varphi)(x) \neq \emptyset$  for every  $x \in S$ . Indeed, in this case  $\varphi$  is Fréchet differentiable on  $S$  due to Proposition 1.87.

Fix  $x_0 \in X$  and  $\varepsilon > 0$ . Since  $\psi(x) := -\varphi(x)$  is continuous, there is a positive number  $\nu < \varepsilon$  such that  $\psi(x) > \psi(x_0) - \varepsilon$  for all  $x \in x_0 + \nu B$ . Thus we have  $\phi(x_0) < \inf_X \phi + 2\varepsilon$ , where the function

$$\phi(x) := \psi(x) + \delta(x; x_0 + \nu B), \quad x \in X,$$

is obviously lower semicontinuous on  $X$ . Applying (b) to the latter function, we find  $\bar{x} \in X$  with  $\|\bar{x} - x_0\| < \nu$  such that  $\widehat{\partial}\phi(\bar{x}) \neq \emptyset$ . This clearly implies that  $\widehat{\partial}\psi(\bar{x}) \neq \emptyset$ , i.e., the set of points  $x \in X$  with  $\widehat{\partial}(-\varphi)(x) \neq \emptyset$  is dense in  $X$ . Hence  $X$  must be Asplund.

(a) $\Rightarrow$ (b). First let us choose  $0 < \tilde{\varepsilon} < \varepsilon$  with  $\varphi(x_0) < \inf_X \varphi + (\varepsilon - \tilde{\varepsilon})$  and put  $\tilde{\lambda} := (2\varepsilon)^{-1}(2\varepsilon - \tilde{\varepsilon})\lambda < \lambda$ . Applying Theorem 2.26(i), we find  $\tilde{x} \in X$  satisfying  $\|\tilde{x} - x_0\| \leq \tilde{\lambda}$ ,  $\varphi(\tilde{x}) \leq \inf_X \varphi + (\varepsilon - \tilde{\varepsilon})$ , and

$$\varphi(\tilde{x}) < \varphi(x) + \tilde{\lambda}^{-1}(\varepsilon - \tilde{\varepsilon})\|x - \tilde{x}\| \quad \text{for all } x \in X \setminus \{\tilde{x}\}. \quad (2.37)$$

Define two closed subsets of  $X \times \mathbb{R}$  by

$$\Omega_1 := \text{epi } \varphi, \quad \Omega_2 := \{(x, \alpha) \in X \times \mathbb{R} \mid \alpha \leq \varphi(\tilde{x}) - \tilde{\lambda}^{-1}(\varepsilon - \tilde{\varepsilon})\|x - \tilde{x}\|\}.$$

It is easy to conclude from (2.37) that  $(\tilde{x}, \varphi(\tilde{x}))$  is a *local extremal point* of the set system  $\{\Omega_1, \Omega_2\}$ ; so we can use the extremal principle.

Consider the norm  $\|(x, \alpha)\| := \|x\| + |\alpha|$  on  $X \times \mathbb{R}$  with the corresponding dual norm  $\|(x^*, \xi)\| = \max\{\|x^*\|, |\xi|\}$  on  $X^* \times \mathbb{R}$ . Applying the approximate extremal principle to the above system, for any  $\hat{\varepsilon} > 0$  we find  $(x_i, \alpha_i) \in \Omega_i$  and  $(x_i^*, \xi_i) \in \widehat{N}((x_i, \alpha_i); \Omega_i)$ ,  $i = 1, 2$ , satisfying

$$\begin{cases} \|x_i - \tilde{x}\| + |\alpha_i - \varphi(\tilde{x})| < \hat{\varepsilon} , \\ \frac{1}{2} - \hat{\varepsilon} < \max \{ \|x_i^*\|, |\xi_i| \} < \frac{1}{2} + \hat{\varepsilon} , \\ \max \{ \|x_1^* + x_2^*\|, |\xi_1 + \xi_2| \} < \hat{\varepsilon} . \end{cases} \quad (2.38)$$

Observe that  $(x_2^*, \xi_2) \neq 0$  when  $\hat{\varepsilon}$  is sufficiently small. It follows from the structure of  $\Omega_2$  that  $\alpha_2 = \varphi(\tilde{x}) - \tilde{\lambda}^{-1}(\varepsilon - \tilde{\varepsilon})\|x_2 - \tilde{x}\|$ , which yields  $\xi_2 > 0$  and thus implies

$$x_2^*/\xi_2 \in \partial\left(\tilde{\lambda}^{-1}(\varepsilon - \tilde{\varepsilon})\|\cdot - \tilde{x}\|\right)(x_2) \quad \text{and} \quad \|x_2^*\|/\xi_2 \leq \tilde{\lambda}^{-1}(\varepsilon - \tilde{\varepsilon}) .$$

Taking (2.38) into account, the latter gives the estimate

$$\xi_2 \geq \min \left\{ \frac{(1 - 2\hat{\varepsilon})\tilde{\lambda}}{2(\varepsilon - \tilde{\varepsilon})}, \frac{1}{2} - \hat{\varepsilon} \right\} , \quad (2.39)$$

which ensures by (2.38) that  $\xi_1 < 0$  when  $\hat{\varepsilon}$  is sufficiently small. This allows us to show that  $\alpha_1 = \varphi(x_1)$ , since the opposite implies  $\xi_1 = 0$  due to  $(x_1^*, \xi_1) \in \hat{N}((x_1, \alpha_1); \text{epi } \varphi)$  and the definition of  $\hat{N}$ . Consequently  $-x_1^*/\xi_1 \in \partial\varphi(x_1)$ .

It follows from (2.39) that  $\hat{\varepsilon}/\xi_2 \rightarrow 0$  as  $\hat{\varepsilon} \downarrow 0$ . Putting all the above together, we have

$$\frac{\|x_1^*\|}{|\xi_1|} < \frac{\|x_2^*\| + \hat{\varepsilon}}{\xi_2 - \hat{\varepsilon}} = \left( \frac{\|x_2^*\|}{\xi_2} + \frac{\hat{\varepsilon}}{\xi_2} \right) / \left( 1 - \frac{\hat{\varepsilon}}{\xi_2} \right) < \frac{\varepsilon}{\tilde{\lambda}}$$

when  $\hat{\varepsilon}$  is sufficiently small. On the other hand, it follows from (2.38) and the choice of  $\tilde{\lambda}$  that

$$\|x_1 - x_0\| < \tilde{\lambda} + \hat{\varepsilon} \quad \text{and} \quad \varphi(x_1) = \alpha_1 < \inf_X \varphi + \varepsilon - \tilde{\varepsilon} + \hat{\varepsilon} .$$

Finally, letting  $\bar{x} := x_1$  and  $x^* := -x_1^*/\xi_1$ , we arrive at all the conclusions in (b) and finish the proof of the theorem.  $\triangle$

One can see that the major difference between the results of Theorem 2.26(i) and Theorem 2.28(b) is that, instead of the minimization condition (c) in the first theorem, we have the “almost stationary” lower subdifferential condition in the second one with the same type of estimates. The latter subdifferential condition carries essential information for local variational analysis and applications, which allows us to treat assertion (b) of Theorem 2.28 as a proper variational principle in Asplund spaces and call it the (lower) *subdifferential variational principle*. Moreover, we’ll see in the next subsection that this result implies smooth variational principles in the conventional minimization/support form under additional smoothness assumptions on Asplund spaces that are *necessary* for the fulfillment of smooth variational principles but are *not needed* in Theorem 2.28.

The subdifferential variational principle of Theorem 2.28 easily implies the dense Fréchet subdifferentiability and related properties of l.s.c. functions that also turn out to be characterizations of Asplund spaces.

**Corollary 2.29 (Fréchet subdifferentiability of l.s.c. functions).** *Let  $\mathcal{A}$  be a class of all proper l.s.c. functions  $\varphi: X \rightarrow \overline{\mathbb{R}}$  on a Banach space  $X$ . The following properties are equivalent:*

- (a)  *$X$  is Asplund.*
- (b) *For every  $\varphi \in \mathcal{A}$  the set of points  $\{(x, \varphi(x)) \in X \times \mathbb{R} \mid \widehat{\partial}\varphi(x) \neq \emptyset\}$  is dense in the graph of  $\varphi$ .*
- (c) *For every  $\varphi \in \mathcal{A}$  there is  $x \in \text{dom } \varphi$  with  $\widehat{\partial}\varphi(x) \neq \emptyset$ .*
- (d) *For every  $\varphi \in \mathcal{A}$  and every  $\varepsilon > 0$  there is  $x \in \text{dom } \varphi$  with  $\widehat{\partial}_{g\varepsilon}\varphi(x) \neq \emptyset$ .*
- (e) *For every  $\varphi \in \mathcal{A}$  and every  $\varepsilon > 0$  there is  $x \in \text{dom } \varphi$  with  $\widehat{\partial}_{a\varepsilon}\varphi(x) \neq \emptyset$ .*

**Proof.** By Theorem 2.28 the smooth variational principle holds in any Asplund space. Take arbitrary  $\varphi \in \mathcal{A}$ ,  $x_0 \in \text{dom } \varphi$ , and  $\varepsilon > 0$ . Following the proof of (b) $\Rightarrow$ (c) in the above theorem, we find  $\bar{x} \in X$  such that  $\|\bar{x} - x_0\| < \varepsilon$ ,  $|\varphi(\bar{x}) - \varphi(x_0)| < 2\varepsilon$ , and  $\widehat{\partial}\varphi(\bar{x}) \neq \emptyset$ . This justifies (a) $\Rightarrow$ (b) in the corollary. Implications (b) $\Rightarrow$ (c) $\Rightarrow$ (d) are obvious, and (d) $\Rightarrow$ (e) easily follows from Theorem 1.86. To justify the concluding implication (e) $\Rightarrow$ (a), it is sufficient to observe that the concave continuous function  $\varphi := -|\cdot|$  from Proposition 2.18 violates (e) for every  $\varepsilon < \vartheta/2$ .  $\triangle$

It follows from the proof of Corollary 2.29 that all the equivalences therein keep holding if the class  $\mathcal{A}$  is replaced by more narrow classes of l.s.c. functions. In particular, one can consider only *concave continuous* functions  $\varphi: X \rightarrow \mathbb{R}$ , or proper l.s.c. functions  $\varphi: X \rightarrow \overline{\mathbb{R}}$  *bounded from below*. The latter follows from the fact that implication (e) $\Rightarrow$ (a) can be verified for the function  $\varphi = 1/|\cdot|$ , where  $|\cdot|$  is taken from Proposition 2.18. Note also that the list of equivalences in Corollary 2.29 can be supplemented by counterparts of (b) and (c) in terms of *basic subgradients*. It immediately follows from the limiting representations (1.55) in Theorem 1.89.

Finally in this subsection, we establish another version of the subdifferential variational principle whose difference from that in Theorem 2.28 consists of using *upper* Fréchet subgradients instead of lower ones as above. The new version, which holds in arbitrary Banach spaces, involves *every* upper subgradient of the function in question, while it generally doesn't guarantee the existence of such subgradients. However, this result has certain essential advantages in comparison with its lower subdifferential counterpart being useful in some applications (particularly for deriving *suboptimality* conditions in constrained minimization) for important classes of functions that admit nonempty Fréchet upper subdifferential at reference points; see Chap. 5 for various results, discussions, and references.

**Theorem 2.30 (upper subdifferential variational principle).** *Let  $X$  be a Banach space, and let  $\varphi: X \rightarrow \mathbb{R}$  be a l.s.c. function bounded from below. Then for every  $\varepsilon > 0$ ,  $\lambda > 0$ , and  $x_0 \in X$  with  $\varphi(x_0) < \inf_X \varphi + \varepsilon$  there is  $\bar{x} \in X$  with  $\|\bar{x} - x_0\| < \lambda$  and  $\varphi(\bar{x}) < \inf_X \varphi + \varepsilon$  such that*

$$\|x^*\| < \varepsilon/\lambda \quad \text{whenever } x^* \in \widehat{\partial}^+\varphi(\bar{x}).$$

**Proof.** Given arbitrary numbers  $\varepsilon > 0$  and  $\lambda > 0$  and applying Ekeland's variational principle to the function  $\varphi$  and the point  $x_0$  under consideration, we find  $\bar{x} \in X$  satisfying  $\|x_0 - \bar{x}\| < \lambda$ ,  $\varphi(\bar{x}) < \inf_X \varphi(x) + \varepsilon$ , and

$$\varphi(\bar{x}) \leq \varphi(x) + \frac{\varepsilon}{\lambda} \|x - \bar{x}\| \quad \text{for all } x \in X.$$

Taking now any  $x^* \in \widehat{\partial}^+ \varphi(\bar{x}) = -\widehat{\partial}(-\varphi)(\bar{x})$  and using the smooth variational description of Fréchet subgradients from Theorem 1.88(i) held in arbitrary Banach spaces, we find a function  $s: X \rightarrow \mathbb{R}$  Fréchet differentiable at  $\bar{x}$  and such that

$$s(\bar{x}) = \varphi(\bar{x}), \quad \nabla s(\bar{x}) = x^* \quad \text{and} \quad s(x) \geq \varphi(x) \quad \text{whenever } x \in X.$$

Combining this with the above global minimization property for the perturbation of  $\varphi$  at  $\bar{x}$ , conclude that the function  $\phi(x) := s(x) + (\varepsilon/\lambda)\|x - \bar{x}\|$  attains its *global minimum* at  $\bar{x}$ . Then it follows from the generalized Fermat rule of Proposition 1.114, the sum rule of Proposition 1.107(i), and subdifferentiating the norm function at zero that

$$0 \in \widehat{\partial} \phi(\bar{x}) = \nabla s(\bar{x}) + \widehat{\partial} \left( \frac{\varepsilon}{\lambda} \|\cdot - \bar{x}\| \right)(\bar{x}) \subset x^* + \frac{\varepsilon}{\lambda} \mathbb{B}^*.$$

This gives  $\|x^*\| < \varepsilon/\lambda$  and completes the proof of the theorem.  $\triangle$

### 2.3.3 Smooth Variational Principles

The crucial condition (c) in Theorem 2.26 can be interpreted as follows: for every proper l.s.c. function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  bounded from below (i.e., such that  $\inf \varphi > -\infty$ ) there exist a point  $\bar{x} \in \text{dom } \varphi$  and a function  $s: X \rightarrow \mathbb{R}$  satisfying

$$\varphi(\bar{x}) = s(\bar{x}) \quad \text{and} \quad \varphi(x) \geq s(x) \quad \text{for all } x \in X. \quad (2.40)$$

The latter means that  $s(\cdot)$  “supports  $\varphi$  from below.” Such a function  $s(\cdot)$  is usually called a *supporting function* belonging to some class  $\mathcal{S}$ . In these words condition (2.40), with  $s(\cdot) \in \mathcal{S}$  for every l.s.c. function  $\varphi$  bounded from below, postulates that the  *$\mathcal{S}$ -variational principle* holds in  $X$ . Thus Ekeland's theorem ensures that, for the class of

$$\mathcal{S} := \left\{ -\varepsilon \|\cdot - \bar{x}\| + c \mid \varepsilon > 0, c \in \mathbb{R} \right\}$$

with arbitrary small positive numbers  $\varepsilon$ , the  $\mathcal{S}$ -variational principle holds in any Banach space. A notable limitation on applications of this result is that the supporting functions are not smooth.

If all  $s(\cdot) \in \mathcal{S}$  are required to be smooth (in some sense), we speak about a *smooth variational principle* in a Banach space  $X$ . An  $\mathcal{S}$ -variational principle is called *concave* if  $\mathcal{S}$  consists of concave functions. The afore-mentioned result



of Borwein and Preiss establishes a concave smooth variational principle provided that  $X$  admits a smooth renorm with respect to some bornology. The corresponding result of Deville, Godefroy and Zizler ensures a smooth (but not concave) variational principle when the smooth renorming assumption is weakened to the existence of a smooth Lipschitzian bump function on  $X$ .

In the following theorem we consider variational principles for the three classes of  $\mathcal{S}$ -smooth functions on  $X$ : Fréchet differentiable ( $\mathcal{S} = \mathcal{F}$ ), Lipschitzian and Fréchet differentiable ( $\mathcal{S} = \mathcal{LF}$ ), and Lipschitzian and continuously differentiable ( $\mathcal{S} = \mathcal{LC}^1$ ). Applying the lower subdifferential variational principle of Theorem 2.28 and then the variational descriptions of Fréchet subgradients established above, we derive  $\mathcal{S}$ -smooth variational principles in some *enhanced forms* under the corresponding smoothness assumptions on the Banach space in question, which inevitably imply the Asplund property of this space. Moreover, we show that the smoothness assumptions on  $X$  are *not only sufficient but also necessary* for the fulfillment of these smooth (resp. concave and smooth) variational principles in Asplund spaces.

**Theorem 2.31 (smooth variational principles in Asplund spaces).** *Let  $X$  be a Banach space, and let  $\mathcal{A}$  stand for the class of all proper l.s.c. functions  $\varphi: X \rightarrow \overline{\mathbb{R}}$  bounded from below. Given arbitrary  $\varepsilon > 0$  and  $\lambda > 0$ , one has the following assertions:*

(i) *If  $X$  admits a Fréchet smooth renorm, then for every  $\varphi \in \mathcal{A}$  and  $x_0 \in X$  with  $\varphi(x_0) < \inf_X \varphi + \varepsilon$  there exist  $\bar{x} \in X$  and a concave Fréchet differentiable function  $s: X \rightarrow \mathbb{R}$  such that*

$$\|\bar{x} - x_0\| < \lambda, \quad \varphi(\bar{x}) < \inf_X \varphi + \varepsilon, \quad (2.41)$$

*$\|\nabla s(\bar{x})\| < \varepsilon/\lambda$ , and*

$$\varphi(\bar{x}) = s(\bar{x}), \quad \varphi(x) \geq s(x) + \|x - \bar{x}\|^2 \text{ for all } x \in X.$$

(ii) *Let  $X$  admit an  $\mathcal{S}$ -smooth bump function, where  $\mathcal{S}$  stands for either  $\mathcal{F}$ ,  $\mathcal{LF}$ , or  $\mathcal{LC}^1$ . Then for every  $\varphi \in \mathcal{A}$  and  $x_0 \in X$  with  $\varphi(x_0) < \inf_X \varphi + \varepsilon$  there exist  $\bar{x} \in X$  satisfying (2.41), an  $\mathcal{S}$ -smooth bump  $b: X \rightarrow \mathbb{R}$ , and a constant  $c \in \mathbb{R}$  such that  $\|\nabla b(\bar{x})\| < \varepsilon/\lambda$  and*

$$\varphi(\bar{x}) = b(\bar{x}) + c, \quad \varphi(x) \geq b(x) + c \text{ for all } x \in X.$$

*Moreover, in this case we can find  $\mathcal{S}$ -smooth functions  $s: X \rightarrow \mathbb{R}$  and  $\theta: X \rightarrow [0, \infty)$  such that  $\|\nabla s(\bar{x})\| < \varepsilon/\lambda$ ,  $\theta(x) = 0$  only for  $x = 0$ ,  $\theta(x) \leq \|x\|^2$  if  $x \in \mathbb{B}$ , and*

$$\varphi(\bar{x}) = s(\bar{x}), \quad \varphi(x) \geq s(x) + \theta(x - \bar{x}) \text{ for all } x \in X.$$

(iii) *Conversely, the concave  $\mathcal{F}$ -smooth variational principle holds in  $X$  only if  $X$  admits a Fréchet smooth renorm, and the  $\mathcal{S}$ -smooth variational principle holds in  $X$  only if  $X$  admits an  $\mathcal{S}$ -smooth bump function for the corresponding classes  $\mathcal{S}$  listed above.*

**Proof.** Assertions (i) and (ii) follow directly from the lower subdifferential variational principle in Theorem 2.28(b) due to the variational descriptions of Fréchet subgradients in Theorem 1.88. Let us justify the converse statements formulated in (iii).

First we prove that the concave  $\mathcal{F}$ -smooth variational principle in  $X$  implies that  $X$  admits a Fréchet smooth renorm. Applying (2.40) to the function  $\varphi(x) := 1/\|x\|$ , we find  $0 \neq v \in X$  and a concave Fréchet differentiable function  $s: X \rightarrow \mathbb{R}$  such that

$$s(x) \leq \varphi(x) = 1/\|x\| < 1/(2\|v\|) \quad \text{if } \|x\| > 2\|v\| ,$$

with  $s(v) = 1/\|v\|$ . Putting

$$p(x) := -s(x + v) + 1/\|v\|, \quad x \in X ,$$

we conclude that  $p$  is convex and Fréchet differentiable on  $X$  due to the corresponding properties of  $s$ . Thus  $p$  is  $\mathcal{C}^1$ -smooth on  $X$ . Moreover, one has  $p(0) = 0$  and

$$p(x) > -1/(2\|v\|) + 1/\|v\| = 1/(2\|v\|) \quad \text{if } \|x\| > 3\|v\| ,$$

since  $\|x + v\| > 2\|v\|$ . Now let us consider the *Minkowski gauge functional*

$$g(x) := \inf \{ \lambda > 0 \mid x \in \lambda \Omega \}, \quad x \in X ,$$

of the set  $\Omega := \{x \in X \mid p(x) \leq 1/(2\|v\|)\}$ . It is easy to see that  $\Omega$  is convex, closed, and bounded with  $0 \in \text{int } \Omega$ . In this case the Minkowski gauge is a continuous sublinear functional with  $g(x) > 0$  for all  $x \neq 0$  and  $\Omega = \{x \in X \mid g(x) \leq 1\}$ . This ensures the existence of  $M > 0$  such that

$$\|x\|/(3\|v\|) \leq g(x) \leq M\|x\| \quad \text{for all } x \in X .$$

Now considering the function

$$n(x) := g(x) + g(-x), \quad x \in X ,$$

we conclude that it defines a *norm* on  $X$  equivalent to the original one  $\|\cdot\|$ . To complete the proof of the first statement in (iii), it remains to justify that  $g$  is Fréchet differentiable on  $X \setminus \{0\}$ . The crucial step for this is to show the Gâteaux differentiability of  $g$  at every nonzero point of  $X$ . Since  $g$  is convex, the latter is equivalent to the fact that its subdifferential  $\partial g(x)$  is a singleton for each  $x \in X \setminus \{0\}$ .

To proceed, we fix an arbitrary  $x \in X$  with  $g(x) = 1$  and pick  $x^* \in \partial g(x)$ . It can be easily derived from the definitions that

$$p(x) = 1/(2\|v\|) \quad \text{and} \quad \langle x^*, x \rangle = g(x) .$$

Now taking any  $t > 0$  and  $h \in X$  with  $\langle x^*, h \rangle = 0$ , one has

$$g(x + th) \geq g(x) + \langle x^*, th \rangle = 1, \quad g(\alpha(x + th)) = \alpha g(x + th) > 1 \text{ if } \alpha > 1,$$

and hence  $\alpha(x + th) \notin \Omega$ . Thus  $p(\alpha(x + th)) > 1/(2\|v\|)$  for all  $\alpha > 1$  and all  $t > 0$ . Passing to the limit as  $\alpha \downarrow 1$ , we get  $p(x + th) \geq 1/(2\|v\|)$  ( $= p(x)$ ) for all  $t > 0$ . Since  $p$  is Gâteaux differentiable at  $x$  with the derivative  $p'(x)$ , this implies that

$$\langle p'(x), h \rangle = \lim_{t \downarrow 0} \frac{p(x + th) - p(x)}{t} \geq 0 \text{ for all } h \in X \text{ with } \langle x^*, h \rangle = 0.$$

The latter gives  $\langle p'(x), h \rangle = 0$  for all such  $h$ , and so  $x^* = \lambda p'(x)$  for some  $\lambda \in \mathbb{R}$ . Therefore

$$1 = g(x) = \langle x^*, x \rangle = \lambda \langle p'(x), x \rangle,$$

which uniquely determines  $x^* \in \partial g(x)$  as  $x^* = p'(x)/\langle p'(x), x \rangle$ . This means that  $g$  is Gâteaux differentiable at  $x$  and  $g'(x) = x^*$  when  $g(x) = 1$ . Considering an arbitrary nonzero  $x \in X$  and taking into account that  $g$  is positively homogeneous and  $g(x) \neq 0$ , we get the following formula for the Gâteaux derivative of  $g$  at  $x$ :

$$g'(x) = \left\langle p'\left(\frac{x}{g(x)}\right), \frac{x}{g(x)} \right\rangle^{-1} p'\left(\frac{x}{g(x)}\right).$$

Since  $p$  is  $\mathcal{C}^1$ -smooth, this formula implies that  $g'$  is norm-to-norm continuous. Thus  $g$  is Fréchet differentiable at every nonzero point of  $X$ , which justifies the first part of (iii).

Next we prove the second part of (iii) simultaneously for each listed  $\mathcal{S}$ . Again pick the function  $\varphi = 1/\|\cdot\|$  and apply to it the supporting condition (2.40) with some  $v = \bar{x}$  and  $\mathcal{S}$ -smooth function  $s : X \rightarrow \mathbb{R}$ . Then consider an arbitrary  $\mathcal{C}^2$ -smooth function  $\tau : \mathbb{R} \rightarrow [0, 1]$  satisfying

$$\tau(t) = 1 \text{ if } t \geq 1/\|v\| \text{ and } \tau(t) = 0 \text{ if } t \leq 1/(2\|v\|).$$

One can easily check that  $b := \tau \circ s$  is an  $\mathcal{S}$ -smooth bump function on  $X$ , which justifies (iii).  $\triangle$

Note that the supporting conditions in assertions (i) and (ii) of Theorem 2.31 carry more information in comparison with the basic supporting condition (2.40) used in the proof of assertion (iii). Observe also that the proof of Theorem 2.31(iii) holds true when the Fréchet smoothness is replaced by the Gâteaux smoothness or, generally, by any  $\beta$ -smoothness with respect to an arbitrary bornology  $\beta$  on  $X$ ; cf. Remark 2.11. This implies that any smooth (resp. concave smooth) variational principle with the supporting condition (2.40) necessarily requires the corresponding smooth renorming/bump function assumption on the underlying Banach space  $X$ .

## 2.4 Representations and Characterizations in Asplund Spaces

In this section we apply the above extremal and variational principles to obtain efficient representations of the generalized differential constructions of Chap. 1 in the case of Asplund spaces. Most of these representations turn out to be characterizations of Asplund spaces. We begin with a subgradient description of the approximate extremal principle, which plays an essential role in the subsequent material. Then we derive characterizations of Asplund spaces in terms of special subdifferential sum rules involving Lipschitzian functions. This leads to simplified representations of basic subgradients, normals, and coderivatives in Asplund spaces similar to those in finite dimensions. In the last subsection we derive convenient representations of singular subgradients of extended-real-valued l.s.c. functions and related results for horizontal normals to graphs of continuous functions on Asplund spaces.

### 2.4.1 Subgradients, Normals, and Coderivatives in Asplund Spaces

Let  $\mathcal{SL}(\bar{x})$  denote the class of pairs  $(\varphi_1, \varphi_2)$  with proper functions  $\varphi_i: X \rightarrow \overline{\mathbb{R}}$  such that  $\varphi_1$  is Lipschitz continuous around  $\bar{x} \in \text{dom } \varphi_1 \cap \text{dom } \varphi_2$  and  $\varphi_2$  is l.s.c. around this point. For brevity we say that the sum  $\varphi_1 + \varphi_2$  is *semi-Lipschitzian* at  $\bar{x}$  if  $(\varphi_1, \varphi_2) \in \mathcal{SL}(\bar{x})$ . The next result provides an equivalent description of the approximate extremal principle in terms of a “fuzzy” subgradient condition for minimum points of semi-Lipschitzian sums.

**Lemma 2.32 (subgradient description of the extremal principle).**

*Given a Banach space  $X$ , one has the following:*

(i) *Let the approximate extremal principle hold for every extremal system of two closed sets in  $X \times \mathbb{R}$ . Assume that  $(\varphi_1, \varphi_2) \in \mathcal{SL}(\bar{x})$  with  $\varphi_i: X \rightarrow \overline{\mathbb{R}}$  and that the sum  $\varphi_1 + \varphi_2$  attains a local minimum at  $\bar{x}$ . Then for any  $\eta > 0$  there are  $x_i \in \bar{x} + \eta\mathbb{B}$  with  $|\varphi_i(x_i) - \varphi_i(\bar{x})| \leq \eta$ ,  $i = 1, 2$ , such that*

$$0 \in \widehat{\partial}\varphi_1(x_1) + \widehat{\partial}\varphi_2(x_2) + \eta\mathbb{B}^* . \quad (2.42)$$

(ii) *Conversely, let for any  $(\varphi_1, \varphi_2) \in \mathcal{SL}(\bar{x})$  with  $\varphi_i: X \rightarrow \overline{\mathbb{R}}$  and for any  $\eta > 0$  there exist  $x_i \in \bar{x} + \eta\mathbb{B}$  with  $|\varphi_i(x_i) - \varphi_i(\bar{x})| \leq \eta$ ,  $i = 1, 2$ , such that (2.42) is fulfilled provided that  $\varphi_1 + \varphi_2$  attains a local minimum at  $\bar{x}$ . Then the approximate extremal principle holds for every extremal system of two closed sets in  $X$ .*

**Proof.** To justify (i), we consider  $(\varphi_1, \varphi_2) \in \mathcal{SL}(\bar{x})$  and assume without loss of generality that  $\bar{x} = 0 \in X$  is a local minimizer for  $\varphi_1 + \varphi_2$  with  $\varphi_1(0) = \varphi_2(0) = 0$ , that  $\varphi_1$  is Lipschitz continuous on  $\eta\mathbb{B}$  with modulus  $\ell > 0$ , and that  $\varphi_2$  is l.s.c. on  $\eta\mathbb{B}$  for the fixed  $\eta > 0$ . Consider the sets

$$\Omega_1 := \text{epi } \varphi_1 \quad \text{and} \quad \Omega_2 := \{(x, \alpha) \in X \times \mathbb{R} \mid \varphi_2(x) \leq -\alpha\} ,$$

which are obviously closed around  $(0, 0) \in X \times \mathbb{R}$ . It is easy to check that  $(0, 0)$  is a local extremal point of the sets  $\{\Omega_1, \Omega_2\}$ , since  $\bar{x} = 0$  is a local minimizer for  $\varphi_1 + \varphi_2$ . Applying the approximate extremal principle to the system  $\{\Omega_1, \Omega_2, (0, 0)\}$ , for any  $\varepsilon > 0$  we find  $(x_i, \alpha_i) \in \Omega_i$  and  $(x_i^*, \lambda_i) \in X^* \times \mathbb{R}$ ,  $i = 1, 2$ , such that

$$(x_1^*, -\lambda_1) \in \widehat{N}((x_1, \alpha_1); \Omega_1), \quad (-x_2^*, \lambda_2) \in \widehat{N}((x_2, \alpha_2); \Omega_2), \quad (2.43)$$

$$\|(x_i, \alpha_i)\| \leq \varepsilon, \quad \frac{1}{2} - \varepsilon \leq \|(x_i^*, \lambda_i)\| \leq \frac{1}{2} + \varepsilon, \quad i = 1, 2, \quad (2.44)$$

$$\|(x_1^*, -\lambda_1) + (-x_2^*, \lambda_2)\| \leq \varepsilon. \quad (2.45)$$

It follows from (2.43) that  $\lambda_i \geq 0$  for  $i = 1, 2$ . Our goal is to show that choosing  $\varepsilon$  to be sufficiently small, we get  $\lambda_i > 0$  and can equivalently transformed (2.43) to subgradient relations with the required estimates. For these purposes it is convenient to define the corresponding norms on  $X \times \mathbb{R}$  and  $X^* \times \mathbb{R}$  by

$$\|(x, \alpha)\| := \max\{\|x\|, |\alpha|\} \quad \text{and} \quad \|(x^*, \lambda)\| := \|x^*\| + |\lambda|.$$

Then choose  $\varepsilon$  in (2.43)–(2.45) satisfying

$$0 < \varepsilon < \min\left\{\frac{1}{4(2+\ell)}, \frac{\eta}{4(1+\ell)^2}\right\}.$$

Since  $\varphi_1$  is Lipschitz continuous on  $\eta\mathcal{B}$ , we get from  $(x_1^*, -\lambda_1) \in \widehat{N}((x_1, \alpha_1); \Omega_1)$  with  $\max\{\|x_1\|, |\alpha_1|\} \leq \varepsilon < \eta$  that  $\|x_1^*\| \leq \ell\lambda_1$ ; see Proposition 1.85(ii). It gives by (2.44) and (2.45) that

$$\lambda_1 \geq \frac{1}{2(1+\ell)} - \frac{\varepsilon}{1+\ell} > 0 \quad \text{and} \quad \lambda_2 \geq \frac{1}{2(1+\ell)} - \varepsilon\left(\frac{2+\ell}{1+\ell}\right) > \frac{1}{4(1+\ell)}$$

by the choice of  $\varepsilon$ . This implies by (2.43) that  $\alpha_1 = \varphi_1(x_1)$ ,  $\alpha_2 = -\varphi_2(x_2)$ , and hence

$$\widetilde{x}_1^* := x_1^*/\lambda_1 \in \widehat{\partial}\varphi_1(x_1), \quad \widetilde{x}_2^* := -x_2^*/\lambda_2 \in \widehat{\partial}\varphi_2(x_2).$$

By (2.44) we have

$$\|x_i\| \leq \varepsilon < \eta \quad \text{and} \quad |\varphi_i(x_i)| = |\alpha_i| \leq \varepsilon < \eta, \quad i = 1, 2.$$

To justify (2.42), it remains to show that  $\|\widetilde{x}_1^* + \widetilde{x}_2^*\| \leq \eta$ . This follows from

$$\begin{aligned} \left\| \frac{x_1^*}{\lambda_1} - \frac{x_2^*}{\lambda_2} \right\| &= \left\| \frac{x_1^*(\lambda_2 - \lambda_1)}{\lambda_1 \lambda_2} + \frac{x_1^* - x_2^*}{\lambda_2} \right\| \leq \frac{\|x_1^*\|}{\lambda_1} \left( \frac{|\lambda_2 - \lambda_1|}{\lambda_2} \right) + \frac{\|x_1^* - x_2^*\|}{\lambda_2} \\ &\leq \ell \frac{\varepsilon}{\lambda_2} + \frac{\varepsilon}{\lambda_2} = \frac{\varepsilon}{\lambda_2} (1 + \ell) < 4\varepsilon(1 + \ell)^2 < \eta \end{aligned}$$

due to the choice of  $\varepsilon$  and the estimates above.

Next let us prove the converse assertion (ii). Take an extremal system  $\{\Omega_1, \Omega_2, \bar{x}\}$  in  $X$  and find a neighborhood  $U$  of  $\bar{x}$  such that, given an arbitrary  $\varepsilon > 0$ , there is  $a \in X$  with  $\|a\| < \varepsilon^2/2$  and  $(\Omega_1 + a) \cap \Omega_2 \cap U = \emptyset$ . Put  $U = X$  for simplicity and define the function  $\varphi: X \times X \rightarrow \mathbb{R}$  by

$$\varphi(u, v) := \frac{1}{2}\|u - v + a\|, \quad (u, v) \in X^2. \quad (2.46)$$

It follows from the local extremality of  $\bar{x}$  that  $\varphi(\bar{x}, \bar{x}) < (\varepsilon/2)^2$  and that  $\varphi(u, v) > 0$  for all  $u \in \Omega_1$  and  $v \in \Omega_2$ .

Now we apply Ekeland's variational principle in Theorem 2.26(i) to the function  $\varphi$  on the complete metric space  $\Omega_1 \times \Omega_2$  whose metric is induced by the norm  $\|(u, v)\| := \|u\| + \|v\|$  on  $X^2$ . This gives points  $(\bar{u}, \bar{v}) \in \Omega_1 \times \Omega_2$  such that  $\|\bar{u} - \bar{x}\| \leq \varepsilon/2$ ,  $\|\bar{v} - \bar{x}\| \leq \varepsilon/2$ , and

$$\varphi(\bar{u}, \bar{v}) \leq \varphi(u, v) + \frac{\varepsilon}{2}(\|u - \bar{u}\| + \|v - \bar{v}\|) \quad \text{for all } (u, v) \in \Omega_1 \times \Omega_2.$$

The latter means that the sum of the functions

$$\varphi_1(u, v) := \varphi(u, v) + \frac{\varepsilon}{2}(\|u - \bar{u}\| + \|v - \bar{v}\|) \quad \text{and} \quad \varphi_2(u, v) := \delta((u, v); \Omega_1 \times \Omega_2)$$

attains at  $(\bar{u}, \bar{v})$  its minimum over  $X^2$ . Observe that  $\varphi_1$  is Lipschitz continuous and *convex* and that  $\varphi_2$  is proper and l.s.c. on  $X^2$ . By the assumptions in (ii) we find  $(y_1, y_2) \in X^2$  and  $(x_1, x_2) \in \Omega_1 \times \Omega_2$  such that  $\|x_1 - \bar{u}\| \leq \varepsilon/2$ ,  $\|x_2 - \bar{v}\| \leq \varepsilon/2$ ,  $\varphi(y_1, y_2) > 0$ , and

$$0 \in \widehat{\partial}\varphi_1(y_1, y_2) + \widehat{\partial}\varphi_2(x_1, x_2) + \frac{\varepsilon}{2}(\mathcal{B}^* \times \mathcal{B}^*).$$

Note that  $\widehat{\partial}\varphi_2(x_1, x_2) = \widehat{N}((x_1, x_2); \Omega_1 \times \Omega_2) = \widehat{N}(x_1; \Omega_1) \times \widehat{N}(x_2; \Omega_2)$  due to Proposition 1.2. Now using the well-known subdifferential formula for the norm function (2.46) at nonzero points, we conclude that

$$\widehat{\partial}\varphi_1(y_1, y_2) = \frac{1}{2}(x^*, -x^*) + \frac{\varepsilon}{2}(\mathcal{B}^* \times \mathcal{B}^*)$$

with some  $x^* \in X^*$  of the unit norm. Finally, putting  $x_1^* := -x^*/2$  and  $x_2^* := x^*/2$ , we get  $x_i^* \in \widehat{N}(x_i; \Omega_i) + \varepsilon\mathcal{B}^*$  with  $x_1^* + x_2^* = 0$  and  $\|x_1^*\| + \|x_2^*\| = 1$ , which justifies (ii).  $\triangle$

Next we obtain two subdifferential sum rules in the semi-Lipschitzian case: the *fuzzy* rule for Fréchet subgradients and  $\varepsilon$ -subgradients and the *exact* one for basic subgradients. Each of these rules applied to all semi-Lipschitzian sums is proved to be a characterization of Asplund spaces.

**Theorem 2.33 (semi-Lipschitzian sum rules).** *Let  $X$  be a Banach space with  $\bar{x} \in X$ . The following properties are equivalent:*

(a)  *$X$  is Asplund.*

(b) For any  $(\varphi_1, \varphi_2) \in \mathcal{SL}(\bar{x})$ , for any  $\varepsilon \geq 0$ , and for any  $\gamma > 0$  one has

$$\widehat{\partial}_\varepsilon(\varphi_1 + \varphi_2)(\bar{x}) \subset \bigcup \left\{ \widehat{\partial}\varphi_1(x_1) + \widehat{\partial}\varphi_2(x_2) \mid x_i \in \bar{x} + \gamma \mathcal{B}, \right. \\ \left. |\varphi_i(x_i) - \varphi_i(\bar{x})| \leq \gamma, i = 1, 2 \right\} + (\varepsilon + \gamma) \mathcal{B}^* .$$

(c) For any  $(\varphi_1, \varphi_2) \in \mathcal{SL}(\bar{x})$  one has

$$\partial(\varphi_1 + \varphi_2)(\bar{x}) \subset \partial\varphi_1(\bar{x}) + \partial\varphi_2(\bar{x}) .$$

**Proof.** First we prove (a) $\Rightarrow$ (b). Observe that if  $X$  is Asplund, then  $X \times \mathbb{R}$  is Asplund as well. By Theorem 2.20 the approximate extremal principle holds in  $X \times \mathbb{R}$ . Hence we have property (2.42) in Lemma 2.32 for any  $(\varphi_1, \varphi_2) \in \mathcal{SL}(\bar{x})$ . Let us derive (b) from this property and from the variational description of analytic  $\varepsilon$ -subgradients in Proposition 1.84. Fix  $(\varepsilon, \gamma)$  in (b) and find  $\eta$  satisfying the relations

$$0 < \eta < \min \{ \gamma/4, \bar{\eta} \}, \text{ where } \bar{\eta}^2 + (2 + \varepsilon)\bar{\eta} - \gamma = 0 .$$

Then pick an arbitrary  $x^* \in \widehat{\partial}_\varepsilon(\varphi_1 + \varphi_2)(\bar{x})$  and conclude by Proposition 1.84(ii) that the sum

$$\left( \varphi_1(x) - \langle x^*, x - \bar{x} \rangle + (\varepsilon + \eta) \|x - \bar{x}\| \right) + \varphi_2(x)$$

attains a local minimum at  $\bar{x}$ . Applying (2.42) with the chosen  $\eta$  to the above sum and then using the elementary sum rule in Proposition 1.107(i), we find  $x_i \in \bar{x} + \eta \mathcal{B}$  and  $x_i^* \in X^*$ ,  $i = 1, 2$ , such that

$$\left| \varphi_1(x_1) + (\varepsilon + \eta) \|x_1 - \bar{x}\| - \varphi_1(\bar{x}) \right| \leq \eta, \quad |\varphi_2(x_2) - \varphi_2(\bar{x})| \leq \eta ,$$

$$x_1^* \in \widehat{\partial} \left( \varphi_1 + (\varepsilon + \eta) \|\cdot - \bar{x}\| \right) (x_1), \quad x_2^* \in \widehat{\partial}\varphi_2(x_2) ,$$

and  $x^* - x_1^* - x_2^* \in \eta \mathcal{B}^*$ . This implies that

$$|\varphi_1(x_1) - \varphi_1(\bar{x})| \leq \eta(\varepsilon + \eta + 1) .$$

Now employing Proposition 1.84(ii) in the case of the Fréchet subgradient  $x_1^*$ , we conclude that the sum  $\varphi_1 + \psi$  with

$$\psi(x) := (\varepsilon + \eta) \|x - \bar{x}\| - \langle x_1^*, x - x_1 \rangle + \eta \|x - x_1\|$$

attains a local minimum at  $x_1$ . Observe that  $\psi$  is convex and continuous on  $X$  with  $\partial\psi(x) \subset -x_1^* + (\varepsilon + 2\eta) \mathcal{B}^*$  for any  $x \in X$ . Applying (2.42) to  $\varphi_1 + \psi$ , we find  $\tilde{x}_1 \in x_1 + \eta \mathcal{B}$  such that

$$|\varphi_1(\tilde{x}_1) - \varphi_1(x_1)| \leq \eta \text{ and } x_1^* \in \widehat{\partial}\varphi_1(\tilde{x}_1) + (\varepsilon + 3\eta) \mathcal{B}^* .$$

We finally have

$$x^* \in \widehat{\partial}\varphi_1(\tilde{x}_1) + \widehat{\partial}\varphi_2(x_2) + (\varepsilon + 4\eta)\mathcal{B}^*$$

with  $\|\tilde{x}_1 - \bar{x}\| \leq 2\eta$  and  $|\varphi_1(\tilde{x}_1) - \varphi_1(\bar{x})| \leq \eta(\varepsilon + \eta + 2)$ . This gives (b) by the choice of  $\eta$ .

Next let us prove that (b) and the Asplund property of  $X$  implies (c). Take an arbitrary  $x^* \in \partial(\varphi_1 + \varphi_2)(\bar{x})$  and by representation (1.55) in Theorem 1.89 find sequences  $\varepsilon_k \downarrow 0$ ,  $x_k \rightarrow \bar{x}$  with  $\varphi_1(x_k) + \varphi_2(x_k) \rightarrow \varphi_1(\bar{x}) + \varphi_2(\bar{x})$ , and  $x_k^* \xrightarrow{w^*} x^*$  such that  $x_k^* \in \widehat{\partial}_{\varepsilon_k}(\varphi_1 + \varphi_2)(x_k)$  as  $k \rightarrow \infty$ . Then employing (b) with  $\gamma_k = \varepsilon_k$ , we get sequences  $x_{ik} \rightarrow \bar{x}$  with  $\varphi_i(x_{ik}) \rightarrow \varphi_i(\bar{x})$  and  $x_{ik}^* \in \widehat{\partial}\varphi_i(x_{ik})$ ,  $i = 1, 2$ , such that

$$\|x_k^* - x_{1k}^* - x_{2k}^*\| \leq 2\varepsilon_k \quad \text{for all } k \in \mathbb{N}. \quad (2.47)$$

Since  $x_k^* \rightarrow x^*$ , this sequence is bounded in  $X^*$  due to the uniform boundedness principle. The sequence  $\{x_{1k}^*\}$  is also bounded by modulus  $\ell$  due to the Lipschitz continuity of  $\varphi_1$  around  $\bar{x}$ ; see Proposition 1.85(ii). Hence  $\{x_{2k}^*\}$  is bounded as well. Using the weak\* sequential compactness of bounded sets in duals to Asplund spaces, we find  $x_i^* \in X^*$  such that  $x_{ik}^* \xrightarrow{w^*} x_i^*$ ,  $i = 1, 2$ , along a subsequence of  $k \rightarrow \infty$ . Again employing Theorem 1.89, we get  $x_i^* \in \partial\varphi_i(\bar{x})$  for  $i = 1, 2$ . Moreover, (2.47) implies that  $x^* = x_1^* + x_2^*$ , which gives (c).

It remains to show that each of the properties (b) and (c) implies that  $X$  is Asplund. Indeed, according to Proposition 2.18 and Example 2.19 for any non-Asplund space  $X$  there is an equivalent norm  $|\cdot|$  on  $X$  such that

$$\widehat{\partial}\varphi(x) = \partial\varphi(x) = \emptyset \quad \text{whenever } x \in X$$

for  $\varphi := -|\cdot|$ . Now we can see that both properties (b) and (c) are violated for the sum  $\varphi_1 + \varphi_2$  with  $\varphi_1 := |\cdot|$  and  $\varphi_2 := -|\cdot|$ .  $\triangle$

The next theorem contains subdifferential characterizations of Asplund spaces via a simplified limiting representation of basic subgradients (like in finite-dimensions) and a related expansion formula for the so-called *limiting  $\varepsilon$ -subdifferential* of  $\varphi: X \rightarrow \overline{\mathbb{R}}$  at  $\bar{x} \in X$  with  $|\varphi(\bar{x})| < \infty$  defined by

$$\partial_\varepsilon\varphi(\bar{x}) := \text{Lim sup}_{x \xrightarrow{\varphi} \bar{x}} \widehat{\partial}_\varepsilon\varphi(x). \quad (2.48)$$

**Theorem 2.34 (subdifferential representations in Asplund spaces).**

Let  $X$  be a Banach space,  $\bar{x} \in X$ , and  $\mathcal{A}(\bar{x})$  be the class of proper functions  $\varphi: X \rightarrow \overline{\mathbb{R}}$  l.s.c. around  $\bar{x} \in \text{dom } \varphi$ . The following properties are equivalent:

- (a)  $X$  is Asplund.
- (b) For every  $\bar{x} \in X$  and every  $\varphi \in \mathcal{A}(\bar{x})$  one has

$$\partial\varphi(\bar{x}) = \text{Lim sup}_{x \xrightarrow{\varphi} \bar{x}} \widehat{\partial}\varphi(x).$$



(c) For every  $\bar{x} \in X$ , every  $\varphi \in \mathcal{A}(\bar{x})$ , and every  $\varepsilon > 0$  one has

$$\partial_\varepsilon \varphi(\bar{x}) = \partial \varphi(\bar{x}) + \varepsilon \mathcal{B}^* .$$

**Proof.** To justify (a) $\Rightarrow$ (b), we use the fuzzy sum rule in Theorem 2.33(b) with  $\varphi_1 = 0$  and  $\varphi_2 = \varphi$ . This gives

$$\widehat{\partial}_\varepsilon \varphi(\bar{x}) \subset \bigcup \left\{ \widehat{\partial} \varphi(x) \mid x \in \bar{x} + \gamma \mathcal{B}, \ |\varphi(x) - \varphi(\bar{x})| \leq \gamma \right\} + (\varepsilon + \gamma) \mathcal{B}^* \quad (2.49)$$

for any  $\varepsilon \geq 0$  and  $\gamma > 0$ . Passing there to the limit as  $\varepsilon = \gamma \downarrow 0$ , we arrive at the subdifferential representation (b).

To prove (a) $\Rightarrow$ (c), observe that the inclusion “ $\supset$ ” in (c) is trivial, and we need to show that the opposite inclusion holds in Asplund spaces. Pick  $x^* \in \partial_\varepsilon \varphi(\bar{x})$  and find by (2.48) sequences  $x_k \xrightarrow{\varphi} \bar{x}$  and  $x_k^* \xrightarrow{w^*} x^*$  such that  $x_k^* \in \widehat{\partial}_\varepsilon \varphi(x_k)$  for all  $k \in \mathbb{N}$ . Taking any  $\gamma_k \downarrow 0$  and using (2.49) with  $\gamma = \gamma_k$ , one gets  $u_k \in x_k + \gamma_k \mathcal{B}$  satisfying  $|\varphi(u_k) - \varphi(x_k)| \leq \gamma_k$  and

$$x_k^* \in \widehat{\partial} \varphi(u_k) + (\varepsilon + \gamma_k) \mathcal{B}^*, \quad k \in \mathbb{N} .$$

This allows us to find  $u_k^* \in \widehat{\partial} \varphi(u_k)$  and  $v_k^* \in (\varepsilon + \gamma_k) \mathcal{B}^*$  such that  $x_k^* = u_k^* + v_k^*$  for all  $k \in \mathbb{N}$ . By the weak\* sequential compactness of  $\mathcal{B}^*$  and the weak\* lower semicontinuity of  $\|\cdot\|$  on  $X^*$  we have  $v^* \in X^*$  satisfying

$$v_k^* \xrightarrow{w^*} v^* \text{ as } k \rightarrow \infty \text{ with } \|v^*\| \leq \liminf_{k \rightarrow \infty} \|v_k^*\| \leq \varepsilon$$

along a subsequence of  $\{k\}$ . This implies the existence of  $u^* \in \partial \varphi(\bar{x})$  such that  $u_k^* \xrightarrow{w^*} u^*$  and hence  $x^* = u^* + v^* \in \partial \varphi(\bar{x}) + \varepsilon \mathcal{B}^*$ , which gives (c).

To justify the opposite inclusion (c) $\Rightarrow$ (a), one has to show that for any non-Asplund space  $X$  there are  $\bar{x} \in X$ ,  $\varphi \in \mathcal{A}(\bar{x})$ , and  $\bar{\varepsilon} > 0$  such that the representation in (c) doesn’t hold. Taking the equivalent norm  $|\cdot|$  on  $X$  and the number  $\vartheta > 0$  in Proposition 2.18, let us show that this representation is violated for  $\varphi = -|\cdot|$ ,  $\bar{x} = 0$ , and  $\bar{\varepsilon} = 1$ . Indeed, it follows from Proposition 2.18 and Definition 1.83(ii) that

$$\widehat{\partial}_\varepsilon \varphi(x) = \emptyset \text{ for all } x \in X \text{ if } 0 \leq \varepsilon < \min \{1, \vartheta/2\} ,$$

which gives  $\partial \varphi(0) = \emptyset$ . On the other hand, one can easily check that  $\widehat{\partial}_1 \varphi(0) \supset \{0\} \neq \emptyset$ . Hence  $\partial_1 \varphi(0) \neq \emptyset$  by (2.48), and thus (c) doesn’t hold. Note that our proof actually shows more: if  $X$  is not Asplund, then for *any given*  $\varepsilon > 0$  there is a function  $\varphi \in \mathcal{A}(0)$  such that the representation in (c) is violated. Indeed, consider the function  $\varphi := -\varepsilon |\cdot|$  in the above arguments.

To finish the proof of the theorem, it remains to justify (b) $\Rightarrow$ (a), i.e., to show that the representation in (b) is violated for some  $\bar{x} \in X$  and some  $\varphi \in \mathcal{A}(\bar{x})$  in any non-Asplund space. Assuming that  $X$  is not Asplund, we take the equivalent norm  $|\cdot|$  in Proposition 2.18,  $\bar{x} = 0$ , and let

$$\varphi(x) := -|x|^2 + \min \{ \langle u^*, x \rangle, \langle v^*, x \rangle \}, \quad x \in X, \quad (2.50)$$

where  $u^*, v^* \in X^*$  with  $u^* \neq v^*$ . Consider a sequence  $\{x_k\} \subset X$  such that  $x_k \rightarrow 0$  and  $\langle u^*, x_k \rangle < \langle v^*, x_k \rangle$  for all  $k \in \mathbb{N}$ . Denote  $\psi(x) := -|x|^2$  and observe that

$$\varphi(x) = \psi(x) + \langle u^*, x \rangle \quad \text{whenever } x \in U_k \text{ and } k \in \mathbb{N}$$

for some neighborhood  $U_k$  of  $x_k$ . Since  $|\cdot| \leq \|\cdot\|$ , we have

$$|\psi(u) - \psi(v)| = (|u| + |v|) \cdot |(|u| - |v|)| \leq 3|x_k| \cdot |u - v|$$

for all  $u, v \in x_k + (\|x_k\|/2)\mathcal{B}$ . This means that the function  $\psi$  is Lipschitzian around  $x_k$  with modulus  $3|x_k|$  for any fixed  $k \in \mathbb{N}$ . It easily follows from the definitions that

$$u^* \in \widehat{\partial}_{3|x_k|}\varphi(x_k) \quad \text{for all } k \in \mathbb{N},$$

where the analytic  $\varepsilon$ -subdifferential is taken with respect to the norm  $|\cdot|$ . Passing to the limit as  $k \rightarrow \infty$  and taking into account that representation (1.55) is invariant with respect to equivalent norms on  $X$ , we get  $u^* \in \partial\varphi(0)$ .

Let us show that  $\widehat{\partial}\varphi(x) = \emptyset$  for all  $x$  near the origin, which violates (b) in the case of  $\varphi$  in (2.50) and  $\bar{x} = 0$ . First check that  $\widehat{\partial}\varphi(0) = \emptyset$ . Assuming the contrary, we get  $x^* \in \widehat{\partial}\varphi(0)$  satisfying

$$\liminf_{h \rightarrow 0} \frac{1}{\|h\|} \left[ -|h|^2 + \min \{ \langle u^*, h \rangle, \langle v^*, h \rangle \} - \langle x^*, h \rangle \right] \geq 0.$$

Since the norms  $|\cdot|$  and  $\|\cdot\|$  are equivalent on  $X$ , we conclude that  $\lim_{h \rightarrow 0} |h|^2/\|h\| = 0$  and hence

$$\liminf_{h \rightarrow 0} \frac{1}{\|h\|} \langle u^* - x^*, h \rangle \geq 0, \quad \liminf_{h \rightarrow 0} \frac{1}{\|h\|} \langle v^* - x^*, h \rangle \geq 0.$$

The latter is possible only when  $u^* = x^* = v^*$ , which contradicts the initial assumption that  $u^* \neq v^*$ ; thus  $\widehat{\partial}\varphi(0) = \emptyset$ .

Let us finally show that  $\widehat{\partial}\varphi(x) = \emptyset$  for any  $x \neq 0$ . If it is not the case, we take  $x^* \in \widehat{\partial}\varphi(x)$  and get from (2.50) that

$$\begin{aligned} \liminf_{h \rightarrow 0} \frac{1}{\|h\|} \left[ -|x+h|^2 + |x|^2 + \min \{ \langle u^*, x+h \rangle, \langle v^*, x+h \rangle \} \right. \\ \left. - \min \{ \langle u^*, x \rangle, \langle v^*, x \rangle \} - \langle x^*, h \rangle \right] \geq 0. \end{aligned}$$

Assume first that  $\langle u^*, x \rangle \leq \langle v^*, x \rangle$ . Then

$$\liminf_{h \rightarrow 0} \frac{1}{\|h\|} \left[ -|x+h|^2 + |x|^2 + \langle u^* - x^*, h \rangle \right] \geq 0,$$

which means that  $\widehat{\partial}(-|\cdot|^2)(x) \neq \emptyset$ . Since  $|\cdot|^2$  is convex and continuous, one always has  $\widehat{\partial}(|\cdot|^2)(x) \neq \emptyset$ . By Proposition 1.87 the function  $|\cdot|^2$  is Fréchet differentiable at  $x$ , which implies the Fréchet differentiability of  $|\cdot|$  at  $x \neq 0$ . The latter contradicts Proposition 2.18. The case of  $\langle u^*, x \rangle > \langle v^*, x \rangle$  can be considered similarly. Thus  $\widehat{\partial}\varphi(x) = \emptyset$  for any  $x \in X$ , which justifies (b) $\Rightarrow$ (a) and completes the proof of the theorem.  $\triangle$

The next result related to Theorem 2.34 gives an efficient representation of basic normals to closed sets via weak\* sequential limits of Fréchet normals at points nearby. It also happens to be a characterization of Asplund spaces.

**Theorem 2.35 (basic normals in Asplund spaces).** *Let  $X$  be a Banach space. The following properties are equivalent:*

- (a)  $X$  is Asplund.
- (b) For every closed set  $\Omega \subset X$  and every  $\bar{x} \in \Omega$  one has the limiting representation

$$N(\bar{x}; \Omega) = \limsup_{x \rightarrow \bar{x}} \widehat{N}(x; \Omega).$$

**Proof.** Implication (a) $\Rightarrow$ (b) follows from (a) $\Rightarrow$ (b) in Theorem 2.34 for the case of set indicator functions  $\varphi(x) = \delta(x; \Omega)$ . It remains to prove that if  $X$  is not Asplund, representation (b) of basic normals doesn't hold for some closed set  $\Omega \subset X$  and  $\bar{x} \in \Omega$ .

Put  $X = Z \times \mathbb{R}$ , where  $Z$  must be non-Asplund as well. Taking two distinct elements  $u^*$  and  $v^*$  of  $Z^*$ , define a Lipschitz function  $\varphi: Z \rightarrow \mathbb{R}$  by (2.50), where  $|\cdot|$  is the equivalent norm on  $Z$  from Proposition 2.18. We proved in Theorem 2.34 that  $\widehat{\partial}\varphi(z) = \emptyset$  for every  $z \in Z$ . Now let us consider the epigraphical set  $\Omega := \text{epi } \varphi \subset X$  generated by this function and show that  $\widehat{N}(x; \Omega) = \{0\}$  for every  $x \in \Omega$ .

It suffices to prove that  $\widehat{N}((z, \varphi(z)); \Omega) = \{(0, 0)\}$  for all  $z \in Z$ . Assuming the contrary and taking into account that  $\varphi$  is Lipschitzian, we find

$$(z^*, \lambda) \in \widehat{N}((z, \varphi(z)); \Omega) \text{ with } \lambda < 0$$

due to Proposition 1.85(ii) as  $\varepsilon = 0$ , which gives  $(-z^*/\lambda) \in \widehat{\partial}\varphi(z)$ . This contradicts the fact that  $\widehat{\partial}\varphi(z) = \emptyset$  proved in Theorem 2.34. Therefore

$$\limsup_{x \rightarrow \bar{x}} \widehat{N}(x; \Omega) = \{0\} \text{ whenever } \bar{x} \in \Omega$$

for the set  $\Omega$  under consideration. On the other hand, from the proof of (b) $\Rightarrow$ (a) in Theorem 2.34 we have  $z_k \in Z$  and  $\varepsilon_k > 0$  such that

$$u^* \in \widehat{\partial}_{\varepsilon_k} \varphi(z_k) \text{ with } \varepsilon_k \downarrow 0 \text{ and } z_k \rightarrow 0 \text{ as } k \rightarrow \infty.$$

It implies that  $(u^*, -1) \in \widehat{N}_{\varepsilon_k}((z_k, \varphi(z_k)); \Omega)$  due to Theorem 1.86 and hence  $(u^*, -1) \in N((0, 0); \Omega)$  by definition (1.3). Thus the basic normal representation in (b) is violated for the above set  $\Omega$  at the point  $\bar{x} = 0$ .  $\triangle$

Note that, for any Asplund space  $X$ , the subdifferential representation in Theorem 2.34(b) follows from the normal cone representation of Theorem 2.35 applied to epigraphical sets in the Asplund space  $X \times \mathbb{R}$ . The latter one is implied by the formula

$$\widehat{N}_\varepsilon(\bar{x}; \Omega) \subset \bigcup \left\{ \widehat{N}(x; \Omega) \mid x \in \Omega \cap (\bar{x} + \gamma \mathbb{B}) \right\} + (\varepsilon + \gamma) \mathbb{B}^* \quad (2.51)$$

held for every  $\varepsilon \geq 0$ ,  $\gamma > 0$ ,  $\bar{x} \in \Omega$ , and every closed subset  $\Omega \subset X$  of an Asplund space. Formula (2.51) immediately follows from (2.49) with  $\varphi = \delta(\cdot; \Omega)$  and, given any  $x^* \in \widehat{N}_\varepsilon(\bar{x}; \Omega)$ , it can also be obtained by the direct application of the approximate extremal principle to the system of two closed sets

$$\Omega_1 := \{(x, \alpha) \in X \times \mathbb{R} \mid x \in \Omega, \alpha \geq 0\},$$

$$\Omega_2 := \{(x, \alpha) \in X \times \mathbb{R} \mid x \in X, \alpha \leq \langle x^*, x - \bar{x} \rangle - (\varepsilon + \gamma) \|x - \bar{x}\|\}$$

for which  $(\bar{x}, 0)$  is a local extremal point.

As a consequence of Theorem 2.35, we have the following simplified representations (with  $\varepsilon = 0$  in Definition 1.32) of both normal and mixed coderivatives for closed-graph multifunctions between Asplund spaces.

**Corollary 2.36 (coderivatives of mappings between Asplund spaces).**

*Let  $F: X \rightrightarrows Y$  be a multifunction between Asplund spaces whose graph is closed around  $(\bar{x}, \bar{y}) \in \text{gph } F$ . Then*

$$D_N^* F(\bar{x}, \bar{y})(\bar{y}^*) = \limsup_{\substack{(x, y) \rightarrow (\bar{x}, \bar{y}) \\ y^* \xrightarrow{w^*} \bar{y}^*}} \widehat{D}^* F(x, y)(y^*), \quad \bar{y}^* \in Y^*,$$

$$D_M^* F(\bar{x}, \bar{y})(\bar{y}^*) = \limsup_{\substack{(x, y) \rightarrow (\bar{x}, \bar{y}) \\ y^* \rightarrow \bar{y}^*}} \widehat{D}^* F(x, y)(y^*), \quad \bar{y}^* \in Y^*.$$

**Proof.** Since both  $X$  and  $Y$  are Asplund, its product  $X \times Y$  is Asplund as well. Hence the representation for  $D_N^* F(\bar{x}, \bar{y})$  follows immediately from (1.26) and the normal cone representation of Theorem 2.35 applied to  $\Omega = \text{gph } F \subset X \times Y$ . To prove the mixed coderivative representation, we pick any  $\bar{x}^* \in D_M^* F(\bar{x}, \bar{y})(\bar{y}^*)$  and find, by Definition 1.32(iii), sequences  $\varepsilon_k \downarrow 0$ ,  $(x_k, y_k, y_k^*) \rightarrow (\bar{x}, \bar{y}, \bar{y}^*)$ , and  $x_k^* \xrightarrow{w^*} \bar{x}^*$  with  $(x_k, y_k) \in \text{gph } F$  and

$$(x_k^*, -y_k^*) \in \widehat{N}_{\varepsilon_k}((x_k, y_k); \text{gph } F) \quad \text{for all } k \in \mathbb{N}.$$

Now using formula (2.51) with  $\varepsilon = \gamma := \varepsilon_k$  and  $\Omega = \text{gph } F$ , we get sequences  $(\tilde{x}_k, \tilde{y}_k) \in \text{gph } F$  and  $(\tilde{x}_k^*, -\tilde{y}_k^*) \in \widehat{N}((\tilde{x}_k, \tilde{y}_k); \text{gph } F)$  such that

$$\|(\tilde{x}_k, \tilde{y}_k) - (x_k, y_k)\| \leq \varepsilon_k \quad \text{and} \quad \|(\tilde{x}_k^*, \tilde{y}_k^*) - (x_k^*, y_k^*)\| \leq 2\varepsilon_k.$$

This implies that  $\tilde{x}_k^* \xrightarrow{w^*} \bar{x}^*$  and that  $(\tilde{x}_k, \tilde{y}_k, \tilde{y}_k^*) \rightarrow (\bar{x}, \bar{y}, \bar{y}^*)$  in the norm topology of  $X \times Y \times Y^*$ , which justifies the representation for  $D_M^* F(\bar{x}, \bar{y})$ .  $\triangle$

### 2.4.2 Representations of Singular Subgradients and Horizontal Normals to Graphs and Epigraphs

In Subsect. 1.3.1 we defined singular subgradients of extended-real-valued functions through horizontal normals to their epigraphs. For a number of applications of singular subgradients it is important to obtain their efficient representations via some limits of Fréchet subgradients and  $\varepsilon$ -subgradients at points nearby, similar to those available for basic subgradients. This issue is related to the possibility of approximating horizontal normals by sequences of sloping (non-horizontal) normals to epigraphs. In this subsection we consider these questions (and related ones for the case of graphs of continuous functions) in the framework of Asplund spaces.

Let us start with the basic lemma ensuring a strong approximation of horizontal Fréchet normals to epigraphs of l.s.c. functions on Asplund spaces by sequences of Fréchet subgradients.

**Lemma 2.37 (horizontal Fréchet normals to epigraphs).** *Let  $X$  be Asplund, and let  $\varphi: X \rightarrow \overline{\mathbb{R}}$  be a proper function l.s.c. around  $\bar{x} \in \text{dom } \varphi$ . Then for every  $x^* \in X^*$  with  $(x^*, 0) \in \widehat{N}((\bar{x}, \varphi(\bar{x})); \text{epi } \varphi)$  there are sequences  $x_k \xrightarrow{\varphi} \bar{x}$ ,  $\lambda_k \downarrow 0$ , and  $x_k^* \in \lambda_k \widehat{\partial} \varphi(x_k)$  such that  $\|x_k^* - x^*\| \rightarrow 0$  as  $k \rightarrow \infty$ .*

**Proof.** Fix  $x^* \in X^*$  satisfying  $(x^*, 0) \in \widehat{N}((\bar{x}, \varphi(\bar{x})); \text{epi } \varphi)$  and assume without loss of generality that  $\bar{x} = 0$ ,  $\varphi(\bar{x}) = 0$ , and  $\|x^*\| = 1$ . Take an arbitrary  $\varepsilon \in (0, 1)$  and choose  $\eta = \eta(\varepsilon) \downarrow 0$  as  $\varepsilon \downarrow 0$  such that

$$\varphi(x) \geq -\varepsilon \text{ on } \eta B \text{ and}$$

$$\langle x^*, x \rangle < \varepsilon(\|x\| + |\lambda|) \text{ if } x \in (\eta B) \setminus \{0\}, (x, \lambda) \in \text{epi } \varphi, \lambda \text{ near } 0. \quad (2.52)$$

Form the closed convex set

$$\Omega_\varepsilon := \{x \in X \mid \langle x^*, x \rangle \geq \varepsilon\|x\|\}$$

and observe that

$$\varphi(x) \geq 0 \text{ for all } x \in \Omega_\varepsilon \cap \eta B.$$

Indeed, otherwise one has  $(x, 0) \in \text{epi } \varphi$ , and hence (2.52) implies that  $\langle x^*, x \rangle < \varepsilon\|x\|$ , which contradicts the fact of  $x \in \Omega_\varepsilon$ . Next we show that

$$\text{dist}(x; \Omega_{2\varepsilon}) \geq \frac{\varepsilon}{1 + 2\varepsilon} \|x\| \text{ for any } x \in \Omega_\varepsilon. \quad (2.53)$$

Assuming the opposite, we find  $\tilde{x} \in \Omega_{2\varepsilon}$  satisfying

$$\|x - \tilde{x}\| < \frac{\varepsilon}{1 + 2\varepsilon} \|x\|.$$

The latter inequality implies that

$$\begin{aligned}
\langle x^*, \tilde{x} \rangle &= \langle x^*, \tilde{x} - x \rangle + \langle x^*, x \rangle \leq \|x^*\| \cdot \|\tilde{x} - x\| + \langle x^*, x \rangle \\
&< \|\tilde{x} - x\| + \varepsilon \|x\| < \frac{\varepsilon}{1 + 2\varepsilon} \|x\| + \varepsilon \|x\| \\
&\leq 2\varepsilon \left[ \|x\| - \frac{\varepsilon}{1 + 2\varepsilon} \|x\| \right] \leq 2\varepsilon [\|x\| - \|x - \tilde{x}\|] \leq 2\varepsilon \|\tilde{x}\|,
\end{aligned}$$

which contradicts the fact of  $\tilde{x} \in \Omega_{2\varepsilon}$ . Now given an arbitrary number  $k \in \mathbb{N}$ , define the function

$$\psi_{k,\varepsilon}(x) = \varepsilon \varphi(x) + k \operatorname{dist}(x; \Omega_{2\varepsilon}) - \langle x^*, x \rangle + 2\varepsilon \|x\|$$

that is l.s.c. and bounded from below on  $\eta\mathcal{B}$ . Taking  $u_{k,\varepsilon} \in \eta\mathcal{B}$  with

$$\psi_{k,\varepsilon}(u_{k,\varepsilon}) \leq \inf_{x \in \eta\mathcal{B}} \psi_{k,\varepsilon}(x) + \frac{1}{k}$$

and applying the Ekeland variational principle (Theorem 2.26) to the function  $\psi_{k,\varepsilon}$  on the metric space  $\eta\mathcal{B}$ , we find  $\bar{u}_{k,\varepsilon} \in \eta\mathcal{B}$  satisfying

$$\psi_{k,\varepsilon}(\bar{u}_{k,\varepsilon}) \leq \psi_{k,\varepsilon}(x) + \frac{1}{k} \|x - \bar{u}_{k,\varepsilon}\| \quad \text{whenever } x \in \eta\mathcal{B}.$$

Putting  $x = 0$ , we arrive at the useful upper estimate

$$\psi_{k,\varepsilon}(\bar{u}_{k,\varepsilon}) \leq \frac{1}{k} \|\bar{u}_{k,\varepsilon}\|,$$

which means, by the construction of  $\psi_{k,\varepsilon}$ , that

$$\varepsilon \varphi(\bar{u}_{k,\varepsilon}) + k \operatorname{dist}(\bar{u}_{k,\varepsilon}; \Omega_{2\varepsilon}) - \langle x^*, \bar{u}_{k,\varepsilon} \rangle + 2\varepsilon \|\bar{u}_{k,\varepsilon}\| \leq \frac{1}{k} \|\bar{u}_{k,\varepsilon}\|.$$

The latter clearly yields  $\operatorname{dist}(\bar{u}_{k,\varepsilon}; \Omega_{2\varepsilon}) \rightarrow 0$  as  $k \rightarrow \infty$ .

Now we show that one can always find  $k = k(\varepsilon) \in \mathbb{N}$  satisfying  $\bar{u}_{k,\varepsilon} \in \operatorname{int}(\eta\mathcal{B})$  whenever  $\varepsilon > 0$ ; note that  $\eta = \eta(\varepsilon)$  also depends on  $\varepsilon$  but we skip this in notation for simplicity. Assume first that  $\bar{u}_{k,\varepsilon} \in \Omega_\varepsilon$ , i.e.,

$$\langle x^*, \bar{u}_{k,\varepsilon} \rangle \geq \varepsilon \|\bar{u}_{k,\varepsilon}\|.$$

Employing (2.52), we have

$$\varepsilon \varphi(\bar{u}_{k,\varepsilon}) + \varepsilon \|u_{k,\varepsilon}\| - \langle x^*, u_{k,\varepsilon} \rangle \geq 0$$

with  $u_{k,\varepsilon}$  chosen above, and hence

$$\psi_{k,\varepsilon}(\bar{u}_{k,\varepsilon}) \geq \varepsilon \|\bar{u}_{k,\varepsilon}\| + k \operatorname{dist}(\bar{u}_{k,\varepsilon}; \Omega_{2\varepsilon}) \geq \varepsilon \|\bar{u}_{k,\varepsilon}\|.$$

Combining this with the preceding upper estimate for  $\psi(\bar{u}_{k,\varepsilon})$ , one gets

$$\varepsilon \|\bar{u}_{k,\varepsilon}\| \leq \frac{1}{k} \|\bar{u}_{k,\varepsilon}\|, \quad \text{and thus } \bar{u}_{k,\varepsilon} = 0$$

for all  $k \in \mathbb{N}$  sufficiently large. If  $\bar{u}_{k,\varepsilon} \notin \Omega_\varepsilon$ , then (2.53) gives

$$\frac{\varepsilon}{1+2\varepsilon} \|\bar{u}_{k,\varepsilon}\| \leq \text{dist}(\bar{u}_{k,\varepsilon}; \Omega_{2\varepsilon}) \rightarrow 0 ,$$

i.e.,  $\bar{u}_{k,\varepsilon} \rightarrow 0$  as  $k \rightarrow \infty$ . Thus there is a sequence of  $k = k_\varepsilon \rightarrow \infty$  as  $\varepsilon \downarrow 0$  for which  $\|\bar{u}_{k,\varepsilon}\| \leq \eta = \eta(\varepsilon)$ . Taking this into account and the fact that  $\bar{u}_\varepsilon := \bar{u}_{k_\varepsilon,\varepsilon}$  is a *minimizer* to the function  $\psi_{k,\varepsilon} + \frac{1}{k_\varepsilon} \|x - \bar{u}_\varepsilon\|$  on  $\eta B$ , one has

$$0 \in \widehat{\partial}(\varepsilon\varphi + \varphi_\varepsilon)(\bar{u}_\varepsilon)$$

by the generalized Fermat rule, where

$$\varphi_\varepsilon(x) := k_\varepsilon \text{dist}(x; \Omega_{2\varepsilon}) - \langle x^*, x \rangle + 2\varepsilon \|x\| + \frac{1}{k_\varepsilon} \|x - \bar{u}_\varepsilon\| . \quad (2.54)$$

Applying the subgradient description of Lemma 2.32 to the above sum, we find elements  $v_\varepsilon$ ,  $w_\varepsilon$ ,  $v_\varepsilon^*$ , and  $w_\varepsilon^*$  satisfying

$$\|v_\varepsilon - \bar{u}_\varepsilon\| \leq \eta, \quad \|w_\varepsilon - \bar{u}_\varepsilon\| \leq \eta ,$$

$$v_\varepsilon^* \in \widehat{\partial}\varphi(v_\varepsilon), \quad w_\varepsilon^* \in \widehat{\partial}\varphi_\varepsilon(w_\varepsilon) ,$$

$$\|\varepsilon v_\varepsilon^* + w_\varepsilon^*\| \leq \varepsilon \quad \text{for all } \varepsilon > 0 .$$

It follows from the structure of the *convex continuous* function  $\varphi_\varepsilon$  in (2.54), by basic convex analysis, that

$$w_\varepsilon^* \in k_\varepsilon \partial \text{dist}(w_\varepsilon; \Omega_{2\varepsilon}) - x^* + \left(2\varepsilon + \frac{1}{k_\varepsilon}\right) B^* .$$

Hence there is  $\bar{w}_\varepsilon^* \in \partial \text{dist}(w_\varepsilon; \Omega_{2\varepsilon})$  such that

$$\|\varepsilon v_\varepsilon^* + k_\varepsilon \bar{w}_\varepsilon^* - x^*\| \leq 2\varepsilon + \frac{1}{k_\varepsilon} . \quad (2.55)$$

To proceed, we consider the following two cases.

**Case 1.** Let  $w_\varepsilon \in \Omega_{2\varepsilon}$ . Then, as well known from convex analysis,

$$\partial \text{dist}(w_\varepsilon; \Omega_{2\varepsilon}) = N(w_\varepsilon; \Omega_{2\varepsilon}) \cap B^* = \text{cone} \{ -x^* + 2\varepsilon B^* \} \cap B^*$$

due to the structure of the set  $\Omega_{2\varepsilon}$ ; cf. Corollary 1.96. Hence

$$\bar{w}_\varepsilon^* = \alpha_\varepsilon (-x^* + 2\varepsilon e_\varepsilon^*) \quad \text{with} \quad \|\bar{w}_\varepsilon^*\| \leq 1 \quad \text{and} \quad \|e_\varepsilon^*\| \leq 1 ,$$

where  $\alpha_\varepsilon \geq 0$  are uniformly bounded due to  $\|x^*\| = 1$ . By (2.55) one has

$$\|\varepsilon v_\varepsilon^* + k_\varepsilon (\alpha_\varepsilon (-x^* + 2\varepsilon e_\varepsilon^*)) - x^*\| \leq 2\varepsilon + \frac{1}{k_\varepsilon} ,$$

which implies the estimate

$$\|\varepsilon v_\varepsilon^* - (k_\varepsilon \alpha_\varepsilon + 1)x^*\| \leq 2\varepsilon k_\varepsilon \alpha_\varepsilon + 2\varepsilon + \frac{1}{k_\varepsilon}.$$

Let  $\tilde{\lambda}_\varepsilon := k_\varepsilon \alpha_\varepsilon + 1$  and observe that

$$\left\| \frac{\varepsilon}{\tilde{\lambda}_\varepsilon} v_\varepsilon^* - x^* \right\| \leq \frac{1}{k_\varepsilon \alpha_\varepsilon + 1} \left( 2\varepsilon k_\varepsilon \alpha_\varepsilon + 2\varepsilon + \frac{1}{k_\varepsilon} \right) \rightarrow 0 \text{ as } \varepsilon \downarrow 0.$$

Finally putting  $\lambda_\varepsilon := \varepsilon / \tilde{\lambda}_\varepsilon$ , we get

$$\|\lambda_\varepsilon v_\varepsilon^* - x^*\| \rightarrow 0 \text{ with } v_\varepsilon^* \in \widehat{\partial}\varphi(w_\varepsilon) \text{ and } w_\varepsilon \rightarrow 0$$

as  $\varepsilon \downarrow 0$ , which justifies the lemma in Case 1 considered.

**Case 2.** Let  $w_\varepsilon \notin \Omega_{2\varepsilon}$ . First note that Theorem 1.99 implies the inclusion

$$\widehat{\partial}\text{dist}(\bar{x}; \Omega) \subset \bigcap_{\nu > 0} \bigcup \left[ \widehat{N}(x; \Omega) + \nu B^* \mid \|x - \bar{x}\| \leq \text{dist}(\bar{x}; \Omega) + \nu \right]$$

for any set  $\Omega \subset X$  in a Banach space and any out-of-set point  $\bar{x} \notin \Omega$ . Putting  $\bar{x} := w_\varepsilon$  and  $\nu := 1/k_\varepsilon$  therein, we find  $\tilde{w}_\varepsilon \in \Omega_{2\varepsilon}$  and  $\tilde{w}_\varepsilon^* \in \widehat{N}(\tilde{w}_\varepsilon; \Omega_{2\varepsilon}) = N(\tilde{w}_\varepsilon; \Omega_{2\varepsilon})$  such that

$$\|\tilde{w}_\varepsilon^* - \bar{w}_\varepsilon^*\| \leq \frac{1}{k_\varepsilon} \text{ and }$$

$$\|\tilde{w}_\varepsilon - w_\varepsilon\| \leq \text{dist}(w_\varepsilon; \Omega_{2\varepsilon}) + \frac{1}{k_\varepsilon} \leq \|w_\varepsilon\| + \frac{1}{k_\varepsilon} \rightarrow 0$$

as  $\varepsilon \downarrow 0$ . Then we have the representation

$$\tilde{w}_\varepsilon^* = \alpha_\varepsilon(-x^* + 2\varepsilon e_\varepsilon^*) \text{ with } e_\varepsilon^* \in B^*,$$

where  $\alpha_\varepsilon$  are uniformly bounded. Thus

$$\begin{aligned} \|\varepsilon v_\varepsilon^* + k_\varepsilon \bar{w}_\varepsilon^* - x^*\| &\leq 2\varepsilon + \frac{1}{k_\varepsilon} \\ \implies \|\varepsilon v_\varepsilon^* + k_\varepsilon \tilde{w}_\varepsilon^* - x^*\| &\leq \frac{1}{k_\varepsilon} + 2\varepsilon + \frac{1}{k_\varepsilon} \leq \frac{2}{k_\varepsilon} + 2\varepsilon \\ \implies \|\varepsilon v_\varepsilon^* + k_\varepsilon(-\alpha_\varepsilon)(-x^* + 2\varepsilon e_\varepsilon^*) - x^*\| &\leq \frac{2}{k_\varepsilon} + 2\varepsilon \\ \implies \|\varepsilon v_\varepsilon^* - (k_\varepsilon \alpha_\varepsilon + 1)x^*\| &\leq 2k_\varepsilon \alpha_\varepsilon \varepsilon + \frac{2}{k_\varepsilon} + 2\varepsilon \\ \implies \left\| \frac{\varepsilon}{k_\varepsilon \alpha_\varepsilon + 1} v_\varepsilon^* - x^* \right\| &\leq \frac{2}{k_\varepsilon \alpha_\varepsilon + 1} \left[ k_\varepsilon \alpha_\varepsilon \varepsilon + \frac{1}{k_\varepsilon} + \varepsilon \right] \rightarrow 0 \text{ as } \varepsilon \downarrow 0. \end{aligned}$$



Finally, letting

$$\lambda_\varepsilon := \frac{\varepsilon}{k_\varepsilon \alpha_\varepsilon + 1}$$

as in Case 1, we justify the required relationships in Case 2 and thus complete the proof of the lemma.  $\triangle$

**Theorem 2.38 (singular subgradients in Asplund spaces).** *Let  $X$  be an Asplund space. Assume that  $\varphi: X \rightarrow \overline{\mathbb{R}}$  is a proper function l.s.c. around some point  $\bar{x} \in \text{dom } \varphi$ . Then the singular subdifferential of  $\varphi$  admits the following limiting representations:*

$$\partial^\infty \varphi(\bar{x}) = \limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ \lambda \downarrow 0}} \lambda \widehat{\partial} \varphi(x) = \limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ \varepsilon, \lambda \downarrow 0}} \lambda \widehat{\partial}_\varepsilon \varphi(x).$$

**Proof.** The equality

$$\limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ \lambda \downarrow 0}} \lambda \widehat{\partial} \varphi(x) = \limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ \varepsilon, \lambda \downarrow 0}} \lambda \widehat{\partial}_\varepsilon \varphi(x)$$

for any l.s.c. function on Asplund spaces follows from formula (2.49) justified above. It remains to prove the inclusion

$$\partial^\infty \varphi(\bar{x}) \subset \limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ \lambda \downarrow 0}} \lambda \widehat{\partial} \varphi(x),$$

since the opposite one is easily implied by the definitions. To proceed, we take an arbitrary  $x^* \in \partial^\infty \varphi(\bar{x})$  for which  $(x^*, 0) \in N((\bar{x}, \varphi(\bar{x})); \text{epi } \varphi)$  by Definition 1.77(ii). Employing Theorem 2.35, we find sequences  $(x_k, \alpha_k) \rightarrow (\bar{x}, \varphi(\bar{x}))$  and  $(x_k^*, \nu_k) \xrightarrow{w^*} (x^*, 0)$  such that  $\alpha_k \geq \varphi(x_k)$  and  $(x_k^*, -\nu_k) \in \widehat{N}((x_k, \alpha_k); \text{epi } \varphi)$ ,  $k \in \mathbb{N}$ . The latter implies that  $\nu_k \geq 0$  for all  $k$ . Thus one has two possibilities for the sequence  $\{(x_k^*, \nu_k)\}$ : either

- (a) there is a subsequence of  $\{\nu_k\}$  consisting of positive numbers, or
- (b)  $\nu_k = 0$  for all  $k$  sufficiently large.

In case (a) we assume without loss of generality that  $\nu_k > 0$  for all  $k \in \mathbb{N}$ , which implies that  $\alpha_k = \varphi(x_k)$  and  $x_k^*/\nu_k \in \widehat{\partial} \varphi(x_k)$ ,  $k \in \mathbb{N}$ . Letting  $\lambda_k := \nu_k$  and  $\tilde{x}_k^* := x_k^*/\nu_k$ , we get  $\lambda_k \tilde{x}_k^* \xrightarrow{w^*} x^*$  and  $\lambda_k \downarrow 0$  as  $k \rightarrow \infty$ .

In case (b) one has  $(x_k^*, 0) \in \widehat{N}((x_k, \varphi(x_k)); \text{epi } \varphi)$  if  $x_k^* \neq 0$ , which we may always assume. Now employing Lemma 2.37 and the standard diagonal process, we get sequences  $\tilde{x}_k \xrightarrow{\varphi} \bar{x}$ ,  $\lambda_k \downarrow 0$ , and  $\tilde{x}_k^* \xrightarrow{w^*} x^*$  such that  $\tilde{x}_k^* \in \lambda_k \widehat{\partial} \varphi(\tilde{x}_k)$  for large  $k$ . This completes the proof.  $\triangle$

Note that analytic  $\varepsilon$ -subgradients in the second representation of Theorem 2.38 can be replaced with  $\varepsilon$ -geometric subgradients due to Theorem 1.86.

We'll see further in the book many applications of both Lemma 2.37 and Theorem 2.38 to various aspects of analysis and optimization in Asplund spaces. Right now let us present a consequence of Lemma 2.37 providing a convenient subdifferential description of the SNEC property for extended-real-valued functions on Asplund spaces; cf. Definition 1.116.

**Corollary 2.39 (subdifferential description of sequential normal epi-compactness).** *Let  $X$  be Asplund, and let  $\varphi: X \rightarrow \overline{\mathbb{R}}$  be a proper function l.s.c. around  $\bar{x} \in \text{dom } \varphi$ . Then  $\varphi$  is SNEC at  $\bar{x}$  if and only if for any sequences  $x_k \xrightarrow{\varphi} \bar{x}$ ,  $\lambda_k \downarrow 0$ , and  $x_k^* \in \lambda_k \widehat{\partial} \varphi(x_k)$  one has*

$$[x_k^* \xrightarrow{w^*} 0] \implies \|x_k^*\| \rightarrow 0 \text{ as } k \rightarrow \infty.$$

**Proof.** Assume that  $\varphi$  is SNEC at  $\bar{x}$ . Take any sequences  $x_k \xrightarrow{\varphi} \bar{x}$ ,  $\lambda_k \downarrow 0$ , and  $x_k^* \in \lambda_k \widehat{\partial} \varphi(x_k)$  with  $x_k^* \xrightarrow{w^*} 0$  as  $k \rightarrow \infty$ . Then

$$(x_k^*, -\lambda_k) \in \widehat{N}((x_k, \varphi(x_k)); \text{epi } \varphi) \text{ for all } k \in \mathbb{N},$$

and the SNEC property of  $\varphi$  at  $\bar{x}$  implies that  $\|x_k^*\| \rightarrow 0$  as  $k \rightarrow \infty$ .

To prove the converse application, pick arbitrary sequences

$$(x_k, \alpha_k) \in \text{epi } \varphi \text{ and } (x_k^*, -\lambda_k) \in \widehat{N}((x_k, \alpha_k); \text{epi } \varphi)$$

with  $(x_k, \alpha_k) \rightarrow (\bar{x}, \varphi(\bar{x}))$ ,  $\lambda_k \rightarrow 0$ , and  $x_k^* \xrightarrow{w^*} 0$ . We need to show  $\|x_k^*\| \rightarrow 0$  as  $k \rightarrow \infty$ ; in fact it is sufficient to justify the latter holds along a subsequence.

Since  $\lambda_k \geq 0$  for all  $k \in \mathbb{N}$ , there are the following two cases to consider:

- (a)  $\lambda_k > 0$  along a subsequence of  $k \in \mathbb{N}$ ;
- (b)  $\lambda_k = 0$  for all large  $k \in \mathbb{N}$ .

Case (a) is simple. Indeed, we easily have  $\alpha_k = \varphi(x_k)$ , and hence

$$\left(\frac{x_k^*}{\lambda_k}, -1\right) \in \widehat{N}((x_k, \varphi(x_k)); \text{epi } \varphi), \text{ i.e., } x_k^* \in \lambda_k \widehat{\partial} \varphi(x_k).$$

Then  $\|x_k^*\| \rightarrow 0$  by the assumption made, which yields that  $\varphi$  is SNEC at  $\bar{x}$ .

Case (b) is more involved requiring the usage of Lemma 2.37. To proceed, we suppose without lost of generality that  $\lambda_k = 0$  and  $\alpha_k = \varphi(x_k)$  for all  $k \in \mathbb{N}$ . Thus  $(x_k^*, 0) \in \widehat{N}((x_k, \varphi(x_k)); \text{epi } \varphi)$ . Applying Lemma 2.37 for each  $k$ , we select subsequences  $\lambda_{n_k}$ ,  $\tilde{x}_{n_k}$ , and  $\tilde{x}_{n_k}^*$  so that

$$0 < \lambda_{n_k} < \frac{1}{k}, \quad \|\tilde{x}_{n_k} - x_k\| \leq \frac{1}{k}, \quad |\varphi(\tilde{x}_{n_k}) - \varphi(x_k)| \leq \frac{1}{k},$$

$$\|\tilde{x}_{n_k}^* - x_k^*\| \leq \frac{1}{k}, \text{ and } \tilde{x}_{n_k}^* \in \lambda_{n_k} \widehat{\partial} \varphi(\tilde{x}_{n_k}).$$

One clearly has  $\tilde{x}_{n_k}^* \xrightarrow{w^*} 0$  due to the construction of  $\tilde{x}_{n_k}^*$  and the assumption on  $x_k^* \xrightarrow{w^*} 0$ . Then  $\|\tilde{x}_{n_k}^*\| \rightarrow 0$  and hence  $\|x_{n_k}^*\| \rightarrow 0$ , which implies the SNEC property and completes the proof of the corollary.  $\triangle$

The concluding result of this section gives an efficient representation of horizontal Fréchet normals to *graphs* of continuous functions in Asplund spaces and provides a refinement of coderivative-subdifferential relations considered in Theorem 1.80.

**Theorem 2.40 (horizontal normals to graphs of continuous functions).** *Let  $X$  be an Asplund space, and let  $\varphi: X \rightarrow \overline{\mathbb{R}}$  be finite and continuous around some point  $x \in X$ . The following hold:*

(i) *If  $(x^*, 0) \in \widehat{N}((\bar{x}, \varphi(\bar{x})); \text{gph } \varphi)$ , then there exist sequences  $x_k \rightarrow \bar{x}$ ,  $\lambda_k \downarrow 0$ , and  $x_k^* \rightarrow x^*$  such that*

$$x_k^* \in \widehat{\partial}(\lambda_k \varphi)(x_k) \cup \widehat{\partial}(-\lambda_k \varphi)(x_k) \text{ for all } k \in \mathbb{N}.$$

(ii)  $D^*\varphi(\bar{x})(0) = \partial^\infty \varphi(\bar{x}) \cup \partial^\infty(-\varphi)(\bar{x})$ .

**Proof.** To justify (i), we proceed similarly to the proof of Lemma 2.37 with a certain modification in constructions and estimates due to the continuity of  $\varphi$ , which makes it possible to derive two-sided formulas. For brevity we skip some details using slightly different notation.

Assume that  $\bar{x} = 0$ ,  $\varphi(\bar{x}) = 0$  and pick an arbitrary  $x^* \in B^* \subset X^*$  with  $(x^*, 0) \in \widehat{N}((0, 0); \text{gph } \varphi)$ . For each  $\varepsilon > 0$  we find  $\eta = \eta(\varepsilon) \downarrow 0$  as  $\varepsilon \downarrow 0$  such that  $\varphi$  is bounded on  $\eta B$  and

$$\langle x^*, x \rangle < \varepsilon(\|x\| + |\varphi(x)|) \text{ for all } x \in \eta B \setminus \{0\}. \quad (2.56)$$

Form the set  $\Omega_\varepsilon$  as in the proof of Lemma 2.37 and observe that either

(a)  $\varphi(x) \geq 0$  for all  $x \in \Omega_\varepsilon \cap (\eta B)$ , or

(b)  $\varphi(x) \leq 0$  for all  $x \in \Omega_\varepsilon \cap (\eta B)$ .

Indeed, if there are  $x_1, x_2 \in \Omega_\varepsilon \cap (\eta B)$  with  $\varphi(x_1) > 0$  and  $\varphi(x_2) < 0$ , then both  $x_1$  and  $x_2$  are nonzero and, by the continuity of  $\varphi$ , there is  $x := \alpha x_1 + (1 - \alpha)x_2 \in \Omega_\varepsilon \cap (\eta B) \setminus \{0\}$  with  $\alpha \in (0, 1)$  and  $\varphi(x) = 0$ . This clearly contradicts (2.56).

For each  $k \in \mathbb{N}$  define the function

$$\psi_{k,\varepsilon}(x) := \begin{cases} \varepsilon \varphi(x) + k \text{dist}(x; \Omega_{2\varepsilon}) - \langle x^*, x \rangle + 2\varepsilon \|x\| & \text{if (a) holds,} \\ -\varepsilon \varphi(x) + k \text{dist}(x; \Omega_{2\varepsilon}) - \langle x^*, x \rangle + 2\varepsilon \|x\| & \text{if (b) holds} \end{cases}$$

and apply the Ekeland variational principle to this function on the metric space  $\eta B$ . In this way we find  $x_{k,\varepsilon} \in \eta B$  that minimizes the function  $\psi_{k,\varepsilon}(x) + \frac{1}{k}\|x - x_{k,\varepsilon}\|$  on  $\eta B$ . In particular,

$$\psi_{k,\varepsilon}(x_{k,\varepsilon}) \leq \psi_{k,\varepsilon}(0) = \frac{1}{k} \|x_{k,\varepsilon}\| \quad \text{and} \quad \text{dist}(x_{k,\varepsilon}; \Omega_{2\varepsilon}) \rightarrow 0 \quad (2.57)$$

as  $k \rightarrow \infty$ . Let us further choose  $k_\varepsilon \rightarrow \infty$  as  $\varepsilon \downarrow 0$  similarly to the proof of Lemma 2.37. If  $x_{k,\varepsilon} \in \Omega_\varepsilon$ , then it follows from (2.56) and (2.57) that  $x_{k,\varepsilon} = 0$  for  $k > 1/\varepsilon$ . If  $x_{k,\varepsilon} \notin \Omega_\varepsilon$ , then  $\|x_{k,\varepsilon}\| \rightarrow 0$  as  $k \rightarrow \infty$  by (2.55) and (2.57). Thus for every  $\varepsilon > 0$  there are  $k = k_\varepsilon$  and  $x_\varepsilon := x_{k_\varepsilon, \varepsilon}$  such that  $k_\varepsilon \rightarrow \infty$  as  $\varepsilon \downarrow 0$ , that  $\|x_\varepsilon\| < \eta/2$ , and that

$$0 \in \widehat{\partial} \left( \psi_\varepsilon + \frac{1}{k} \|\cdot - x_\varepsilon\| \right) (x_\varepsilon),$$

where  $\psi_\varepsilon(x) := \psi_{k_\varepsilon, \varepsilon}(x)$ . Applying Lemma 2.32 and taking into account the structure of  $\psi_\varepsilon$ , we find  $u_\varepsilon \in \eta B$ ,  $v_\varepsilon \in \eta B$ ,  $u_\varepsilon^* \in \widehat{\partial}\varphi(u_\varepsilon) \cup \widehat{\partial}(-\varphi)(u_\varepsilon)$ , and  $v_\varepsilon^* \in \partial \text{dist}(v_\varepsilon; \Omega_{2\varepsilon})$  with

$$\|v_\varepsilon^*\| \leq 1 \quad \text{and} \quad \|\varepsilon u_\varepsilon^* + k v_\varepsilon^* - x^*\| \leq 2(\varepsilon + 1/k). \quad (2.58)$$

Consider again the two possible cases:  $v_\varepsilon \in \Omega_{2\varepsilon}$  and  $v_\varepsilon \notin \Omega_{2\varepsilon}$ . In the first case we employ the representation of  $\partial \text{dist}(v_\varepsilon; \Omega_{2\varepsilon})$  from convex analysis and get  $\alpha_\varepsilon > 0$  and  $e^* \in B^*$  such that  $v_\varepsilon^* + \alpha_\varepsilon x^* = 2\varepsilon \alpha_\varepsilon e^*$ . This implies that the sequence  $\{\alpha_\varepsilon\}$  is bounded as  $\varepsilon \downarrow 0$ . From (2.58) one has the estimates

$$\begin{aligned} \|\varepsilon u_\varepsilon^* - (k\alpha_\varepsilon + 1)x^*\| &\leq \|\varepsilon u_\varepsilon^* + k v_\varepsilon^* - x^*\| + k \|v_\varepsilon^* + \alpha_\varepsilon x^*\| \\ &\leq 2(\varepsilon + 1/k) + 2k\alpha_\varepsilon \varepsilon. \end{aligned}$$

Dividing this by  $k\alpha_\varepsilon + 1$  and denoting  $\lambda_\varepsilon := \varepsilon/(k\alpha_\varepsilon + 1)$ ,  $x_\varepsilon^* := \lambda_\varepsilon u_\varepsilon^*$ , we obtain  $x_\varepsilon^* \in \widehat{\partial}(\lambda_\varepsilon \varphi)(u_\varepsilon) \cup \widehat{\partial}(-\lambda_\varepsilon \varphi)(u_\varepsilon)$  with  $\|x_\varepsilon^* - x^*\| \rightarrow 0$  and  $\lambda_\varepsilon \downarrow 0$  as  $\varepsilon \downarrow 0$ . In the case of  $v_\varepsilon \notin \Omega_{2\varepsilon}$  we proceed similarly to the proof of Lemma 2.37 based on the upper estimate of  $\widehat{\text{dist}}(\bar{x}; \Omega)$  with  $\bar{x} \notin \Omega$  from Theorem 1.99. This completes the proof of assertion (i) in the theorem.

To justify the inclusion “ $\subset$ ” in (ii), we argue as in the proof of Theorem 2.38. The opposite inclusion follows from Theorem 1.80.  $\triangle$

## 2.5 Versions of Extremal Principle in Banach Spaces

We have shown in the previous section that the above versions of the extremal principle and most of the related results are not only valid in Asplund spaces but happen to provide *characterizations* for this general class of Banach spaces. To cover other classes of Banach spaces, one therefore needs to employ different constructions of generalized normals involving in formulations of the extremal principle. In this section we detect those properties of axiomatically defined normal and subgradient structures that allow us to derive approximate and exact versions of the abstract extremal principle valid in appropriate classes of Banach spaces.

### 2.5.1 Axiomatic Normal and Subdifferential Structures

First we define an abstract *prenormal* structure on a Banach space that supports an *approximate* version of the extremal principle.

**Definition 2.41 (prenormal structures).** *Let  $X$  be a Banach space. We say that  $\widehat{N}$  defines a PRENORMAL STRUCTURE on  $X$  if it associates, with every nonempty set  $\Omega \subset X$ , a set-valued mapping  $\widehat{N}(\cdot; \Omega): X \rightrightarrows X^*$  such that  $\widehat{N}(x; \Omega) = \emptyset$  for  $x \notin \Omega$ ,  $\widehat{N}(x; \Omega) = \widehat{N}(x; \widetilde{\Omega})$  when  $\Omega$  and  $\widetilde{\Omega}$  are the same near  $x \in \Omega$ , and the following property holds:*

**(H)** *Given any small  $\varepsilon > 0$ ,  $a \in X$  with  $\|a\| \leq \varepsilon$ , and closed sets  $\Omega_1, \Omega_2 \subset X$ , assume that  $(\bar{x}_1, \bar{x}_2) \in \Omega_1 \times \Omega_2$  is a local minimizer for the function*

$$\psi(x_1, x_2) := \|x_1 - x_2 + a\| + \varepsilon(\|x_1 - \bar{x}_1\| + \|x_2 - \bar{x}_2\|) \quad (2.59)$$

*relative to the set  $\Omega_1 \times \Omega_2$  with  $\bar{x}_1 - \bar{x}_2 + a \neq 0$ . Then there are  $\tilde{x}_i \in \bar{x}_i + \varepsilon \mathcal{B}$ ,  $i = 1, 2$ , and  $x^* \in X^*$  with  $\|x^*\| = 1$  such that*

$$(-x^*, x^*) \in \widehat{N}(\tilde{x}_1; \Omega_1) \times \widehat{N}(\tilde{x}_2; \Omega_2) + \gamma(\mathcal{B}^* \times \mathcal{B}^*) \text{ for all } \gamma > \varepsilon. \quad (2.60)$$

We can easily check by the results above that property (H) holds for the prenormal (Fréchet normal) cone  $\widehat{N}$  in Asplund spaces; cf. the proof of Lemma 2.32(ii). In general this property postulates the ability of the prenormal structure  $\widehat{N}$  to describe *first-order necessary optimality conditions* for minimizing functions of the norm type (2.59) over arbitrary sets. Note that (2.60) provides a “fuzzy” optimality condition, since it involves points  $(\tilde{x}_1, \tilde{x}_2)$  close to the given minimizer with  $\gamma > \varepsilon$  in (2.60).

Let us show that property (H) always holds for subdifferentially generated *prenormal cones* under a minimal amount of natural requirements in the corresponding Banach spaces. Given a Banach space  $X$ , we say that  $\widehat{\mathcal{D}}$  defines an (abstract) *presubdifferential* on  $X \times X$  if it associates, with every proper function  $\varphi: X \times X \rightarrow \overline{\mathbb{R}}$ , a set-valued mapping  $\widehat{\mathcal{D}}\varphi: X \times X \rightrightarrows X^* \times X^*$  such that  $\widehat{\mathcal{D}}\varphi(z) = \emptyset$  for  $z \notin \text{dom } \varphi$ ,  $\widehat{\mathcal{D}}\varphi(z) = \widehat{\mathcal{D}}\phi(z)$  if  $\varphi$  and  $\phi$  coincide around  $z$ , and one has the following:

**(S1)** Suppose that  $\bar{z}$  provides a local minimum for the sum  $\varphi_1 + \varphi_2$  of two functions finite at  $\bar{z}$ , where  $\varphi_1$  is a convex continuous function of type (2.59) and where  $\varphi_2$  is a l.s.c. function of the set indicator type. Then for any  $\eta > 0$  there are  $u, v \in \bar{z} + \eta \mathcal{B}$  such that  $\varphi_2(v) \leq \varphi_2(\bar{z}) + \eta$  and

$$0 \in \widehat{\mathcal{D}}\varphi_1(u) + \widehat{\mathcal{D}}\varphi_2(v) + \eta(\mathcal{B}^* \times \mathcal{B}^*).$$

**(S2)**  $\widehat{\mathcal{D}}\varphi(z)$  is contained in the subdifferential of convex analysis for convex continuous function of type (2.59).

**(S3)** If  $\varphi(x_1, x_2) = \varphi_1(x_1) + \varphi_2(x_2)$ , then  $\widehat{\mathcal{D}}\varphi(\bar{x}_1, \bar{x}_2) \subset \widehat{\mathcal{D}}\varphi_1(\bar{x}_1) \times \widehat{\mathcal{D}}\varphi_2(\bar{x}_2)$  for any  $\bar{x}_i \in \text{dom } \varphi_i$ ,  $i = 1, 2$ .

**Proposition 2.42 (prenormal cones from presubdifferentials).** *Given a Banach space  $X$ , let  $\widehat{\mathcal{D}}$  be an arbitrary presubdifferential on  $X \times X$ . Then  $\widehat{\mathcal{N}}(x; \Omega) := \widehat{\mathcal{D}}\delta(x; \Omega)$  is a cone for any closed set  $\Omega \subset X \times X$  and any  $x \in \Omega$ , and  $\widehat{\mathcal{N}}$  defines a prenormal structure on  $X$ .*

**Proof.** The set  $\widehat{\mathcal{N}}(x; \Omega)$  is a cone, since  $\alpha\delta(x; \Omega) = \delta(x; \Omega)$  for every  $\alpha > 0$ . Obviously  $\widehat{\mathcal{N}}(x; \Omega) = \emptyset$  if  $x \notin \Omega$ . We need to show that  $\widehat{\mathcal{N}}$  satisfies property (H) in Definition 2.41. To proceed, take  $\bar{z} = (\bar{x}_1, \bar{x}_2) \in \Omega_1 \times \Omega_2$  that provides a local minimum for  $\psi$  in (2.59) relative to  $\Omega_1 \times \Omega_2$  with given  $\varepsilon > 0$  and  $\bar{x}_1 - \bar{x}_2 + a \neq 0$ . Observe that  $\bar{z}$  is a local minimizer for the function

$$\varphi(x_1, x_2) := \psi(x_1, x_2) + \delta((x_1, x_2); \Omega_1 \times \Omega_2), \quad (x_1, x_2) \in X \times X,$$

with no additional constraints. Pick any  $\gamma > \varepsilon$  and put

$$\eta := \gamma - \varepsilon \quad \text{with} \quad \eta \leq \min \{\varepsilon, \nu/2\}, \quad \nu := \|\bar{x}_1 - \bar{x}_2 + a\|. \quad (2.61)$$

Applying (S1) with  $\varphi_1 = \psi$  and  $\varphi_2 = \delta(\cdot; \Omega_1 \times \Omega_2)$  and using the construction of  $\widehat{\mathcal{N}}$ , we find  $u = (x'_1, x'_2) \in X^2$  and  $v = (\tilde{x}_1, \tilde{x}_2) \in \Omega_1 \times \Omega_2$  such that

$$\max \{\|x'_1 - \bar{x}_1\|, \|x'_2 - \bar{x}_2\|, \|\tilde{x}_1 - \bar{x}_1\|, \|\tilde{x}_2 - \bar{x}_2\|\} \leq \eta \leq \varepsilon, \quad (2.62)$$

$$0 \in \widehat{\mathcal{D}}\psi(x'_1, x'_2) + \widehat{\mathcal{N}}((\tilde{x}_1, \tilde{x}_2); \Omega_1 \times \Omega_2) + \eta(\mathcal{B}^* \times \mathcal{B}^*).$$

Due to (2.61) and (2.62) we get

$$\|x'_1 - x'_2\| \geq \|\bar{x}_1 - \bar{x}_2 + a\| - (\|x'_1 - \bar{x}_1\| + \|x'_2 - \bar{x}_2\|) = \nu - 2\eta > 0.$$

Observe also that (S3) yields

$$\widehat{\mathcal{N}}((\bar{x}_1, \bar{x}_2); \Omega_1 \times \Omega_2) \subset \widehat{\mathcal{N}}(\bar{x}_1; \Omega_1) \times \widehat{\mathcal{N}}(\bar{x}_2; \Omega_2).$$

By (S2) and the subdifferential formulas of convex analysis for function (2.59) one has the inclusion

$$\widehat{\mathcal{D}}\psi(x'_1, x'_2) \subset (x^*, -x^*) + \varepsilon(\mathcal{B}^* \times \mathcal{B}^*) \quad \text{with} \quad \|x^*\| = 1. \quad (2.63)$$

Putting the above together and taking into account that  $\gamma = \varepsilon + \eta$ , we arrive at (2.60) and finish the proof.  $\triangle$

The result obtained describes an important class of prenormal structures given by subdifferentially generated conic sets. Observe that condition (2.60) with  $\|x^*\| = 1$  doesn't necessarily require that  $\widehat{\mathcal{N}}(x; \Omega)$  are cones or even unbounded sets. Note also that a prenormal structure  $\widehat{\mathcal{N}}$  doesn't need to be subdifferentially generated.

Let us describe another class of prenormal structures on  $X$  involving bounded sets  $\widehat{\mathcal{N}}(x; \Omega)$  associated with presubdifferentials of distance functions

under minimal requirements. Fix an arbitrary number  $\ell > 0$  and consider the class of Lipschitz continuous functions  $\varphi: X \times X \rightarrow \mathbb{R}$  with modulus  $\ell$ . We say that  $\widehat{\mathcal{D}}\varphi(\cdot)$  defines an  $\ell$ -presubdifferential on this class of functions if it satisfies the above presubdifferential assumptions, where (S1) and (S3) are required to hold, respectively, for functions  $\varphi_2$  and  $\varphi_i$ ,  $i = 1, 2$ , of this class. Then we define  $\widehat{\mathcal{N}}$  on  $X$  by

$$\widehat{\mathcal{N}}(x; \Omega) := \begin{cases} \widehat{\mathcal{D}}(\ell \operatorname{dist}(x; \Omega)) & \text{if } x \in \Omega, \\ \emptyset & \text{otherwise} \end{cases} \quad (2.64)$$

for every closed set  $\Omega \subset X$ , where  $\widehat{\mathcal{D}}(\ell \operatorname{dist}(x; \Omega)) := \widehat{\mathcal{D}}(\ell \operatorname{dist}(\cdot; \Omega))(x)$ .

**Proposition 2.43 (prenormal structures from  $\ell$ -presubdifferentials).**

Let  $\widehat{\mathcal{D}}$  be an  $\ell$ -presubdifferential with some  $\ell > 1$ . Then (2.64) defines a prenormal structure on a Banach space  $X$ .

**Proof.** Let us prove that property (H) holds for (2.64) if  $\varepsilon > 0$  is sufficiently small. Fix  $\ell > 1$  and take  $0 < \varepsilon \leq (\ell - 1)/2$ . Since  $(\bar{x}_1, \bar{x}_2)$  is a local minimizer of the function  $\psi$  in (2.59) over the set  $\Omega_1 \times \Omega_2$ , we find neighborhoods  $U_1$  of  $\bar{x}_1$  and  $U_2$  of  $\bar{x}_2$  such that  $\psi$  attains its global minimum over  $(\Omega_1 \cap U_1) \times (\Omega_2 \cap U_2)$  at  $(\bar{x}_1, \bar{x}_2)$ . One can easily see that  $\psi$  is Lipschitz continuous on  $X^2$  with modulus  $1 + 2\varepsilon \leq \ell$ . It is well known that the function

$$\varphi(x_1, x_2) := \psi(x_1, x_2) + \ell \operatorname{dist}((x_1, x_2); (\Omega_1 \cap U_1) \times (\Omega_2 \cap U_2)) \quad (2.65)$$

attains its minimum over the whole space  $X^2$  at  $(\bar{x}_1, \bar{x}_2)$ ; see Proposition 2.4.3 from Clarke [255]. Observe that

$$\operatorname{dist}((x_1, x_2); (\Omega_1 \cap U_1) \times (\Omega_2 \cap U_2)) = \operatorname{dist}(x_1; \Omega_1 \cap U_1) + \operatorname{dist}(x_2; \Omega_2 \cap U_2)$$

due to  $\|(x_1, x_2)\| = \|x_1\| + \|x_2\|$ . Similarly to the proof of Proposition 2.42 we pick  $\gamma > 0$  and take positive numbers  $\eta$  and  $\nu$  satisfying (2.61). By the above property (S1) for the  $\ell$ -presubdifferential  $\widehat{\mathcal{D}}$  of the sum in (2.65) we find points  $u = (x'_1, x'_2) \in X^2$  and  $v = (\tilde{x}_1, \tilde{x}_2) \in X^2$  satisfying (2.62) so that

$$0 \in \widehat{\mathcal{D}}\psi(x'_1, x'_2) + \widehat{\mathcal{D}}(\ell \operatorname{dist}(\tilde{x}_1; \Omega_1 \cap U_1) + \ell \operatorname{dist}(\tilde{x}_2; \Omega_2 \cap U_2)) + \eta(\mathbb{B}^* \times \mathbb{B}^*).$$

If  $\varepsilon$  is sufficiently small, one has

$$\operatorname{dist}(x; \Omega_i \cap U_i) = \operatorname{dist}(x; \Omega_i), \quad i = 1, 2,$$

for all  $x$  in some neighborhoods of  $\tilde{x}_1$  and  $\tilde{x}_2$ , respectively. Thus

$$0 \in \widehat{\mathcal{D}}\psi(x'_1, x'_2) + \widehat{\mathcal{N}}(\tilde{x}_1; \Omega_1) \times \widehat{\mathcal{N}}(\tilde{x}_2; \Omega_2) + (\gamma - \varepsilon)(\mathbb{B}^* \times \mathbb{B}^*)$$

by (2.64) and (S3). Using (S2) and (2.63), we arrive at (2.60).  $\triangle$

As we mentioned above, the basic property (H) of prenormal structures reflects the ability of  $\widehat{\mathcal{N}}$  to describe “fuzzy” necessary optimality conditions in constrained optimization. To get “exact” conditions corresponding to  $\tilde{x}_i = \bar{x}_i$ ,  $i = 1, 2$ , and  $\gamma = \varepsilon$  in (2.60), one needs to employ more *robust* normal constructions. The latter can be obtained by using *limiting procedures* based on prenormals. Let us consider two kinds of such limiting constructions involving the sequential Painlevé-Kuratowski upper limit described in (1.1) and its topological closure.

**Definition 2.44 (sequential and topological normal structures).** *Let  $\widehat{\mathcal{N}}$  be an arbitrary prenormal structure on a Banach space  $X$ . We say that  $\mathcal{N}$  defines a SEQUENTIAL NORMAL STRUCTURE on  $X$  generated by  $\widehat{\mathcal{N}}$  if*

$$\mathcal{N}(\bar{x}; \Omega) = \limsup_{x \rightarrow \bar{x}} \widehat{\mathcal{N}}(x; \Omega) \quad (2.66)$$

for any nonempty set  $\Omega \subset X$  and any  $\bar{x} \in X$ . If (2.66) is replaced with

$$\overline{\mathcal{N}}(\bar{x}; \Omega) = \text{cl}^* \left\{ \limsup_{x \rightarrow \bar{x}} \widehat{\mathcal{N}}(x; \Omega) \right\}, \quad (2.67)$$

then  $\overline{\mathcal{N}}$  defines the corresponding TOPOLOGICAL NORMAL STRUCTURE on  $X$ .

It immediately follows from the definitions that  $\mathcal{N}(\bar{x}; \Omega) = \overline{\mathcal{N}}(\bar{x}; \Omega) = \emptyset$  for  $\bar{x} \notin \Omega$  and, moreover, one may consider only  $x \in \Omega$  in (2.66) and (2.67). Obviously  $\mathcal{N}(\bar{x}; \Omega) \subset \overline{\mathcal{N}}(\bar{x}; \Omega)$ . However, sequential normal structures are mostly useful in Banach spaces  $X$  whose unit dual balls  $\mathcal{B}^* \subset X^*$  are weak\* sequentially compact, while topological normal structures don’t need such an assumption; see, e.g., Subsect. 2.5.3.

Similarly we can define sequential and topological *subdifferential* constructions generated by presubdifferentials. It follows from Proposition 1.31 that our basic normal cone (1.3) is smaller than any other sequential (and hence topological) normal structure in Banach spaces under natural requirements. The next proposition gives a counterpart of this minimality result for the basic subdifferential in Definition 1.77(i).

**Proposition 2.45 (minimality of the basic subdifferential).** *Let  $X$  be a Banach space, and let  $\widehat{\mathcal{D}}\varphi: X \rightrightarrows X^*$  satisfy the following properties on the class of proper l.s.c. functions  $\varphi: X \rightarrow \overline{\mathbb{R}}$ :*

(M1)  $\widehat{\mathcal{D}}\varphi(u) = \widehat{\mathcal{D}}\varphi(x + u)$  for  $\phi(u) := \varphi(x + u)$  and  $x, u \in X$ .

(M2)  $\widehat{\mathcal{D}}\varphi(x)$  is contained in the subdifferential of convex analysis for convex continuous functions in the form

$$\varphi(x) := \langle x^*, x \rangle + \varepsilon \|x\|, \quad x^* \in X^*, \varepsilon > 0. \quad (2.68)$$

(M3) For any  $\eta > 0$  and any functions  $\varphi_i$ ,  $i = 1, 2$ , such that  $\varphi_1$  is convex of type (2.68) and the sum  $\varphi_1 + \varphi_2$  attains a local minimum at  $x = 0$  there are  $x_1, x_2 \in \eta \mathcal{B}$  with  $|\varphi_2(x_2) - \varphi_2(0)| \leq \eta$  and



$$0 \in \widehat{\mathcal{D}}\varphi_1(x_1) + \widehat{\mathcal{D}}\varphi_2(x_2) + \eta \mathcal{B}^* .$$

Then for every  $\bar{x} \in \text{dom } \varphi$  one has the inclusion

$$\partial\varphi(\bar{x}) \subset \limsup_{x \xrightarrow{\varphi} \bar{x}} \widehat{\mathcal{D}}\varphi(x) .$$

**Proof.** Take  $x^* \in \partial\varphi(\bar{x})$  and by Theorem 1.89 find  $\varepsilon_k \downarrow 0$ ,  $x_k \xrightarrow{\varphi} \bar{x}$ , and  $x_k^* \xrightarrow{w^*} x^*$  satisfying  $x_k^* \in \widehat{\partial}_{\varepsilon_k} \varphi(x_k)$  for all  $k \in \mathbb{N}$ . Thus there are neighborhoods  $U_k$  of  $x_k$  such that

$$\varphi(x) - \varphi(x_k) - \langle x_k^*, x - x_k \rangle \geq -2\varepsilon_k \|x - x_k\| \quad \text{for all } x \in U_k, \quad k \in \mathbb{N} .$$

The latter means that for any fixed  $k$  the function

$$\psi_k(x) := \varphi(x_k + x) - \langle x_k^*, x \rangle + 2\varepsilon_k \|x\|$$

attains a local minimum at  $x = 0$ . Denoting  $\varphi_1(x) := \varphi(x_k + x)$  and  $\varphi_2(x) := -\langle x_k^*, x \rangle + 2\varepsilon_k \|x\|$ , we represent  $\psi_k$  as the sum of two functions satisfying the assumptions in (M3). Employ (M3) with  $\eta = \varepsilon_k$  and then (M1) and (M2). This gives  $u_k \in X$  such that  $\|u_k\| \leq \varepsilon_k$ ,  $|\varphi(x_k + u_k) - \varphi(x_k)| \leq \varepsilon_k$ , and

$$x_k^* \in \widehat{\mathcal{D}}\varphi(x_k + u_k) + 3\mathcal{B}^*, \quad k \in \mathbb{N} .$$

Passing to the limit as  $k \rightarrow \infty$ , we arrive at the desired conclusion.  $\triangle$

It follows from the above proof that  $\widehat{\mathcal{D}}$  may be an  $\ell$ -presubdifferential on the class of Lipschitz continuous function  $\varphi: X \rightarrow \mathbb{R}$  with modulus  $\ell > 0$  if property (M3) is required to hold only for such functions. When  $\varphi = \delta(\cdot; \Omega)$ , the minimality property in Proposition 2.45 corresponds to the result of Proposition 1.31 for the case of subdifferentially generated normal structures, while the latter result ensures the minimality of the basic normal cone without such an assumption.

### 2.5.2 Specific Normal and Subdifferential Structures

As proved in Subject. 2.4.1, our basic normal cone and subdifferential provide a constructively defined class of *sequential* normal and subdifferential structures generated by Fréchet normals and subgradients in arbitrary Asplund spaces. Let us discuss some other remarkable classes of generalized normals and subgradients that satisfy the above requirements to abstract (pre)normal and (pre)subdifferential structures on appropriate Banach space.

**A. Convex-Valued Constructions by Clarke.** We start with Clarke's constructions of generalized normals to sets and subgradients of extended-real-valued functions that produce *topological* normal and subdifferential structures

on arbitrary *Banach spaces* by the following *four-step* procedure; see Clarke [255] for more details and proofs. First let  $\varphi$  be Lipschitz continuous around  $\bar{x} \in X$  with modulus  $\ell$ . The *generalized directional derivative* of  $\varphi$  at  $\bar{x}$  in the direction  $h$  is

$$\varphi^\circ(\bar{x}; v) := \limsup_{\substack{x \rightarrow \bar{x} \\ t \downarrow 0}} \frac{\varphi(x + tv) - \varphi(x)}{t}. \quad (2.69)$$

The function  $\varphi^\circ(\bar{x}; \cdot): X \rightarrow \mathbb{R}$  happens to be convex for any Lipschitzian  $\varphi$ ; moreover, (2.69) is upper semicontinuous in both variables with  $\varphi^\circ(\bar{x}; -v) = (-\varphi)^\circ(\bar{x}; v)$  and  $|\varphi^\circ(\bar{x}; v)| \leq \ell \|v\|$  for all  $v \in X$ . Then the *generalized gradient* of a locally Lipschitzian function is defined by

$$\partial_C \varphi(\bar{x}) := \{x^* \in X^* \mid \langle x^*, v \rangle \leq \varphi^\circ(\bar{x}; v) \text{ for any } v \in X\}. \quad (2.70)$$

It follows from (2.70) and the properties of  $\varphi^\circ$  that  $\partial_C \varphi(\bar{x})$  is a nonempty, weak\* compact, convex subset of  $X^*$  with  $\|x^*\| \leq \ell$  for all  $x^* \in \partial_C \varphi(\bar{x})$  and the classical plus-minus symmetry

$$\partial_C(-\varphi)(\bar{x}) = -\partial_C \varphi(\bar{x}) \text{ for Lipschitzian } \varphi. \quad (2.71)$$

The next step is to define the *Clarke normal cone* to  $\Omega \subset X$  by

$$N_C(\bar{x}; \Omega) := \text{cl}^* \left\{ \bigcup_{\lambda > 0} \lambda \partial_C \text{dist}(\bar{x}; \Omega) \right\}, \quad \bar{x} \in \Omega, \quad (2.72)$$

through the generalized gradient of the Lipschitzian distance function, with  $N_C(\bar{x}; \Omega) := \emptyset$  for  $\bar{x} \notin \Omega$ . Finally, the *Clarke subdifferential* of a function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  is defined by

$$\partial_C \varphi(\bar{x}) := \{x^* \in X^* \mid (x^*, -1) \in N_C((\bar{x}, \varphi(\bar{x})); \text{epi } \varphi)\} \quad (2.73)$$

if  $|\varphi(\bar{x})| < \infty$  and  $\partial_C \varphi(\bar{x}) := \emptyset$  if  $|\varphi(\bar{x})| = \infty$ . Clearly the sets (2.72) and (2.73) are convex and weak\* closed in  $X^*$ . The two basic facts ensuring that (2.72) defines a *topological* normal structure on  $X$  generated by  $\bigcup_{\lambda > 0} \lambda \partial_C \text{dist}(\bar{x}; \Omega)$  are the following: the sum rule

$$\partial_C(\varphi_1 + \varphi_2)(\bar{x}) \subset \partial_C \varphi_1(\bar{x}) + \partial_C \varphi_2(\bar{x}) \quad (2.74)$$

if  $\varphi_1$  is locally Lipschitzian and  $\varphi_2$  is l.s.c. around  $\bar{x}$ , and that the graph of  $\partial_C \varphi(\cdot)$  is closed in the norm  $\times$  weak\* topology of  $X \times X^*$  if  $\varphi$  is Lipschitz continuous. Moreover, these facts imply by Proposition 2.43 that for any fixed  $\lambda > 0$  the sets  $\lambda \partial_C \text{dist}(\bar{x}; \Omega)$  define a topological normal structure on  $X$ . Note however that there are generally *strict* inclusions

$$N_C(\bar{x}; \Omega) \subset \limsup_{x \rightarrow \bar{x}} N_C(x; \Omega) \subset \text{cl}^* \left\{ \limsup_{x \rightarrow \bar{x}} N_C(x; \Omega) \right\},$$

where the first one may be strict even in finite dimensions unless  $\Omega$  is epi-Lipschitzian at  $\bar{x}$ ; see Rockafellar [1146]. Note also that the Clarke normal

cone may be too large, especially for graphs of Lipschitzian functions when it is actually a *linear subspace*; see the proof of Theorem 1.46 and its infinite-dimensional generalizations in Subsect. 3.2.4. In particular, for  $\Omega = \text{gph } |x| \subset \mathbb{R}^2$  one has

$$N_C(0; \Omega) = \mathbb{R}^2, \quad \text{while } N(0; \Omega) = \{(v_1, v_2) \mid v_2 \leq -|v_1|\} \cup \{(v_1, v_2) \mid v_2 = v_1\}$$

for the basic normal cone  $N$ . It follows from Proposition 2.45 that

$$\partial\varphi(\bar{x}) \subset \partial_C\varphi(\bar{x}) \quad \text{and} \quad N(\bar{x}; \Omega) \subset N_C(\bar{x}; \Omega)$$

in general Banach spaces. More precise relationships between these objects will be obtained in Subsect. 3.2.3 in the Asplund space setting.

**B. Approximate Normals and Subgradients.** Another type of *topological* normal and subdifferential structures was developed by Ioffe, under the name of “approximate normals and subgradients,” as an extension of Mordukhovich’s construction to arbitrary *Banach spaces*; see remarks and references in Subsect. 1.4.7 and the corresponding results of Subsect. 3.2.3 on close connections with our basic constructions in the Asplund space setting. It doesn’t seem that the adjective “approximate” reflects the essence of these constructions, while its usage in this context clearly contradicts the regular use of this word in the book; see Subsect. 1.4.7 and also remarks in Rockafellar and Wets [1165, p. 347] for motivations of the word “approximate” appearing in this setting. On the other hand, it has been widely spread in nonsmooth analysis. In what follows we put quotation marks when referring to “approximate” normals and subdifferentials in this context.

Let us describe the multistep procedure for these constructions from the paper of Ioffe [599], where the reader can find proofs, more discussions, and references. Given  $\varphi: X \rightarrow \overline{\mathbb{R}}$  finite at  $\bar{x}$ , the constructions

$$d^-\varphi(\bar{x}; v) := \liminf_{\substack{z \rightarrow v \\ t \downarrow 0}} \frac{\varphi(\bar{x} + tz) - \varphi(\bar{x})}{t},$$

$$\partial_\varepsilon^-\varphi(\bar{x}) := \{x^* \in X^* \mid \langle x^*, v \rangle \leq d^-\varphi(\bar{x}; v) + \varepsilon\|v\|\}$$

are called the *lower Dini* (or Dini-Hadamard) *directional derivative* and the *Dini  $\varepsilon$ -subdifferential* of  $\varphi$  at  $\bar{x}$ , respectively. As usual, we put  $\partial^-\varphi(\bar{x}) := \emptyset$  if  $|\varphi(\bar{x})| = \infty$ . Note that the sets  $\partial_\varepsilon^-\varphi(\bar{x})$  are always convex, while the function  $d^-\varphi(\bar{x}; \cdot)$  is not. One can check that  $\partial_\varepsilon^-\varphi(\bar{x})$  reduces to the analytic  $\varepsilon$ -subdifferential from Definition 1.83(ii) if  $\dim X < \infty$ . In general, the *A-subdifferential* of  $\varphi$  at  $\bar{x}$  is defined via topological limits involving finite-dimensional reductions of  $\varepsilon$ -subgradients as

$$\partial_A\varphi(\bar{x}) := \bigcap_{\substack{L \in \mathcal{L} \\ \varepsilon > 0}} \overline{\limsup_{x \xrightarrow{\varphi} \bar{x}} \partial_\varepsilon^-(\varphi + \delta(\cdot; L))(x)} \quad (2.75)$$

where  $\mathcal{L}$  is the collection of all finite-dimensional subspaces of  $X$  and where  $\overline{\text{Lim sup}}$  stands for the *topological* counterpart of the Painlevé-Kuratowski upper limit (1.1) with sequences replaced by nets. Further, the  $G$ -normal cone  $N_G$  and its *nucleus*  $\tilde{N}_G$  to  $\Omega$  at  $\bar{x} \in \Omega$  are defined by

$$N_G(\bar{x}; \Omega) := \text{cl}^* \tilde{N}_G(\bar{x}; \Omega) \quad \text{and} \quad \tilde{N}_G(\bar{x}; \Omega) := \bigcup_{\lambda > 0} \lambda \partial_A \text{dist}(\bar{x}; \Omega), \quad (2.76)$$

respectively, with  $N_G(\bar{x}; \Omega) = \tilde{N}_G(\bar{x}; \Omega) = \emptyset$  if  $\bar{x} \notin \Omega$ . Finally, the  $G$ -subdifferential of  $\varphi$  at  $\bar{x}$  is defined geometrically as

$$\partial_G \varphi(\bar{x}) := \{x^* \in X^* \mid (x^*, -1) \in N_G((\bar{x}, \varphi(\bar{x})); \text{epi } \varphi)\}, \quad (2.77)$$

while its  $G$ -nucleus  $\tilde{\partial}_G \varphi(\bar{x})$  corresponds to (2.77) with  $N_G$  replaced by  $\tilde{N}_G$ . One always has

$$\tilde{\partial}_G \varphi(\bar{x}) \subset \partial_G \varphi(\bar{x}) \subset \partial_A \varphi(\bar{x}),$$

where equalities hold if  $\varphi$  is locally Lipschitzian around  $\bar{x}$ . For closed sets  $\Omega$  the graph of  $N_G(\cdot; \Omega)$  is closed in the norm  $\times$  weak\* topology of  $X \times X^*$ . Moreover, both  $\partial_G \varphi$  and  $\tilde{\partial}_G \varphi$  satisfy the sum rule in form (2.74) if  $\varphi_1$  is locally Lipschitzian and  $\varphi_2$  is l.s.c. around  $\bar{x}$ . Hence  $N_G(\cdot; \Omega)$  and  $\lambda \partial_A \text{dist}(\cdot; \Omega)$  provide topological normal structures on  $X$  and

$$\partial \varphi(\bar{x}) \subset \tilde{\partial}_G \varphi(\bar{x}), \quad N(\bar{x}; \Omega) \subset \tilde{N}_G(\bar{x}; \Omega)$$

by Proposition 2.45. Note that the latter inclusions may be *strict*, even in the case of Lipschitz continuous functions on spaces with Fréchet smooth renorms; see Example 3.61. In Subsect. 3.2.3 we obtain more precise relationships between these constructions in the general case of Asplund spaces.

**C. Viscosity Subdifferentials.** Next we consider normal and subgradient constructions related to the so-called *viscosity subdifferentials* that generally make sense in *smooth* Banach spaces admitting smooth renorms (or bump functions) with respect to some bornology; see Remark 2.11. The following description is based on the paper by Borwein, Mordukhovich and Shao [151], where one can find more details and references on the genesis and applications of such constructions; see also the book by Borwein and Zhu [164].

Given a bornology  $\beta$  on a Banach space  $X$ , we denote by  $X_\beta^*$  the dual space  $X^*$  endowed with the topology of uniform convergence on  $\beta$ -sets. The latter convergence agrees with the norm convergence in  $X^*$  when  $\beta$  is the (strongest) Fréchet bornology, and with the weak\* convergence in  $X^*$  when  $\beta$  is the (weakest) Gâteaux bornology. A function  $\theta: X \rightarrow \overline{\mathbb{R}}$  is  $\beta$ -differentiable at  $\bar{x}$  with  $\beta$ -derivative  $\nabla_\beta \theta(\bar{x}) \in X^*$  provided that

$$t^{-1} \left( \theta(\bar{x} + tv) - \theta(\bar{x}) - t \langle \nabla_\beta \theta(\bar{x}), v \rangle \right) \rightarrow 0$$

as  $t \rightarrow 0$  uniformly in  $v \in V$  for every  $V \in \beta$ . This function is said to be  $\beta$ -smooth around  $\bar{x}$  if it is  $\beta$ -differentiable at each point of a neighborhood  $U$

of  $\bar{x}$  and  $\nabla_\beta \theta: X \rightarrow X^*$  is continuous on  $U$ . The latter requirement is essential; in the case of  $\beta = \mathcal{F}$ , the Fréchet bornology on  $X$ , it means that  $\nabla \theta: X \rightarrow X^*$  is norm-to-norm continuous around  $\bar{x}$ . Note that in the Fréchet case the  $\beta$ -smoothness of  $\theta$  implies its Lipschitz continuity around  $\bar{x}$ , which may not happen for weaker bornologies  $\beta < \mathcal{F}$ .

Now, given  $\varphi: X \rightarrow \overline{\mathbb{R}}$  finite at  $\bar{x}$ , its *viscosity  $\beta$ -subdifferential of rank  $\lambda > 0$*  at  $\bar{x}$  is the set  $\partial_\beta^\lambda \varphi(\bar{x})$  of all  $x^* \in X^*$  with the following properties: there are a neighborhood  $U$  of  $\bar{x}$  and a  $\beta$ -smooth function  $\theta: U \rightarrow \mathbb{R}$  such that  $\theta$  is Lipschitz continuous on  $U$  with modulus  $\lambda$ ,  $\nabla_\beta \theta(\bar{x}) = x^*$ , and  $\varphi - \theta$  attains a local minimum at  $\bar{x}$ . The corresponding set of  *$\beta$ -normals of rank  $\lambda$*  to  $\Omega \subset X$  at  $\bar{x} \in \Omega$  is defined by  $N_\beta^\lambda(\bar{x}; \Omega) := \partial_\beta^\lambda \delta(\bar{x}; \Omega)$ . The unions

$$\partial_\beta \varphi(\bar{x}) := \bigcup_{\lambda > 0} \partial_\beta^\lambda \varphi(\bar{x}), \quad N_\beta(\bar{x}; \Omega) := \bigcup_{\lambda > 0} N_\beta^\lambda(\bar{x}; \Omega) \quad (2.78)$$

are called the *viscosity  $\beta$ -subdifferential* of  $\varphi$  at  $\bar{x}$  and the *viscosity  $\beta$ -normal cone* of  $\Omega$  at  $\bar{x}$ , respectively. Note that  $\theta(\cdot)$  in the above definition can be equivalently chosen to be *concave* if  $X$  admits a  $\beta$ -smooth renorm.

Employing the variational descriptions of Fréchet normals and subgradients in Theorems 1.30 and 1.88, we conclude that

$$\partial_{\mathcal{F}} \varphi(\bar{x}) = \widehat{\partial} \varphi(\bar{x}) \quad \text{and} \quad N_{\mathcal{F}}(\bar{x}) = \widehat{N}(\bar{x}; \Omega)$$

if  $X$  admits an  $\mathcal{F}$ -smooth bump function. These constructions may be different in more general settings of Banach and Asplund spaces. Note that, in contrast to  $\widehat{\partial} \varphi(\cdot)$  and  $\widehat{N}(\cdot; \Omega)$ , the viscosity constructions (2.78) don't reveal useful properties without smoothness assumptions on the space in question.

It follows from the results of the afore-mentioned paper [151] that  $\partial_\beta^\lambda \varphi(\cdot)$  defines a *presubdifferential* structure on a  $\beta$ -smooth space  $X$  for any  $\lambda > 1$ . Hence  $N_\beta^\lambda(\cdot; \Omega)$  defines the corresponding prenormal structure under these conditions. By Proposition 2.45 we have

$$\partial \varphi(\bar{x}) \subset \limsup_{x \xrightarrow{\varphi} \bar{x}} \partial_\beta \varphi(x), \quad N(\bar{x}; \Omega) \subset \limsup_{x \xrightarrow{\Omega} \bar{x}} N_\beta(x; \Omega) \quad (2.79)$$

in  $\beta$ -smooth spaces. It doesn't seem to be true that viscosity subdifferentials (2.78) and their *sequential* limits in (2.79) enjoy the semi-Lipschitzian sum rules of the corresponding types (b) and (c) in Proposition 2.33 on  $\beta$ -smooth spaces with  $\beta < \mathcal{F}$ . On the other hand,

$$\widetilde{\partial}_G \varphi(\bar{x}) = \bigcup_{\lambda > 0} \text{cl}^* \left\{ \limsup_{x \xrightarrow{\varphi} \bar{x}} \partial_\beta^\lambda \varphi \right\}, \quad \partial_A \varphi(\bar{x}) = \overline{\limsup_{x \xrightarrow{\varphi} \bar{x}} \partial_\beta \varphi(x)}$$

for the *nucleus* of the  $G$ -subdifferential (2.77) and for the  $A$ -subdifferential (2.75) of any l.s.c. function on an arbitrary  $\beta$ -smooth Banach space; cf. Borwein and Ioffe [147, Theorem 2] and Mordukhovich, Shao and Zhu [954, Theorem 6.1], respectively.

**D. Proximal Constructions.** Let us consider the *Hilbert space* setting that is the closest to finite dimensions and allows one to construct prenormal and presubdifferential structures defined through the *Euclidean metric*. Given a closed subset  $\Omega \subset X$  of a Hilbert space and the Euclidean projector  $\Pi(\cdot; \Omega)$ , the conic set

$$N_P(\bar{x}; \Omega) := \text{cone} [\Pi^{-1}(\bar{x}; \Omega) - \bar{x}] \quad (2.80)$$

is the *proximal normal cone* to  $\Omega$  at  $\bar{x} \in \Omega$ . It follows from the Euclidean norm properties (cf. the proof of Theorem 1.6 above) that  $x^* \in N_P(\bar{x}; \Omega)$  if and only if there is  $\alpha > 0$  such that

$$\langle x^*, x - \bar{x} \rangle \leq \alpha \|x - \bar{x}\|^2 \quad \text{for all } x \in \Omega.$$

This obviously implies that  $N_P(\bar{x}; \Omega)$  is a convex subcone of  $\widehat{N}(\bar{x}; \Omega)$ . In contrast to the latter one,  $N_P(\bar{x}; \Omega)$  may not be closed even in finite dimensions; moreover, its closure may be different from  $\widehat{N}(\bar{x}; \Omega)$ . A simple example is provided by the epigraph of a smooth function:

$$\Omega = \text{epi } \varphi \subset \mathbb{R}^2 \quad \text{with } \varphi(x) = -|x|^{3/2} \quad \text{at } \bar{x} = (0, 0),$$

where  $N_P(\bar{x}; \Omega) = \{(0, 0)\}$  and  $\widehat{N}(\bar{x}; \Omega) = \{(v_1, v_2) \mid v_1 = 0, v_2 \leq 0\}$ .

A functional counterpart of the proximal normal cone (2.70) is the *proximal subdifferential* of a proper l.s.c. function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  at  $\bar{x} \in \text{dom } \varphi$  defined as

$$\partial_P \varphi(\bar{x}) := \left\{ x^* \in X^* \mid \liminf_{x \rightarrow \bar{x}} \frac{\varphi(x) - \varphi(\bar{x}) - \langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|^2} > -\infty \right\}, \quad (2.81)$$

which is a convex subset of the Fréchet subdifferential  $\widehat{\partial} \varphi(\bar{x})$  and can be equivalently described by  $(x^*, -1) \in N_P((\bar{x}, \varphi(\bar{x})); \text{epi } \varphi)$ . Note that the proximal subdifferential may be *empty* even for *smooth* functions as in the above example, where  $\partial_P \varphi(0) = \emptyset$  while  $\widehat{\partial} \varphi(0) = \{0\}$ . Nevertheless, for every proper l.s.c. function  $\varphi$  finite at  $\bar{x}$  the following holds: given any  $x^* \in \widehat{\partial} \varphi(\bar{x})$ , there are sequences  $x_k \xrightarrow{\varphi} \bar{x}$  and  $x_k^* \in \partial_P \varphi(x_k)$  such that  $\|x_k^* - x^*\| \rightarrow 0$  as  $k \rightarrow \infty$ ; see Loewen [802, Theorem 5.5]. Therefore

$$\partial \varphi(\bar{x}) = \limsup_{x \xrightarrow{\varphi} \bar{x}} \partial_P \varphi(x) \quad \text{and} \quad N(\bar{x}; \Omega) = \limsup_{x \xrightarrow{\Omega} \bar{x}} N_P(x; \Omega).$$

A crucial fact ensuring that (2.81) defines a *presubdifferential* structure on a Hilbert space  $X$  (hence  $N_P(\cdot; \Omega)$  defines the corresponding prenormal structure) follows from the fuzzy sum rule for  $\partial_P \varphi(\cdot)$  proved in Ioffe and Rockafellar [616, Theorem 2] and in Clarke et al. [265, Theorem 1.8.3].

**E. Derivate Sets.** In conclusion of this subsection we compare our subdifferential constructions with generalized derivatives based on the idea of *uniformly approximating* nonsmooth functions by smooth (finitely differentiable) functions. Recall that a mapping  $f: X \rightarrow Y$  between Banach spaces is *finitely*

*differentiable* at  $\bar{x}$  with the derivative  $\nabla f(\bar{x})$  if for every finite-dimensional subspace  $X \subset X$  the mapping  $z \rightarrow f(x+z): Z \rightarrow Y$  is differentiable at the origin and its derivative agrees with the restriction of  $\nabla f(\bar{x})$  to  $Z$ .

Given  $\varphi: X \rightarrow \overline{\mathbb{R}}$  on a Banach space  $X$  and a point  $\bar{x} \in X$  with  $|\varphi(\bar{x})| < \infty$ , we denote by  $\mathcal{A}\varphi(\bar{x})$  a subset of  $X^*$  with the following properties: for any  $\varepsilon, \alpha > 0$  there are  $\gamma \in (0, \alpha]$  and a continuously finitely differentiable function  $\psi: X \rightarrow \mathbb{R}$  such that

$$|\varphi(x) - \psi(x)| \leq \varepsilon\gamma \quad \text{and} \quad \nabla\psi(x) \in \mathcal{A}\varphi(\bar{x}) \quad \text{for all } x \in \bar{x} + \gamma\mathcal{B}.$$

The *derivate set*  $\mathcal{A}\varphi(\bar{x})$  is a derivative-like object, which is not uniquely defined. If  $\varphi$  is continuous around  $\bar{x}$  and can be represented as the uniform limit of a sequence of continuously finitely differentiable functions  $\varphi_i$ ,  $i \in \mathbb{N}$ , then for any  $\gamma > 0$  and  $j \in \mathbb{N}$  one can take

$$\mathcal{A}\varphi(\bar{x}) = \bigcup_{\substack{\|x - \bar{x}\| \leq \gamma \\ i \geq j}} \{\nabla\varphi_i(x)\}.$$

The following result shows that for every function  $\varphi$  the Fréchet subdifferential of  $\varphi$  at  $\bar{x}$  is contained in the *norm* closure of *any* derivate set  $\mathcal{A}\varphi(\bar{x})$  obtained via a uniform approximation by finitely smooth functions.

**Theorem 2.46 (derivate sets and Fréchet subgradients).** *Let  $X$  be a Banach space, and let  $\mathcal{A}\varphi(\bar{x})$  be a derivate set of  $\varphi: X \rightarrow \overline{\mathbb{R}}$  finite at  $\bar{x}$ . Then*

$$\widehat{\partial}\varphi(\bar{x}) \subset \text{cl } \mathcal{A}\varphi(\bar{x}) \quad \text{if } \mathcal{A}\varphi(\bar{x}) \neq \emptyset.$$

**Proof.** Let  $\bar{x}^* \notin \text{cl } \mathcal{A}\varphi(\bar{x})$ . Then there is  $\eta > 0$  such that

$$\|\bar{x}^* - x^*\| > \eta \quad \text{for all } x^* \in \mathcal{A}\varphi(\bar{x}). \quad (2.82)$$

Put  $\bar{\varepsilon} := \eta/4$  and for each  $k \in \mathbb{N}$  select a number  $\gamma_k$  and a function  $\psi_k$  according to the definition of the derivate set  $\mathcal{A}\varphi(\bar{x})$  with  $\varepsilon = \bar{\varepsilon}/4$  and  $\alpha = 1/k$ .

Next we define, for some positive integer  $N_k$ , a finite set of points  $x_i \in X$ ,  $i = 0, 1, \dots, N_k$ , from the following conditions:

- (a)  $x_0 = \bar{x}$ ,  $x_{i+1} = x_i + hz_i$ ,  $i = 0, 1, \dots, N_k - 1$ ;
- (b)  $\|z_i\| = 1$ ,  $i = 0, 1, \dots, N_k - 1$ ;
- (c)  $h = \gamma_k/(2N_k)$ ;
- (d)  $\langle \bar{x}^* - \nabla\psi_k(x_i), z_i \rangle > \eta$ ,  $i = 0, 1, \dots, N_k - 1$ .

Note that it is possible to find  $z_i$  satisfying (d) because  $\psi$  is finitely differentiable at  $x_i$  with  $\nabla = \psi(x_i) \in \mathcal{A}\varphi(\bar{x})$ , (2.82) holds, and

$$\|x_i - \bar{x}\| \leq N_k h = \gamma_k/2 \quad \text{for } i = 1, \dots, N_k \quad (2.83)$$

due to (a), (b), and (c). When  $N_k$  is sufficiently large, one has

$$\begin{aligned}
 & \psi_k(x_{N_k}) - \psi_k(\bar{x}) - \langle \bar{x}^*, x_{N_k} - \bar{x} \rangle \\
 &= \sum_{i=0}^{N_k-1} \left( \int_0^h \langle \nabla \psi_k(x_i + tz_i), z_i \rangle dt - h \langle \bar{x}^*, z_i \rangle \right) \\
 &\leq h \sum_{i=0}^{N_k} \langle \psi_k(x_i) - \bar{x}^*, z_i \rangle + \frac{\eta \gamma_k}{4}.
 \end{aligned}$$

This implies, by (d) and (c), that

$$\psi_k(x_{N_k}) - \psi_k(\bar{x}) - \langle \bar{x}^*, x_{N_k} - \bar{x} \rangle < -\eta \gamma_k / 2 = \bar{\varepsilon} \gamma_k. \quad (2.84)$$

Now recall that  $\psi_k$  approximates the original function  $\varphi$  by

$$|\varphi(x) - \psi_k(x)| \leq \bar{\varepsilon} \gamma_k / 4 \quad \text{whenever } x \in \bar{x} + \gamma_k \mathcal{B}.$$

Combining this with (2.83) and (2.84), we finally get

$$\varphi(x_{N_k}) - \varphi(\bar{x}) - \langle \bar{x}^*, x_{N_k} - \bar{x} \rangle \leq \bar{\varepsilon} \gamma_k / 2 \leq -\bar{\varepsilon} \|x_{N_k} - \bar{x}\|.$$

Since  $x_{N_k} \rightarrow \bar{x}$  as  $k \rightarrow \infty$ , the latter means that  $\bar{x}^* \notin \widehat{\partial} \varphi(\bar{x})$ , which ends the proof of the theorem.  $\triangle$

Theorem 2.46 concerns relationships between Fréchet subgradients and derivate sets of real-valued functions that can be approximated by smooth functions near the point under consideration. It easily implies corresponding results for mappings  $f: X \rightarrow Y$  involving their scalarization. In particular, we deduce from Theorem 2.46 the following relationship between Fréchet subgradients and screens introduced by Halkin [544] for mappings between finite-dimensional spaces.

Recall that, given  $f: U \rightarrow \mathbb{R}^m$  defined on an open subset  $U \subset \mathbb{R}^n$ , a nonempty set  $\mathcal{U} \subset \mathbb{R}^{mn}$  is called a *screen* of  $f$  at  $\bar{x} \in U$  if for every  $\varepsilon, \alpha > 0$  there exist  $\gamma > 0$  and a  $\mathcal{C}^1$  mapping  $g: B_\gamma^n(\bar{x}) \rightarrow \mathbb{R}^m$  such that  $B_\gamma^n(\bar{x}) \subset U$ ,

$$\|f(x) - g(x)\| \leq \varepsilon \gamma, \quad \text{and} \quad \nabla g(x) \in \mathcal{U} + \varepsilon \mathcal{B}^{mn} \quad \text{for all } x \in B_\gamma^n(\bar{x}),$$

where  $B_\gamma^n(\bar{x}) := \bar{x} + \gamma \mathcal{B}_{\mathbb{R}^n}$  and  $\mathcal{B}^{mn}$  stands for the closed unit ball in  $\mathbb{R}^{mn}$ .

**Corollary 2.47 (relationship between Fréchet subgradients and screens).** *Let  $\mathcal{U} \subset \mathbb{R}^{mn}$  be a screen of a mapping  $f: U \rightarrow \mathbb{R}^m$  at  $\bar{x} \in U \subset \mathbb{R}^n$ . Then*

$$\widehat{\partial} \langle y^*, f \rangle(\bar{x}) \subset \text{cl} \{A^* y^* \mid A \in \mathcal{U}\} \quad \text{for all } y^* \in \mathbb{R}^m.$$

**Proof.** Given  $y^* \in \mathbb{R}^m$  and a screen  $\mathcal{U}$  of  $f$  at  $\bar{x}$ , it is not hard to check that the set  $\{A^* y^* \mid A \in \mathcal{U}\}$  satisfies all the above properties of the derivate set  $\mathcal{A} \varphi(\bar{x})$  for the scalarized function  $\varphi(x) := \langle y^*, f \rangle(x)$  at  $\bar{x}$ .  $\triangle$



A screen of a mapping is not uniquely defined. Particular examples of screens are given by *derivate containers* of Warga [1316], which include Clarke's *generalized Jacobian* for locally Lipschitzian mappings between finite-dimensional spaces. Warga [1319] also introduced the concept of *directional derivate containers* for mappings between infinite-dimensional spaces. Theorem 2.46 allows us to obtain the following relationships between the latter construction for mappings (see the afore-mentioned papers by Warga for the exact definition) and Fréchet subgradients of their scalarizations.

**Corollary 2.48 (relationship between Fréchet subgradients and derivate containers).** *Consider a directional derivate container  $\{\Lambda^\varepsilon f(\bar{x}) \mid \varepsilon > 0\}$  of a mapping  $f: \Omega \rightarrow Y$  at  $\bar{x} \in \text{int } \Omega$ , where  $\Omega \subset X$  is a convex compact set, and where the spaces  $X$  and  $Y$  are Banach. Then for any  $y^* \in Y^*$ ,  $\varepsilon > 0$ , and  $\eta > 0$  there is  $\gamma > 0$  such that*

$$\widehat{\partial}\langle y^*, f \rangle(x) \subset \{A^* y^* \mid A \in \Lambda^\varepsilon f(\bar{x})\} + \eta B^* \text{ whenever } x \in \bar{x} + \gamma B.$$

Note that the assumption  $\bar{x} \in \text{int } \Omega$  is essential for the validity of the latter result. Indeed, for the function  $f: [0, 1] \rightarrow \mathbb{R}$  with  $f \equiv 0$  extended by  $\infty$  outside of  $[0, 1]$ , we clearly have  $\widehat{\partial}f(1) = [0, \infty)$ , while the singleton  $\{0\}$  is a directional derivate container of  $f$  at  $\bar{x} = 1$ .

Observe that the derivative-like constructions in Theorem 2.46 and Corollaries 2.47 and 2.48 are generally related to *presubdifferential* structures, which lead to *robust subdifferentials* and corresponding generalized derivatives of mappings via some *regularization* procedure. To this end let us recall the definition of the *minimal derivate container* by Warga

$$\begin{aligned} \Lambda^0 f(\bar{x}) &:= \limsup_{\substack{x \rightarrow \bar{x} \\ k \rightarrow \infty}} \{\nabla f_k(x)\} \\ &= \bigcap_{j=1}^{\infty} \bigcap_{\gamma > 0} \text{cl} \bigcup_{\substack{\|x - \bar{x}\| \leq \gamma \\ i \geq j}} \{\nabla f_i(x)\} \end{aligned}$$

for a continuous mapping  $f: X \rightarrow Y$  between finite-dimensional spaces that admits a uniform approximation by a sequence of  $C^1$  mappings  $f_k$ . It follows from the results obtained that

$$\partial\langle y^*, f \rangle(\bar{x}) \subset \{A^* y^* \mid A \in \Lambda^0 f(\bar{x})\} \text{ for all } y^* \in Y^*,$$

which gives the inclusion

$$\partial^0 \varphi(\bar{x}) := \partial \varphi(\bar{x}) \cup \partial^+ \varphi(\bar{x}) \subset \Lambda^0 \varphi(\bar{x}) \quad (2.85)$$

for the two-sided/symmetric generalized differential (1.46) of a real-valued function  $\varphi$  continuous around  $\bar{x}$ . The following example illustrates (2.85) and other relationships between various subgradients studied above.

**Example 2.49 (computing subgradients of Lipschitzian functions).**

Consider the function

$$\varphi(x) := |x_1| + x_2, \quad x = (x_1, x_2) \in \mathbb{R}^2,$$

which is Lipschitz continuous on  $\mathbb{R}^2$ . Based on representation (1.51), we compute Fréchet subgradients of  $\varphi$  at every point  $x \in \mathbb{R}^2$  as follows:

$$\widehat{\partial}\varphi(x) = \begin{cases} (1, 1) & \text{if } x_1 > 0, x_1 + x_2 > 0, \\ (-1, -1) & \text{if } x_1 > 0, x_1 + x_2 < 0, \\ (-1, 1) & \text{if } x_1 < 0, x_1 - x_2 < 0, \\ (1, -1) & \text{if } x_1 < 0, x_1 - x_2 > 0, \\ \{(v, 1) \mid -1 \leq v \leq 1\} & \text{if } x_1 = 0, x_2 > 0, \\ \{(v, v) \mid -1 \leq v \leq 1\} & \text{if } x_1 > 0, x_1 + x_2 = 0, \\ \{(v, -v) \mid -1 \leq v \leq 1\} & \text{if } x_1 < 0, x_1 - x_2 = 0, \\ \{(v_1, v_2) \mid |v_1| \leq v_2 \leq 1\} & \text{if } x_1 = 0, x_2 = 0, \\ \emptyset & \text{if } x_1 = 0, x_2 < 0. \end{cases}$$

Similarly, based on representation (1.52), we compute Fréchet *upper* subgradients of the above function by

$$\widehat{\partial}^+\varphi(x) = \begin{cases} (1, 1) & \text{if } x_1 > 0, x_1 + x_2 > 0, \\ (-1, -1) & \text{if } x_1 > 0, x_1 + x_2 < 0, \\ (-1, 1) & \text{if } x_1 < 0, x_1 - x_2 < 0, \\ (1, -1) & \text{if } x_1 < 0, x_1 - x_2 > 0, \\ \{(v, -1) \mid -1 \leq v \leq 1\} & \text{if } x_1 = 0, x_1 - x_2 < 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

Now using the limiting representation (1.56) of the basic subdifferential in Theorem 1.89 and the symmetric representation of upper subgradients, we arrive at the subgradient sets

$$\partial\varphi(0) = \{(v_1, v_2) \mid |v_1| \leq v_2 \leq 1\} \cup \{(v_1, v_2) \mid v_2 = -|v_1|, -1 \leq v_1 \leq 1\},$$

$$\partial^+\varphi(0) = \{(v, -1) \mid -1 \leq v \leq 1\} \cup \{(1, -1), (1, 1)\},$$

$$\partial^0\varphi(0) = \partial\varphi(0) \cup \{(v, -1) \mid -1 \leq v \leq 1\}.$$

Warga's minimal derivate container for this function is the nonconvex set

$$\Lambda^0\varphi(0) = \{\alpha(v, 1) \mid \alpha, v \in [-1, 1]\},$$

which is the union of two triangles with vertices at  $(0,0)$ ,  $(1,1)$ ,  $(-1, 1)$  and  $(0,0)$ ,  $(1, -1)$ ,  $(-1, 1)$ , respectively. Clarke's generalized gradient is the whole unit square  $[-1, 1] \times [-1, 1]$ .

### 2.5.3 Abstract Versions of Extremal Principle

In the conclusion of this section we establish approximate and exact versions of the extremal principle valid, respectively, for abstract prenormal and normal structures considered in Subsect. 2.5.1. They hold, in particular, for the specific classes of generalized normals in appropriate Banach spaces described in Subsect. 2.5.2.

We'll see that an approximate version of the extremal principle doesn't impose any requirements on abstract prenormal structures in addition to those formulated in Definition 2.41. In contrast to Theorem 2.22, we obtain the exact extremal principle in Banach spaces in two limiting forms—sequential and topological—involving sequential and topological normal structures, respectively. Note that both limiting forms hold under the following *sequential* normal compactness condition formulated in terms of the corresponding prenormal structure similarly to Definition 1.20.

**Definition 2.50 (abstract sequential normal compactness).** *Let  $\widehat{\mathcal{N}}$  define a prenormal structure on a Banach space  $X$ . We say that  $\Omega \subset X$  is  $\widehat{\mathcal{N}}$ -SEQUENTIALLY NORMALLY COMPACT at  $\bar{x} \in \Omega$  if for any sequence  $(x_k, x_k^*) \in X \times X^*$  satisfying*

$$x_k^* \in \widehat{\mathcal{N}}(x_k; \Omega), \quad x_k \rightarrow \bar{x}, \quad x_k^* \xrightarrow{w^*} 0$$

*one has  $\|x_k^*\| \rightarrow 0$  as  $k \rightarrow \infty$ .*

This property obviously holds in finite-dimensional spaces for any prenormal structure  $\widehat{\mathcal{N}}$ . When  $\widehat{\mathcal{N}} = \widehat{N}$ , the prenormal cone of Definition 1.1(i), we studied the SNC property and its modification in Subsect. 1.1.3 for arbitrary Banach spaces. In particular, we established the relationships with the compactly epi-Lipschitzian (CEL) property of sets. In addition to Remark 1.27, let us mention that, for any closed set  $\Omega$  in a Banach space  $X$ , the CEL property

is *equivalent* to the *topological* counterpart of the SNC property in Definition 2.50, where sequences  $(x_k, x_k^*)$  are replaced with *bounded nets* and the prenormal structure  $\widehat{\mathcal{N}}$  is given by the nucleus of the  $G$ -normal cone in (2.76). It is proved by Ioffe [607, Theorem 3] and holds also for prenormal structures defined by the viscosity  $\beta$ -normal cones (2.78) on Banach spaces admitting a Lipschitzian  $\beta$ -smooth bump function. Let us call the net counterpart of the SNC property in Definition 2.50 by the *topological normal compactness* (TNC) of  $\Omega$  at  $\bar{x}$  with respect to  $\widehat{\mathcal{N}}$  and observe that  $\text{CEL} \not\Rightarrow \text{TNC}$  for the case of Clarke's normal cone (2.72), as follows from Example 4.1 in Borwein [138] for  $X = \ell^\infty$ .

Obviously  $\text{TNC} \Rightarrow \text{SNC}$  for any  $\widehat{\mathcal{N}}$ . It is proved by Fabian and Morukhovich [422] that these properties coincide on Banach spaces  $X$  that are *weakly compactly generated* (WCG), i.e.,  $X = \text{cl}(\text{span } K)$  for some weakly compact set  $K \subset X$ . This class includes all reflexive spaces as well as all separable Banach spaces. On the other hand, the SNC property may be *strictly weaker* than its TNC counterpart in general Banach (and Asplund) space settings, even for the case of convex sets; see examples in [422].

**Theorem 2.51 (abstract versions of the extremal principle).** *Let  $\{\Omega_1, \Omega_2, \bar{x}\}$  be an extremal system of closed sets in a Banach space  $X$ , and let  $\widehat{\mathcal{N}}$  define a prenormal structure on  $X$ . The following hold:*

(i) *For every  $\varepsilon > 0$  there are  $x_i \in \Omega_i \cap (\bar{x} + \varepsilon \mathcal{B})$ ,  $i = 1, 2$ , and  $x^* \in X^*$  with  $\|x^*\| = 1$  such that*

$$x^* \in (\widehat{\mathcal{N}}(x_1; \Omega_1) + \varepsilon \mathcal{B}^*) \cap (-\widehat{\mathcal{N}}(x_2; \Omega_2) + \varepsilon \mathcal{B}^*). \quad (2.86)$$

(ii) *Assume that one of the sets  $\Omega_i$ ,  $i = 1, 2$ , is  $\widehat{\mathcal{N}}$ -sequentially normally compact at  $\bar{x}$ . Then there is  $x^* \in \mathcal{B}^* \setminus \{0\}$  such that*

$$x^* \in \overline{\mathcal{N}}(\bar{x}; \Omega_1) \cap (-\overline{\mathcal{N}}(\bar{x}; \Omega_2)), \quad (2.87)$$

where  $\overline{\mathcal{N}}$  stands for the topological normal structure (2.67) generated by  $\widehat{\mathcal{N}}$ . If in addition the dual ball  $\mathcal{B}^* \subset X^*$  is weak\* sequentially compact, then

$$x^* \in \mathcal{N}(\bar{x}; \Omega_1) \cap (-\mathcal{N}(\bar{x}; \Omega_2)) \quad (2.88)$$

for some  $x^* \in \mathcal{B}^* \setminus \{0\}$ , where  $\mathcal{N}$  stands the sequential normal structure (2.66) generated by  $\widehat{\mathcal{N}}$ .

**Proof.** First justify (i) following basically the procedure in the proof of Lemma 2.32(ii). Fix an arbitrary  $\varepsilon > 0$ . Given a local extremal point  $\bar{x}$  of the set system  $\{\Omega_1, \Omega_2\}$ , we find a neighborhood  $U$  of  $\bar{x}$  and  $a \in X$  such that  $\|a\| \leq \epsilon := \varepsilon/2$  and  $(\Omega_1 + a) \cap \Omega_2 \cap U = \emptyset$ . One can always assume that  $\bar{x} + \epsilon \mathcal{B} \subset U$ . Form the function

$$\varphi(x_1, x_2) := \|x_1 - x_2 + a\| \quad \text{for } (x_1, x_2) \in X^2$$

and observe that  $\varphi(\bar{x}, \bar{x}) = \|a\| \leq \epsilon$  and

$$\varphi(x_1, x_2) > 0 \text{ if } (x_1, x_2) \in Z := [\Omega_1 \cap (\bar{x} + \epsilon \mathcal{B})] \times [\Omega_2 \cap (\bar{x} + \epsilon \mathcal{B})] .$$

We see that  $Z$  is a complete metric space with the metric induced by the sum norm on  $X^2$ , and that  $\varphi$  is continuous on  $Z$ . Applying Ekeland's variational principle in Theorem 2.26(i) to  $\varphi$  on  $Z$ , we find  $(\bar{x}_1, \bar{x}_2) \in Z$  such that

$$\varphi(\bar{x}_1, \bar{x}_2) \leq \varphi(x_1, x_2) + \epsilon (\|x_1 - \bar{x}_1\| + \|x_2 - \bar{x}_2\|) \text{ for all } (x_1, x_2) \in Z .$$

The latter implies that  $(\bar{x}_1, \bar{x}_2) \in \Omega_1 \times \Omega_2$  is a local minimizer of the function

$$\psi(x_1, x_2) := \|x_1 - x_2 + a\| + \epsilon (\|x_1 - \bar{x}_1\| + \|x_2 - \bar{x}_2\|)$$

relative to the set  $\Omega_1 \times \Omega_2$  with  $\bar{x}_1 - \bar{x}_2 + a \neq 0$ . Now applying property (H) of the prenormal structure  $\widehat{\mathcal{N}}$  in Definition 2.41 with  $\gamma := \epsilon > \epsilon$ , we find  $\tilde{x}_i \in \bar{x} + \epsilon \mathcal{B}$ ,  $i = 1, 2$ , and  $x^* \in X^*$  with  $\|x^*\| = 1$  such that

$$(-x^*, x^*) \in \widehat{\mathcal{N}}(\tilde{x}_1; \Omega_1) \times \widehat{\mathcal{N}}(\tilde{x}_2; \Omega_2) + \epsilon (\mathcal{B}^* \times \mathcal{B}^*) .$$

It follows from the constructions above that  $(\tilde{x}_1, \tilde{x}_2) \in \Omega_1 \times \Omega_2$  and  $\tilde{x}_i \in \bar{x} + \epsilon \mathcal{B}$ ,  $i = 1, 2$ . Thus we get all the relationships of the approximate extremal principle in (i).

To prove (ii), we need to pass to the limit in (i) as  $\epsilon \downarrow 0$ . Let us first justify the sequential version of the exact extremal principle in (ii) assuming that the dual ball  $\mathcal{B}^* \subset X^*$  is weak\* sequentially compact. Take a sequence  $\epsilon_k \downarrow 0$  and consider the corresponding sequences  $(x_{1k}, x_{2k}, x_k^*)$  satisfying the conclusions of (i). We have  $x_{1k} \rightarrow \bar{x}$  and  $x_{2k} \rightarrow \bar{x}$  as  $k \rightarrow \infty$ . Since  $\mathcal{B}^*$  is weak\* sequentially compact, we select a subsequence of  $\{x_k^*\}$  (without relabeling) that converges weak\* to some  $x^* \in \mathcal{B}^*$ . By (2.86) there are  $x_{ik}^* \in \widehat{\mathcal{N}}(x_{ik}; \Omega_i)$  and  $b_{ik}^* \in \mathcal{B}^*$ ,  $i = 1, 2$ , such that

$$x_k^* = x_{1k}^* + \epsilon_k b_{1k}^*, \quad x_k^* = -x_{2k}^* + \epsilon_k b_{2k}^* \text{ for all } k \in \mathbb{N} . \quad (2.89)$$

This implies that  $x_{ik}^* \xrightarrow{w^*} x^*$  and  $x_{2k}^* \xrightarrow{w^*} -x^*$  as  $k \rightarrow \infty$ . The latter gives, due to definition (2.66), that  $x^*$  satisfies (2.88).

To justify (ii) in the sequential case, it remains to show that  $x^* \neq 0$  under the SNC assumption imposed. On the contrary, assume that  $x^* = 0$ , which gives  $x_{ik}^* \xrightarrow{w^*} 0$  for the sequences  $x_{ik}^* \in \widehat{\mathcal{N}}(x_{ik}; \Omega_i)$ ,  $i = 1, 2$ . Since one of the sets  $\Omega_i$  (say  $\Omega_1$ ) is  $\widehat{\mathcal{N}}$ -sequentially normally compact at  $\bar{x}$ , we get  $\|x_{1k}^*\| \rightarrow 0$ . This clearly implies that  $\|x_k^*\| \rightarrow 0$ , which contradicts the condition  $\|x_k^*\| = 1$  for all  $k \in \mathbb{N}$  and ends the proof of (ii) in the sequential case.

Let us finally consider the case of general Banach spaces and justify the topological version (2.87) of the exact extremal principle under the *sequential* normal compactness condition imposed. We follow the procedure in the sequential case but now don't assume anymore that  $\mathcal{B}^*$  is weak\* sequentially

compact, using instead the well-known fact that  $\mathcal{B}^*$  is (topologically) weak\* compact in arbitrary Banach spaces. This allows us to conclude that the above sequence  $\{x_k^*\}$  has a weak\* cluster point  $x^* \in \text{cl}^*\{x_k^* \mid k \in \mathbb{N}\} \cap \mathcal{B}^*$ . It follows from representation (2.89) with  $x_{ik}^* \in \widehat{\mathcal{N}}(x_{ik}; \Omega_i)$ ,  $i = 1, 2$ , and from definition (2.67) that  $x^*$  satisfies (2.87), where  $\overline{\mathcal{N}}$  is the topological normal structure generated by  $\widehat{\mathcal{N}}$ . This holds for any cluster point  $x^* \in \text{cl}^*\{x_k^* \mid k \in \mathbb{N}\}$ .

It remains to show that  $x^* \neq 0$  for some  $x^* \in \text{cl}^*\{x_k^* \mid k \in \mathbb{N}\}$  if one of the sets  $\Omega_i$ ,  $i = 1, 2$ , is  $\widehat{\mathcal{N}}$ -sequentially normally compact at  $\bar{x}$ . Indeed, the opposite means the  $x^* = 0$  is the only weak\* cluster point of  $\{x_k^*\}$ . The latter yields that the whole sequence  $\{x_k^*\}$  converges weak\* to zero. Then it follows from (2.89) that  $x_{ik}^* \xrightarrow{w^*} 0$ ,  $i = 1, 2$ , as  $k \rightarrow \infty$ . Hence  $\|x_{ik}^*\| \rightarrow 0$  for either  $i = 1$  or  $i = 2$ , which is impossible due to  $\|x_k^*\| = 1$ . This contradiction completes the proof of the theorem.  $\triangle$

As an immediate corollary of Theorem 2.51 we derive the following generalized versions of the Bishop-Phelps and supporting hyperplane theorems in terms of abstract prenormal and normal structures on Banach spaces.

**Corollary 2.52 (prenormal and normal structures at boundary points).** *Let  $\Omega$  be a proper closed subset of a Banach space  $X$ , and let  $\bar{x}$  be a boundary point of  $\Omega$ . Consider an arbitrary prenormal structure  $\widehat{\mathcal{N}}$  on  $X$  and the corresponding sequential normal structure  $\mathcal{N}$  and topological normal structure  $\overline{\mathcal{N}}$  generated by  $\widehat{\mathcal{N}}$ . Then one has:*

- (i) *Given any  $\varepsilon > 0$ , there is  $x \in \Omega \cap (\bar{x} + \varepsilon \mathcal{B})$  such that  $\widehat{\mathcal{N}}(x; \Omega) \neq \{0\}$ .*
- (ii) *Assume that the set  $\Omega$  is  $\widehat{\mathcal{N}}$ -sequentially normally compact at  $\bar{x}$ . Then  $\overline{\mathcal{N}}(\bar{x}; \Omega) \neq \{0\}$ . If in addition the dual ball  $\mathcal{B}^*$  is weak\* sequentially compact, then  $\mathcal{N}(\bar{x}; \Omega) \neq \{0\}$ .*

**Proof.** Follows from Theorem 2.51 with  $\Omega_1 := \Omega$  and  $\Omega_2 := \{\bar{x}\}$ .  $\triangle$

By the results of Subsect. 2.5.1 the abstract versions of the extremal principle in Theorem 2.51 and their corollaries hold for subdifferentially generated prenormal and normal structures under the mild requirements (S1)–(S3) on the corresponding presubdifferentials. These requirements are used in the proof of Lemma 2.32(ii) for the case of Fréchet normals and subgradients. As follows from the proof of the other statement (i) in Lemma 2.32, it holds for any presubdifferential  $\widehat{\mathcal{D}}\varphi(\cdot)$  on the class of proper l.s.c. functions  $\varphi: X \rightarrow \overline{\mathbb{R}}$  generated by a prenormal cone  $\widehat{\mathcal{N}}$  on  $X \times \mathbb{R}$  as

$$\widehat{\mathcal{D}}\varphi(x) := \{x^* \in X^* \mid (x^*, -1) \in \widehat{\mathcal{N}}((x, \varphi(x)); \text{epi } \varphi)\}, \quad x \in \text{dom } \varphi,$$

provided that  $\widehat{\mathcal{N}}(z; \Omega) \subset \{0\}$  if  $z \in \text{int } \Omega$  and that  $\|x^*\| \leq \ell$  for all  $x^* \in \widehat{\mathcal{D}}\varphi(x)$  if  $\varphi$  is locally Lipschitzian around  $x$  with modulus  $\ell$ . Thus both statements in Lemma 2.32 are valid for general classes of normals and subgradients. It is not the case for Theorem 2.33 and most of the other material in this chapter,

where the specific structure of Fréchet-like subdifferential constructions and geometric properties of Asplund spaces are essentially exploited. Note also that the structural properties of our basic constructions are utilized in Chap. 1 to build the generalized differential theory in Banach spaces.

In the subsequent chapters of this book we apply basic principles and results of the first two chapters to develop a comprehensive generalized differential calculus in Asplund spaces and give its applications to important problems in nonlinear analysis, optimization, and economics. Most of the results are formulated in terms of Fréchet-like normals/subgradients/coderivatives and their sequential limits, which is essential in the statements and proofs. As follows from the proofs (and will be explicitly mentioned in some cases), a part of the results obtained holds also for other normal and subgradient structures by the above discussions.

## 2.6 Commentary to Chap. 2

**2.6.1. The Origin of the Extremal Principle.** The chapter collects the fundamental material that is crucial for the subsequent parts of the book, in both aspects of basic theory and applications of variational analysis. Roughly speaking, all the essentials of variational analysis developed in this book largely revolve around the *extremal principle* comprehensively studied in Chap. 2. The extremal principle can be viewed as a local *variational* counterpart of the classical separation in the case of nonconvex sets; it actually plays the same role in variational analysis as separation theorems do in the presence of convexity, i.e., in the framework of convex analysis and its applications.

The term “extremal principle” was coined by Mordukhovich [910], while its first versions (in both approximate/fuzzy and exact/limiting forms of Definition 2.5) were established by Kruger and Mordukhovich [718] under the name of “generalized Euler equations” for local extremal points of finitely many sets in Fréchet smooth spaces. The essence of the exact extremal principle can be traced to the early paper by Mordukhovich [887], where the key method of *metric approximations* has been initiated in the framework of optimal control.

The properties of extremal systems and their connection with separation properties of convex and nonconvex sets presented in Subsect. 2.1.1 can be found in Kruger and Mordukhovich [719] and Mordukhovich [901]. The relationships between extremality and supporting properties from Subsect. 2.1.2 were fully investigated by Fabian and Mordukhovich [421]. To this end we mention a remarkable study of *boundary points* for sums of sets undertaken by Borwein and Jofré [148]. The latter boundary property of a set sum is actually equivalent to the local extremality of another set system; see also the recent paper by Kruger [715] for more details.

In Subsect. 2.1.3 we give a self-contained proof of the exact extremal principle in *finite-dimensional* spaces based on the *method of metric approximations*. As mentioned, this method was originated by Mordukhovich [887] and

then developed in [889, 892, 719, 901, 907] in several finite-dimensional settings; see also the comments below for its infinite-dimensional counterparts with significantly more involved variational arguments. Note that the method of metric approximations contains a *constructive procedure* to study local extremal points of set systems (in particular, local solutions to various problems of constrained optimization and equilibria) based on their *symmetric approximation* by sequences of *smooth* problems of *unconstrained minimization*. The realization of this procedure as in the proof of Theorem 2.8 has actually led us to constructing the *basic/limiting normal cone* in order to describe the (exact) *generalized Euler equation*. Observe that the latter appeared in the process of passing to the limit after applying the classical *Fermat stationary rule* in the sequence of approximating problems; cf. [887]. All this indicates close relationships between *classical and modern* tools and concepts of variational analysis: the novelty comes from applying appropriate *approximation/perturbation* techniques.

**2.6.2. The Extremal Principle in Fréchet Smooth Spaces and Separable Reduction.** Although there are no crucial differences between finite-dimensional and infinite-dimensional settings from conceptional viewpoints, *infinite-dimensional* extensions of the above approach to the extremal principle are technically much more involved requiring the usage of refined variational arguments and delicate geometric properties of Banach spaces. There are the following three most crucial features of *finite dimensionality* significantly exploited in the construction and realization of the *metric approximation* method employed to prove the exact extremal principle in Subsect. 2.1.3:

- (a) *intrinsic variational properties* of the *Euclidean norm*;
- (b) the *equivalence* of *any norm* in finite dimensions to the Euclidean norm, which is *smooth* away from the origin;
- (c) *compactness* of the closed unit ball (as well as the unit sphere), which is a *characterization* of finite-dimensional spaces.

Appropriate counterparts of these properties in infinite dimensions, which have nothing to do with the Euclidean norm, are among the key ingredients in deriving both approximate and exact versions of the extremal principle in the general framework of *Asplund spaces* presented in Sect. 2.2. To establish the *approximate* extremal principle in Asplund spaces, we develop a *two-step* procedure therein: first giving a direct proof of the extremal principle in Banach spaces admitting an equivalent *Fréchet smooth* norm (away from the origin), and then “rising up” the result from Fréchet smooth spaces to the general Asplund space setting by using the *method of separable reduction*.

The *variational arguments* employed in Subsect. 2.2.1 to justify the approximate extremal principle in Banach spaces with *smooth Fréchet renorms* were first developed, to the best of our knowledge, by Li and Shi [785] (preprint of



1990) in their proof of variational principles of the Ekeland and Borwein-Preiss types and then used, e.g., in [159, 265, 266, 688, 809] in parallel variational settings. We combine these arguments with the device in Mordukhovich and Shao [948] and with the subsequent induction. As mentioned in Remark 2.11, a similar device can be employed to establish the approximate extremal principle in Banach spaces admitting *smooth renorms of any kind*, with respect to natural bornologies. We refer the reader to the survey paper by Averbukh and Smolyanov [68] and to the book by Phelps [1073] for more information about bornologies. Appropriate versions of the approximate extremal principle in other (non-Fréchet) bornologically smooth spaces can be found in the paper by Borwein, Mordukhovich and Shao [151].

The method of *separable reduction* developed in Subsect. 2.2.2 in order to apply it to deriving the approximate extremal principle is probably the most difficult device given in this book. It is taken from the paper by Fabian and Mordukhovich [421], while its origin goes back to Preiss [1103] in the theory of Fréchet differentiability. Then versions of separable reduction were used by Fabian and Zhivkov [423], Fabian [413, 415], and Fabian and Mordukhovich [420, 421] in applications to various aspects of nonlinear analysis and generalized differentiability. It seems that the *Fréchet-type* differentiability and sub-differentiability is very essential in the theory and applications of this method.

**2.6.3. Asplund spaces.** The *Asplund property* of Banach spaces formulated in Subsect. 2.2.3 plays a crucial role in the theory and applications of variational analysis developed in this book. Although a number of important results and applications presented in the book hold in arbitrary Banach spaces, the most *comprehensive theory* of generalized differentiation, at the *same level of perfection* as in finite dimensions, is given in the Asplund space setting.

The remarkable class of Banach spaces, now called Asplund spaces, was introduced by Asplund in his 1968 paper [43] as “strong differentiability spaces.” The name “Asplund spaces” was coined by Namioka and Phelps [992] soon after Asplund’s death (1974). The original Asplund definition was the same one presented in Subsect. 2.2.3 with the only difference that the dense set of Fréchet differentiability points was postulated to be  $G_\delta$ . The latter requirement can be *equivalently omitted* due to the fact that Fréchet differentiability points always form a  $G_\delta$  set; see, e.g., Phelps [1073]. It is worth mentioning that, although the main contents of the original Asplund’s paper [43] concerned the geometric theory of Banach spaces, there were nice *variational* applications therein establishing *generic* existence and unique theorems for optimal solutions to some linearly *perturbed* variational problems particularly related to Moreau’s *proximal mappings* in Hilbert spaces [982].

Asplund spaces, which include all *reflexive* and many other remarkable Banach spaces, have been comprehensively investigated in the geometric theory of Banach spaces and its applications, with discovering a great number of impressive characterizations and properties; the reader may find a partial list

of them in the beginning of Subsect. 2.2.3 and in the references therein. Although the Asplund property is generally related to *Fréchet differentiability*, there are Asplund spaces that fail to have *even a Gâteaux smooth renorm*; see striking examples in Haydon [553] and in Deville, Godefroy and Zizler [331]. Note that, in contrast to the class of Asplund spaces that is one of the most beautiful objects in analysis and probably in all mathematics, *weak Asplund* spaces similarly defined in [43] with the replacement of Fréchet differentiability by Gâteaux differentiability are too far from being beautiful admitting only a modest number of satisfactory results; see the book by Fabian [416]. There is an intermediate class of *Asplund generated spaces*, known also in the literature as Grothendieck-Šmulian generated spaces, which particularly include all *weakly compactly generated* (hence all separable) spaces, strongly studied geometrically in the afore-mentioned Fabian's book. An on-going research project by Fabian, Loewen and Mordukhovich [418] is devoted to certain aspects of generalized differentiation and variational analysis in the framework of Asplund generated spaces; see Remark 3.103 for some results and discussions.

**2.6.4. The Extremal Principle in Asplund Spaces.** The extremal characterizations of Asplund spaces in Theorem 2.20 via the two (equivalent) versions of the *approximate extremal principle* were established by Mordukhovich and Shao [948], while the presented proof is taken from the later papers by Fabian and Mordukhovich: from [421] for the *sufficiency* of the Asplund property to ensure the extremal principle via separable reduction and from [420], via Example 2.19 reproduced in Subsect. 2.2.3, for the *necessity* of this property to have the extremal principle. Yet another proof (actually the first one) of the validity of the approximate extremal principle in general Asplund spaces can be found in Mordukhovich and Shao [949] via a *coderivative criterion* for the covering property established in their previous paper [946].

The *boundary characterizations* of Asplund spaces from Corollary 2.21 were obtained by Fabian and Mordukhovich [420] via separable reduction, with no appeal to the extremal principle. On the other hand, assertion (c) of this corollary, which is a far-going *nonconvex extension* of the celebrated *Bishop-Phelps theorem* [116] in the framework of Asplund spaces, was first deduced by Mordukhovich and Shao [948] from the extremal principle; cf. also Borwein and Strójas [156, 157] for other counterparts of the Bishop-Phelps theorem in nonconvex settings with other proofs. In the paper by Mordukhovich and B. Wang [960] the reader can find more *variational characterizations* of Asplund spaces via Fréchet normals and  $\varepsilon$ -normals, as well as different proofs of those mentioned above. Various subdifferential characterizations of Asplund spaces will be discussed below in the commentary to this chapter. We also refer the reader to the recent paper by Wang [1304] who derived some analogs of the afore-mentioned results and characterizations of the *reflexivity* of locally uniformly convex Banach spaces with Fréchet differentiable renorms via the *approximate extremal principle* involving *proximal* normals and subgradients.

The validity of the *exact extremal principle* in Asplund spaces under the *sequential normal compactness* conditions of Theorem 2.22 was established by Mordukhovich and Shao [949] extending the result of Kruger and Mordukhovich [718] obtained under the epi-Lipschitzian assumptions in Fréchet smooth spaces; see also the subsequent publications [707, 901]. The *converse* assertion of Theorem 2.22 was proved by Fabian and Mordukhovich [419]. Example 2.23 on the failure of the exact extremal principle in the absence of normal compactness is taken from Borwein and Zhu [162]. The nontriviality results on basic normals and subgradients from Corollaries 2.24 and 2.25, which immediately follow from the exact extremal principle, were first observed by Mordukhovich and Shao [949].

**2.6.5. The Ekeland Variational Principle.** According to the conventional terminology of modern nonlinear analysis, the expression “variational principle” stands for an assertion ensuring that, given a *lower semicontinuous* and *bounded from below* function  $\varphi$  and its arbitrary  $\varepsilon$ -minimal point  $x_0$ , there is a *small perturbation* of  $\varphi$  such that the perturbed function attains its *exact minimum* at some point close to  $x_0$ . The first variational principle in this sense was discovered by Ekeland in 1972 (see [396, 397, 399]) in general complete metric spaces. The exact statement of Ekeland’s variational principle is presented in Theorem 2.26(i). Note that the original Ekeland’s proof [396, 397] was rather complicated involving transfinite induction arguments via Zorn’s lemma. It was largely similar to the proof of the Bishop-Phelps theorem [116] mentioned above, which was called by Ekeland [399] “the grandfather of it all.” The much simplified proof presented in Theorem 2.26 follows the lines of Crandall’s arguments reproduced in Ekeland [399] as a personal communication. The *converse* statement of Theorem 2.26(ii) ensuring that the Ekeland principle is actually a characterization of the completeness property of metric spaces is due to Sullivan [1232]. There are so many applications of Ekeland’s variational principle to various areas in mathematics and related disciplines that it doesn’t seem to be possible of even mentioning a great part of them in this book. The reader can find a partial list of the most important early applications with their detailed analysis in the excellent survey by Ekeland [399] of 1979.

It is worth emphasizing that among the *main motivations* for the Ekeland original study was the result of Corollary 2.27, which ensures the fulfillment of the “almost stationary” condition for “almost optimal” (*suboptimal* in our terminology) solutions to a *smooth* unconstrained minimization problem. Results of this kind are especially important for optimization problems in infinite dimensions, where optimal solutions may often not exist. Thus the principal issue of both theoretical and practical importance is to derive necessary conditions for *suboptimal solutions*, of about the same type as for optimal solutions, that eventually lead to numerical algorithms for solving optimization problems. From this viewpoint, necessary suboptimality conditions applied to solutions that always exist are not worse than those for exact optimality,

which may not be reachable. We pay a strong attention to this topic throughout the book; see particularly Chaps. 5 and 6.

**2.6.6. Subdifferential Variational Principles.** The main result of Subsect. 2.3.2 called the *lower subdifferential variational principle* (Theorem 2.28) is a far-going development of Ekeland's  $\varepsilon$ -stationary condition in Corollary 2.27 from smooth functions to extended-real-valued l.s.c. functions; it can be applied therefore to problems of constrained optimization. This result established by Mordukhovich and B. Wang [962] is different from conventional variational principles in *only one* aspect: *instead of a perturbed minimization condition*, it contains a (lower) *subdifferential* condition of the  $\varepsilon$ -stationary type, which is actually a necessary condition for suboptimal solutions. The first result of this type for nonsmooth functions was obtained by Rockafellar [1147] via Clarke subgradients in Banach spaces, while for *convex* functions it actually goes back to the early work by Brøndsted and Rockafellar [179] that preceded Ekeland's variational principle; cf. also [154, 186, 501, 1165] for related results and discussions. As proved in the afore-mentioned paper [962], the subdifferential variational principle of Theorem 2.28 occurred to be an *equivalent analytic counterpart* of the *approximate extremal principle* giving hence yet another variational characterization of Asplund spaces.

The variational results of Theorem 2.28 easily imply the *subdifferential characterizations* of Asplund spaces listed in Corollary 2.29. These characterizations were first established via different devices by: Fabian [415] for (b), Fabian and Mordukhovich [419] for (c), and Fabian and Zhivkov [423] for (e); characterizations (d) follows from (e) due to Theorem 1.86. Note also that implication (e) $\Rightarrow$ (a) was proved earlier by Ioffe [593], while the related fact that the density of the set  $x \in \text{dom } \varphi$  with  $\widehat{\partial}_{\text{as}} \varphi(x) \neq \emptyset$  for any l.s.c. function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  yields the Asplund property of  $X$  goes back to Ekeland and Lebourg [400].

The *upper subdifferential variational principle* of Theorem 2.30 taken from the paper by Mordukhovich, Nam and Yen [938] is substantially different from the lower one being generally less powerful, since it applies only to special classes of functions that admit *upper* Fréchet subgradients at the points in question. However, for such classes of functions (which have been well recognized and investigated in nonsmooth analysis; see Chap. 5) the upper version involving *every* upper subgradient, has certain significant advantages in comparison with its lower counterpart from Theorem 2.28. It is particularly useful in developing necessary *suboptimality* conditions for various classes of constrained minimization problems; see Subsect. 5.1.4 for some results in this direction.

**2.6.7. Smooth Variational Principles.** Concerning the *conventional line* in developing variational principles, observe that the minimization condition in Ekeland's variational principle of Theorem 2.26 can be interpreted as follows: for every l.s.c. function  $\varphi: X \rightarrow \overline{\mathbb{R}}$  with  $\inf \varphi > -\infty$  there exists a

function  $s: X \rightarrow \overline{\mathbb{R}}$  that *supports  $\varphi$  from below* at some point  $\bar{x} \in \text{dom } \varphi$ , i.e.,

$$\varphi(\bar{x}) = s(\bar{x}) \text{ and } \varphi(x) \geq s(x) \text{ whenever } x \in X.$$

Then Ekeland's principle ensures, in the framework of arbitrary Banach spaces, that the support  $s(\cdot)$  can be chosen as a small perturbation by functions of the *norm* type. A clear disadvantage of this results is the intrinsic *nonsmoothness* of such perturbations, and so a natural question arises about conditions ensuring *smooth* perturbations, i.e., about *smooth variational principles*.

The first result of this type was obtained by Stegall in his 1978 paper [1224] who showed that, for any l.s.c. function satisfying some growth condition as  $\|x\| \rightarrow \infty$  on a Banach space with the *Radon-Nikodým property* (in particular, on a reflexive space), a supporting function  $s(\cdot)$  could be chosen as a *linear* functional with an arbitrarily small norm.

A more powerful smooth variational principle, in essentially more general settings, was established in the 1987 paper by Borwein and Preiss [154] who proved, assuming the existence of a *bornologically smooth renorm* on the Banach space in question, that supporting functions could be chosen as *concave and smooth* with respect to the same bornology. The Borwein-Preiss smooth variational principle was extended in some directions by Deville, Godefroy and Zizler [330, 331] who showed, in particular, that supporting functions could be chosen as bornologically *smooth* (but not concave anymore) under the more general assumption on the existence of a *smooth Lipschitzian bump* function with respect to some bornology. We refer the reader to [45, 70, 164, 265, 417, 419, 530, 531, 547, 619, 620, 785, 790, 809, 1243, 1356] among other publications for additional information about variational principles, their recent developments, and applications.

The results of Subsect. 2.3.3 are taken from the paper by Fabian and Mordukhovich [419]. Assertions (i) and (ii) of Theorem 2.31 establish *enhanced versions* of the Borwein-Preiss and Deville-Godefroy-Zizler smooth variational principles, respectively, with more *information* about supporting functions in comparison with the original versions in [154, 330]. Observe that the proof given in Theorem 2.31(i,ii) is essentially different from those of [154, 330]; it is based on the *lower subdifferential variational principle* from Theorem 2.28 and *smooth variational descriptions* of Fréchet subgradients from Theorem 1.88.

The *converse* assertion (iii) is *indeed remarkable*: it shows that the *smooth norm* and *smooth bump* assumptions in smooth variational principles of the Borwein-Preiss and Deville-Godefroy-Zizler types, respectively, are not only sufficient but also *necessary* for the validity of such results. As discussed at the end of Subsect. 2.3.3, the *Fréchet* smoothness is *not* essential for these conclusions, which hold true for *any bornology*. Observe again in this respect that *no smoothness* assumption is necessary for the fulfillment of the extremal principle and of the lower subdifferential variational principle. Furthermore, as proved in Borwein, Mordukhovich and Shao [151] (resp. in Mordukhovich [919]), the approximate extremal principle is *equivalent* to certain localized

versions of the Borwein-Preiss and Deville-Godefroy-Zizler variational principles *provided* that the Banach space in question admits a Fréchet smooth renorm (resp. a Fréchet smooth and Lipschitzian bump function).

**2.6.8. Limiting Normal and Subgradient Representations in Asplund Spaces.** It has been mentioned above that the main results of variational analysis and its applications developed in this book are derived from the extremal principle. Section 2.4 contains the first set of results in this direction showing, in particular, that the usage of the *approximate extremal principle* and its subgradient descriptions in Asplund spaces allows us to justify simplified and *convenient representations* of basic normals, subgradients, and coderivatives in the general Asplund setting similar to those established in finite dimensions on the base of specific properties of the Euclidean norm. The power of the extremal principal and its equivalents make it possible to replace the previous arguments without any appeal to either finite dimensionality, or to the Euclidean norm, or even to smooth renorming. Moreover, the Asplund space setting happens to be also *necessary* for such representations provided that they are required for *all* sets, functions, and set-valued mappings belonging to reasonably broad families.

The subdifferential description of the approximate extremal principle given in Lemma 2.32 plays a crucial role in establishing the main results of Sect. 4. This lemma was established by Mordukhovich and Shao [948], while the essence of assertion (i) can be traced to Ioffe [600]; cf. the proof of Step 2 in Lemma 2 therein.

Results of form (2.42) known as *fuzzy sum rules* (or “zero fuzzy sum rules,” or “fuzzy principles”) were initiated by Ioffe [593, 594] for  $\varepsilon$ -subdifferentials ( $\varepsilon > 0$ ) of both Fréchet and Dini types. For the case of Fréchet subgradients ( $\varepsilon = 0$ ) on Asplund spaces, the semi-Lipschitzian result (2.42) was first established by Fabian [415] based on the Borwein-Preiss smooth variational principle and on separable reduction; cf. Ioffe [599] for Fréchet smooth spaces. There are several modifications of such fuzzy rules; all of them happens to be equivalent. The latter was first proved by Zhu [1371] for the so-called  $\beta$ -subdifferentials that are valuable on bornologically smooth spaces and then by Ioffe [606] and Lassonde [747] in more general settings; see also the recent book by Borwein and Zhu [164].

The *full* (not “zero”) semi-Lipschitzian fuzzy sum rule of Theorem 2.33(b) was derived by Fabian first in [413] for  $\varepsilon > 0$  and then in [415] for  $\varepsilon = 0$  in the general Asplund space setting. Note that the structure of *Fréchet subgradients* seems to be very *essential* for this full fuzzy rule, in contrast to its zero counterpart (2.42). Some *topological* modifications of the full fuzzy sum rule (with a weak\* neighborhood of the origin in  $X^*$  instead of a small dual ball) were earlier considered by Ioffe [593] who introduced Banach spaces with such properties as “trustworthy spaces” and proved that any space admitting a Fréchet smooth bump function fell into the trustworthy category. Implication (b) $\Rightarrow$ (a) in Theorem 2.33 can be also deduced from [593]. We refer the reader

to the afore-mentioned publications and also to [147, 151, 158, 160, 163, 164, 257, 265, 329, 413, 414, 607, 614, 616, 622, 802, 952] for more results, equivalent statements, and discussions in this direction.

The *exact/limiting* semi-Lipschitzian sum rule of Theorem 2.33(c) as well as the representations of basic subgradients and normals from Theorems 2.34 and 2.35 in Asplund spaces were established by Mordukhovich and Shao [949], while the *converse* assertions therein are due to Fabian and Mordukhovich [419]. Extended sum rules based on the extremal principle are presented in Chap. 3, where the reader can find comprehensive calculus results with more discussions.

The *limiting  $\varepsilon$ -subdifferential*  $\partial_\varepsilon\varphi(\bar{x})$  in (2.48) for  $\varepsilon > 0$  was defined by Jofré, Luc and Théra [634] (preprint of 1995) motivated by applications to  $\varepsilon$ -monotonicity and related issues. As observed by Mordukhovich and Shao [949, Proposition 2.11], this construction happened to be an  $\varepsilon$ -*enlargement* of our *basic subdifferential* (see Theorem 2.34) for any l.s.c. function on Asplund spaces; moreover, such an enlargement representation of  $\partial_\varepsilon\varphi(\bar{x})$  *characterizes* the class of Asplund spaces as proved by Fabian and Mordukhovich [419].

The *singular subdifferential* limiting representation

$$\partial^\infty\varphi(\bar{x}) = \limsup_{\substack{x \xrightarrow{\varphi} \bar{x} \\ \lambda \downarrow 0}} \lambda \hat{\partial}\varphi(x) \quad (2.90)$$

from Theorem 2.38 was first obtained by Rockafellar [1150] in finite dimensions, with the *proximal subdifferential*  $\partial_P\varphi(x)$  of (2.81) replacing  $\hat{\partial}\varphi(x)$  in (2.90). The latter representation was actually accepted in [1150] as the definition of  $\partial^\infty\varphi(\bar{x})$ . Representation (2.90) was proved by Ioffe [600] for Fréchet smooth Banach spaces, and then the full statement of Theorem 2.38 in Asplund spaces was given by Mordukhovich and Shao [949] following the approach of [600]. The proof of the preceding Lemma 2.37 presented in the book is a clarification of Ioffe's proof in [600, Theorem 4] being different from it in several significant aspects.

Assertion (i) of Theorem 2.40 on *horizontal normals* to graphs and the inclusion

$$D^*\varphi(x)(0) \subset \partial^\infty\varphi(\bar{x}) \cup \partial^\infty(-\varphi)(\bar{x})$$

for continuous functions on Asplund spaces was established by Ngai and Théra [1008]. The *opposite* inclusion to the latter one and hence the equality in the *coderivative representation* of Theorem 2.40(ii) follow from Theorem 1.80. We refer the reader to the recent papers by Zhu [1373] and Ivanov [622] (see also the book by Borwein and Zhu [164]) for other proofs of the above results and their counterparts involving  $\beta$ -subdifferentials in *bornologically smooth* Banach spaces.

**2.6.9. Other Subdifferential Structures and Abstract Versions of the Extremal Principle.** Abstract normal and subdifferential structures of



Subsect. 2.5.1 were defined and studied by Mordukhovich [920] motivated by recognizing *minimal* normal and subdifferential properties needed for deriving the extremal principle in general Banach spaces. Various axiomatic constructions of this type, with generally different properties and applications, were considered by Aussel, Corvellec and Lassonde [61], Correa, Jofré and Thibault [292], Ioffe [599, 606, 607], Ioffe and Penot [614], Lassonde [747], Mordukhovich [901], Mordukhovich and Shao [949], Thibault and Zagrodny [1254], etc. The minimality result for the basic subdifferential from Proposition 2.45 was observed by Mordukhovich and Shao [949], while the essence of such theorems (under less general assumptions) should be traced to the early work by Ioffe [596, 599] and Mordukhovich [894, 901]; see more discussions in [949, Sect. 9]. Note that Ioffe's minimality result [599] *doesn't imply*, as mistakenly stated in [599, Proposition 8.2], that the nucleus  $\partial_G\varphi(\bar{x})$  of his  $G$ -subdifferential belongs to our basic subdifferential  $\partial\varphi(\bar{x})$  for l.s.c. functions on Fréchet smooth spaces. The point is that the mapping  $\partial\varphi(\cdot)$  may *not* be of *closed-graph* for Lipschitz continuous functions as claimed in [599]. In fact, the *opposite inclusion*

$$\partial\varphi(\bar{x}) \subset \widetilde{\partial}_G\varphi(\bar{x}) \quad (2.91)$$

is fulfilled for any l.s.c. function defined on an Asplund space, where equality holds for locally Lipschitzian functions provided that the space  $X$  is weakly compactly generated (and hence automatically Fréchet smooth); see Subsect. 3.2.3 below and comments to it in Subsect. 3.4.7. Moreover, it follows from examples by Borwein and Fitzpatrick [141] that the inclusion in (2.91) may be *strict* even for concave Lipschitz continuous functions defined on some special spaces admitting  $C^\infty$ -smooth renorms but not being weakly compactly generated; cf. Example 3.61 below.

Subsection 2.5.2 presents an overview of some remarkable normal and subdifferential structures important in the theory and applications of variational analysis via generalized differentiation. The main attention is paid to generalized normals and subgradients related to the basic constructions adopted in this book. The descriptions in Subsect. 2.5.2 are self-contained with the corresponding references to publications, where the reader can find more details and discussions; see also Commentary to Chap. 1. We just make some comments to (the last) part E of this subsection regarding the concepts and results formulated and proved therein.

The generalized differential construction  $\mathcal{A}\varphi(\bar{x})$  labeled here as the “derivate set” of  $\varphi$  at  $\bar{x}$  is inspired by Warga's *derivate containers* introduced in [1316] and then developed in many publications; see, e.g., [1317, 1318, 1319, 1320, 1321, 1370] and the more recent papers by Ermoliev, Norkin and Wets [408] and by Sussmann [1236, 1237, 1238] with the references and discussions therein. Theorem 2.46 in the form presented in this book was established by Kruger [713], while its essence and proof go back to the early work by Kruger and Mordukhovich [719] showing that the Fréchet subdifferential (and hence both lower and upper basic subdifferentials) is *smaller* than *any* Warga's



derivate container for continuous functions on finite-dimensional spaces; see also [99, 304, 596, 646, 705, 901] for modifications, extensions, and applications of the latter result and its variants.

Subsection 2.5.3 is based on the paper by Mordukhovich [920], where the approximate and exact versions of the *abstract extremal principle* were derived. Previous results on the fulfillment of the *approximate* extremal principle in non-Asplund (but mostly in bornologically smooth) spaces and on its equivalence to some other basic rules of generalized differentiation were obtained by Borwein, Mordukhovich and Shao [151], Borwein, Treiman and Zhu [159], Ioffe [606], and Zhu [1371]; see also Borwein and Zhu [163, 164] for more discussions.

Regarding the *exact* version of the abstract extremal principle, observe that both its *sequential* and *topological* modifications were established in [920] under an abstract version of the *sequential* normal compactness condition. A similar observation that just a sequential compactness property is sufficient to deal with a limiting topological structure was made by Ioffe [607] in the context of metric regularity.

Variational Analysis and Generalized Differentiation I

Basic Theory

Mordukhovich, B.S.

2006, XXII, 579 p., Hardcover

ISBN: 978-3-540-25437-9