Chapter 6
Nondifferentiable and Bilevel Optimization

It is not accidental that we unify the exposition of these two areas of optimization theory in one chapter. It has been widely recognized that problems of *nondifferentiable/nonsmooth optimization* (i.e., those containing nondifferentiable functions and/or sets with nonsmooth boundaries in their objectives and/or constraints) naturally and frequently appear in different aspects of variational analysis and numerous applications while being very challenging from both theoretical and algorithmic viewpoints. On the other hand, problems of *bilevel optimization* are *intrinsically nonsmooth*, even in the case of fully smooth data at their lower and upper levels. In fact, they can be reduced to single-level optimization problems, but the price to pay is the unavoidable presence of nonsmooth functions as a result of such reductions, regardless of smoothness assumptions imposed on the given data.

The main emphasis of this chapter is obtaining efficient first-order *necessary optimality conditions* for problems of *nondifferentiable programming* and then applying them to *bilevel programs* with smooth and nonsmooth functions at both levels of optimization. To proceed in these directions, we rely on the constructions and results of variational analysis and generalized differentiation developed in the previous chapters of the book.

6.1 Problems of Nondifferentiable Programming

We start with deriving necessary optimality conditions for problems of nonsmooth *minimization* with geometric constraints given by closed sets and then extend them to general problems of nondifferentiable programming with functional constraints described by finitely many inequalities and equalities.

B. S. Mordukhovich, *Variational Analysis and Applications*, Springer Monographs in Mathematics, https://doi.org/10.1007/978-3-319-92775-6_6

6.1.1 Lower and Upper Subdifferential Conditions

Given $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\Omega \subset \mathbb{R}^n$, consider the problem:

$$\text{minimize} \;\; \varphi(x) \;\; \text{subject to} \;\; x \in \Omega. \tag{6.1}$$

Our goal here is to obtain necessary conditions for (feasible) local minimizers $\bar{x} \in \operatorname{dom}\varphi \cap \Omega$ in (6.1). We derive two different types of necessary optimality conditions. Conditions of the first type, called the *lower subdifferential optimality conditions*, are expressed in terms of the basic subdifferential (1.24) under appropriate qualification conditions formulated in terms of the singular subdifferential (1.25). Conditions of the second type, called the *upper subdifferential optimality conditions*, make use of the upper regular subdifferential (1.76) of the cost function φ that is equivalently described as

$$\widehat{\partial}^+\varphi(\bar{x}) = -\widehat{\partial}(-\varphi)(\bar{x}), \quad |\varphi(\bar{x})| < \infty. \tag{6.2}$$

Note that (6.2) may be empty for broad classes of nonsmooth functions (e.g., for convex functions nondifferentiable at $\bar{x}$) while giving more selective necessary conditions for minimization than the lower subdifferential ones in certain "upper regular" settings; see the results, examples, and discussions below.

As before, we always assume without loss of generality that cost functions are *l.s.c.* around the reference points (although it is not needed for upper subdifferential conditions) and constraint sets are locally *closed* around them.

The following theorem contains necessary optimality conditions of both types for problem (6.1). Observe that both of them are derived from the *variational/extremal principles*. Indeed, the upper subdifferential conditions are induced by the smooth variational description of regular subgradients. To establish the lower subdifferential optimality conditions, we employ the basic subdifferential sum rule, which follows from the extremal principle. In fact, the extremal principle can be used directly; see, e.g., the proof of Theorem 6.5 below for problems involving functional and geometric constraints.

Theorem 6.1 (Optimality Conditions for Problems with a Single Geometric Constraint). *Let $\bar{x} \in \operatorname{dom}\varphi \cap \Omega$ be a local optimal solution to the minimization problem* (6.1). *The following assertions hold:*

(i) *The entire set of upper regular subgradients satisfies the inclusions*

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset \widehat{N}(\bar{x};\Omega), \quad -\widehat{\partial}^+\varphi(\bar{x}) \subset N(\bar{x};\Omega). \tag{6.3}$$

(ii) *Under the qualification condition*

$$\partial^\infty\varphi(\bar{x}) \cap \big(-N(\bar{x};\Omega)\big) = \{0\} \tag{6.4}$$

, we have the lower subdifferential relationships

$$\partial\varphi(\bar{x}) \cap \big(-N(\bar{x};\Omega)\big) \neq \emptyset, \;\; \textit{i.e.,} \;\; 0 \in \partial\varphi(\bar{x}) + N(\bar{x};\Omega). \tag{6.5}$$

Proof. To justify assertion (i), it suffices to verify only the first inclusion in (6.3) since $\widehat{N}(\bar{x};\Omega) \subset N(\bar{x};\Omega)$ by Theorem 1.6. To proceed with this task, suppose that $\widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset$ (there is nothing to prove otherwise), and pick any $v \in \widehat{\partial}^+\varphi(\bar{x})$. Using (6.2) and applying the first part of Theorem 1.27 (which holds without the l.s.c. assumption on φ), we find a function $\psi\colon \mathbb{R}^n \to \mathbb{R}$ with $\psi(\bar{x}) = \varphi(\bar{x})$ and $\psi(x) \geq \varphi(x)$ whenever $x \in \mathbb{R}^n$ such that ψ is (Fréchet) differentiable at $\bar{x}$ and $\nabla\psi(\bar{x}) = v$. It gives us

$$\psi(\bar{x}) = \varphi(\bar{x}) \leq \varphi(x) \leq \psi(x) \ \text{ for all } \ x \in \Omega \ \text{ close to } \ \bar{x}$$

showing therefore that $\bar{x}$ is a local minimizer of the constrained problem:

$$\text{minimize } \ \psi(x) \ \text{ subject to } \ x \in \Omega,$$

where the cost function is differentiable at $\bar{x}$. This problem can be equivalently written in the form of unconstrained optimization:

$$\text{minimize } \ \psi(x) + \delta(x;\Omega), \quad x \in \mathbb{R}^n.$$

Employing in the latter setting the generalized Fermat rule from Proposition 1.30(i) and then the regular subdifferential sum rule from Proposition 1.30(ii) with taking into account that $\nabla\psi(\bar{x}) = v$, we get

$$0 \in \widehat{\partial}\big(\psi + \delta(\cdot;\Omega)\big)(\bar{x}) = \nabla\psi(\bar{x}) + \widehat{N}(\bar{x};\Omega) = v + \widehat{N}(\bar{x};\Omega).$$

This yields $-v \in \widehat{N}(\bar{x};\Omega)$ for any $v \in \widehat{\partial}^+\varphi(\bar{x})$ and thus verifies (i).

To prove assertion (ii), we apply the generalized Fermat rule to the local optimal solution $\bar{x}$ of problem (6.1) written in the unconstrained form:

$$\text{minimize } \ \varphi(x) + \delta(x;\Omega), \quad x \in \mathbb{R}^n,$$

and then deduce from the basic subdifferential sum rule of Theorem 2.19 that

$$0 \in \partial\big(\varphi + \delta(\cdot;\Omega)\big)(\bar{x}) \subset \partial\varphi(\bar{x}) + N(\bar{x};\Omega)$$

provided the validity of the qualification condition (6.4) due to Proposition 1.19. This verifies (6.5) and completes the proof of the theorem. △

Let us discuss some particular features of the lower and upper subdifferential conditions from Theorem 6.1 and relationships between them.

Remark 6.2 (Upper vs. Lower Subdifferential Optimality Conditions).

(i) Note first that in the case where φ is (Fréchet) differentiable at $\bar{x}$, the optimality conditions in (6.3) reduce to

$$-\nabla\varphi(\bar{x}) \in \widehat{N}(\bar{x};\Omega), \quad -\nabla\varphi(\bar{x}) \in N(\bar{x};\Omega),$$

while only the second inclusion can be derived from (6.5) provided that φ is *strictly* differentiable at $\bar{x}$. On the other hand, the upper subdifferential conditions in (6.3) are trivial when $\widehat{\partial}^+\varphi(\bar{x}) = \emptyset$, which is the case of, e.g., convex continuous functions nondifferentiable at $\bar{x}$. In contrast, the lower subdifferential condition (6.5) is nontrivial for broad collections of nonsmooth functions including, e.g., every locally Lipschitzian function φ for which $\partial\varphi(\bar{x}) \neq \emptyset$ and the qualification condition (6.4) holds due to $\partial^\infty\varphi(\bar{x}) = \{0\}$ by Theorem 1.22.

(ii) Note also that the *triviality condition* $\widehat{\partial}^+\varphi(\bar{x}) = \emptyset$ itself is an easy checkable *necessary condition for optimality* in (6.1) provided that φ is nondifferentiable at $\bar{x}$ and $\Omega = \mathbb{R}^n$. Indeed, in this case, we have the inclusion $0 \in \widehat{\partial}\varphi(\bar{x}) \neq \emptyset$ by the generalized Fermat rule and hence $\widehat{\partial}^+\varphi(\bar{x}) = \emptyset$ by the simple observation from Exercise 1.76(ii).

(iii) Recall that φ is *upper regular* at $\bar{x}$ if $\widehat{\partial}^+\varphi(\bar{x}) = \partial^+\varphi(\bar{x})$. Note that, besides concave functions and differentiable ones, this class includes, e.g., a rather large class of *semiconcave* functions important in various applications to optimization and control; see, e.g., [136, 523]. If φ is upper regular at $\bar{x}$ and locally Lipschitzian around this point, we have $\widehat{\partial}^+\varphi(\bar{x}) = -\partial(-\varphi)(\bar{x}) \neq \emptyset$ by Theorem 1.22, i.e., the upper subdifferential conditions in (6.3) definitely give us a nontrivial information. Furthermore, in this case, we also have $\overline{\partial}\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x})$ for Clarke's generalized gradient due to its plus-minus symmetry (1.79). Taking into account that the inclusions in (6.3) are valid for the *entire set* of upper subgradients, these observations show that the upper subdifferential optimality conditions may have sizable advantages over the lower subdifferential ones from Theorem 6.1(ii).

(iv) Let us consider in more detail problems of *concave minimization*, i.e., when the cost function φ is concave in (6.1). This class is of significant interest for various aspects of optimization theory and applications; in particular, from the viewpoints of *global* optimization; see, e.g., [355]. When φ is concave and continuous around $\bar{x}$, it follows from Exercise 1.77 that

$$\partial\varphi(\bar{x}) \subset \partial^+\varphi(\bar{x}) = \widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset.$$

Then comparing the second inclusion in (6.3) (which is even weaker than the first inclusion therein) with the lower subdifferential condition in (6.5), we see that the necessary condition of Theorem 6.1(i) requires that *every* element v of the set $\widehat{\partial}^+\varphi(\bar{x})$ must belong to $-N(\bar{x}; \Omega)$, instead of that *some* element v from the *smaller* set $\partial\varphi(\bar{x})$ belongs to $-N(\bar{x}; \Omega)$ in Theorem 6.1(ii). Let us illustrate it by the following simple *example*:

$$\text{minimize } \ \varphi(x) := -|x| \ \text{ subject to } \ x \in \Omega := [-1, 0] \subset \mathbb{R}.$$

Obviously $\bar{x} = 0$ is not an optimal solution to this problem. However, it cannot be taken away by the lower subdifferential condition (6.5) due to

$$\partial\varphi(0) = \{-1, 1\}, \quad N(0; \Omega) = [0, \infty), \quad \text{and} \quad -1 \in -N(0; \Omega).$$

On the other hand, checking the upper subdifferential condition (6.3) gives us

$$\widehat{\partial}^+\varphi(0) = [-1, 1] \quad \text{and} \quad [-1, 1] \not\subset N(0; \Omega),$$

which confirms that $\bar{x} = 0$ is not optimal in (6.1), and thus (6.3) is a more selective necessary condition for optimality in the problem under consideration.

Observe further that minimization problems for *differences of two convex* (*DC*) *functions* can be equivalently reduced to minimizing concave functions subject to convex constraints. This allows us to deduce necessary conditions for such problems from the upper subdifferential conditions of Theorem 6.1(i).

Proposition 6.3 (DC Optimization Problems). *Consider the problem:*

$$\text{minimize } \varphi_1(x) - \varphi_2(x), \quad x \in \mathbb{R}^n, \tag{6.6}$$

where $\varphi_1, \varphi_2 \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ are convex under the convention that $\infty - \infty := \infty$. Then $\bar{x}$ is a local minimizer of (6.6) *if and only if the pair $(\bar{x}, \varphi_1(\bar{x}))$ gives a local minimum to the following problem on minimizing a concave function subject to convex geometric constraints:*

$$\text{minimize } \psi(x, \alpha) := \alpha - \varphi_2(x) \text{ subject to } (x, \alpha) \in \operatorname{epi} \varphi_1. \tag{6.7}$$

Moreover, the upper subdifferential condition (6.3) *for* (6.7) *reduces to the* (*lower*) *subdifferential inclusion $\partial\varphi_2(\bar{x}) \subset \partial\varphi_1(\bar{x})$.*

Proof. If $\bar{x}$ solves (6.6) locally, i.e., there is a neighborhood U of $\bar{x}$ such that

$$\varphi_1(\bar{x}) - \varphi_2(\bar{x}) \le \varphi_1(x) - \varphi_2(x) \text{ for all } x \in U,$$

then for $\bar{\alpha} := \varphi_1(\bar{x})$, we obviously have

$$\bar{\alpha} - \varphi_2(\bar{x}) \le \alpha - \varphi_2(x) \text{ whenever } (x, \alpha) \in (U \times \mathbb{R}) \cap \operatorname{epi} \varphi_1,$$

which means that $(\bar{x}, \bar{\alpha})$ locally solves problem (6.7). Conversely, suppose that there exist $\varepsilon > 0$ and a neighborhood U of $\bar{x}$ such that

$$\varphi_1(\bar{x}) - \varphi_2(\bar{x}) \le \alpha - \varphi_2(x) \text{ for all } \alpha \ge \varphi_1(x),\ x \in U,\ |\alpha - \varphi_1(\bar{x})| < \varepsilon.$$

Since φ_1 is convex and finite around $\bar{x}$ by the above, it is (Lipschitz) continuous around this point. Thus there is a neighborhood $\widetilde{U}$ of $\bar{x}$ on which

$$|\varphi_1(x) - \varphi_1(\bar{x})| < \varepsilon, \text{ and so } \varphi_1(\bar{x}) - \varphi_2(\bar{x}) \le \varphi_1(x) - \varphi_2(x), \quad x \in \widetilde{U}.$$

This verifies that $\bar{x}$ is a local solution to (6.6).

It remains to show that the upper subdifferential optimality condition

$$-\widehat{\partial}^+\psi\big(\bar{x}, \varphi_1(\bar{x})\big) \subset N\big((\bar{x}, \varphi_1(\bar{x})); \operatorname{epi} \varphi_1\big) \tag{6.8}$$

for (6.7) reduces to the subdifferential inclusion claimed in the proposition. Indeed, we get by the direct calculations that

$$-\widehat{\partial}^+\psi\big(\bar{x},\varphi_1(\bar{x})\big)=\widehat{\partial}\big(\varphi_2-\alpha\big)\big(\bar{x},\varphi_1(\bar{x})\big)=\partial\varphi_2(\bar{x})\times\{0\}+\{0\}\times\{-1\}$$
$$=\partial\varphi_2(\bar{x})\times\{-1\}.$$

Hence the upper subdifferential inclusion (6.8) implies that

$$(v,-1)\in N\big((\bar{x},\varphi_1(\bar{x}));\operatorname{epi}\varphi_1\big)\ \text{ for all }\ v\in\partial\varphi_2(\bar{x}),$$

which is equivalent to $v\in\partial\varphi_1(\bar{x})$ for all $v\in\partial\varphi_2(\bar{x})$ and thus justifies the claimed necessary optimality condition $\partial\varphi_2(\bar{x})\subset\partial\varphi_1(\bar{x})$ in (6.6). △

The crucial advantage of the second upper subdifferential inclusion in (6.3) in comparison with the first one and also a strong feature of the lower subdifferential qualification and optimality conditions are well-developed *calculus rules* available for basic normals and subgradients in contrast to their regular counterparts. In particular, calculus results obtained in Chapter 2 allow us to derive various consequences of both assertions (i) and (ii) of Theorem 6.1 in cases where Ω is represented as a product and a sum of finitely many sets, as an inverse image of another set under a set-valued mapping, as a system of inequalities and/or equalities, etc. Qualification conditions that ensure the validity of the obtained representations of $N(\bar{x};\Omega)$ are transferred in this way into *constraint qualifications* under which the corresponding necessary optimality conditions hold in the *qualified/normal/KKT* (*Karush-Kuhn-Tucker*) *form*, i.e., with no (=1) multiplier associated with the cost function; see below.

Next we present both upper and lower subdifferential optimality conditions obtained in this scheme for problems with finitely many geometric constraints.

Proposition 6.4 (Optimality Conditions for Problems with Many Geometric Constraints). *Consider the problem:*

$$\textit{minimize}\ \ \varphi(x)\ \ \textit{subject to}\ \ x\in\Omega_i\ \ \textit{for}\ \ i=1,\ldots,s, \tag{6.9}$$

and suppose that $\bar{x}\in\operatorname{dom}\varphi\cap\Omega_1\cap\ldots\cap\Omega_s$ *is a local minimizer for* (6.9). *Then the following upper subdifferential and lower subdifferential necessary optimality conditions hold at* $\bar{x}$*:*

(i) *Under the validity of the constraint qualification*

$$\big[v_1+\ldots+v_s=0,\ v_i\in N(\bar{x};\Omega_i)\big]\Longrightarrow v_i=0\ \textit{for all}\ \ i=1,\ldots,s, \tag{6.10}$$

we have the upper subdifferential inclusion

$$-\widehat{\partial}^+\varphi(\bar{x})\subset N(\bar{x};\Omega_1)+\ldots+N(\bar{x};\Omega_s).$$

(ii) *Under the validity of the qualification condition*

$$\Big[v+\sum_{i=1}^{s} v_i=0 \;\textit{for}\; v\in\partial^{\infty}\varphi(\bar{x}),\; v_i\in N(\bar{x};\Omega_i)\Big]\Longrightarrow v=v_1=\ldots=v_s=0$$

stronger than (6.10)*, we have the lower subdifferential inclusion*

$$0\in\partial\varphi(\bar{x})+N(\bar{x};\Omega_1)+\ldots+N(\bar{x};\Omega_s).$$

Proof. Necessary optimality conditions in both assertions (i) and (ii) follow directly from the corresponding results of Theorem 6.1 and the normal intersection rule for finitely many sets given in Corollary 2.17. △

6.1.2 Finitely Many Inequality and Equality Constraints

Let us consider here the problem of *nondifferentiable programming*:

$$\begin{cases}\text{minimize }\; \varphi_0(x)\; \text{ subject to}\\ \varphi_i(x)\le 0, \quad i=1,\ldots,m,\\ \varphi_i(x)=0, \quad i=m+1,\ldots,m+r,\\ x\in\Omega\subset\mathbb{R}^n\end{cases} \tag{6.11}$$

with finitely many inequality and equality constraints while keeping geometric constraints as well. In what follows we derive various necessary optimality conditions of both lower subdifferential and upper subdifferential types for local solutions to program (6.11) depending on assumptions imposed on their initial data and proof techniques. Our first theorem presents general necessary optimality conditions of the lower subdifferential type expressed via normals and subgradients of each function and set in (6.11) *separately*. The proof is based on the direct application of the *extremal principle* from Theorem 2.3. Recall that, unless otherwise stated, all the functions in question are assumed to be *lower semicontinuous* around the reference points.

Theorem 6.5 (Lower Subdifferential Conditions via Normals and Subgradients of Separate Constraints). *Let $\bar{x}$ be a feasible solution to* (6.11)*, that is, a local minimizer for this problem. The following necessary optimality conditions hold at $\bar{x}$:*

(i) *Assume that the equality constraint functions φ_i are continuous around $\bar{x}$ for all $i=m+1,\ldots,m+r$. Then there are elements $(v_i,\lambda_i)\in\mathbb{R}^{n+1}$ for $i=0,\ldots,m+r$, not equal to zero simultaneously, and a vector $v\in\mathbb{R}^n$ such that $\lambda_i\ge 0$ for $i=0,\ldots,m$ and*

$$(v_0,-\lambda_0)\in N\big((\bar{x},\varphi_0(\bar{x}));\operatorname{epi}\varphi_i\big), \quad v\in N(\bar{x};\Omega), \tag{6.12}$$

$$(v_i,-\lambda_i)\in N\big((\bar{x},0);\operatorname{epi}\varphi_i\big), \quad i=1,\ldots,m, \tag{6.13}$$

$$(v_i, -\lambda_i) \in N\big((\bar{x}, 0); \operatorname{gph} \varphi_i\big), \quad i = m+1, \ldots, m+r, \tag{6.14}$$

$$v + \sum_{i=0}^{m+r} v_i = 0. \tag{6.15}$$

If in addition the function φ_i is u.s.c. at $\bar{x}$ for some $i \in \{1, \ldots, m\}$ with $\varphi_i(\bar{x}) < 0$, then $\lambda_i = 0$. If this happens for all $i = 1, \ldots, m$, then we have the complementary slackness conditions for the inequality constraints

$$\lambda_i \varphi_i(\bar{x}) = 0, \quad i = 1, \ldots, m. \tag{6.16}$$

(ii) *Assume that the functions φ_i are Lipschitz continuous around $\bar{x}$ for all $i = 0, \ldots, m+r$. Then there are multipliers $(\lambda_0, \ldots, \lambda_{m+r}) \neq 0$ such that*

$$0 \in \sum_{i=0}^{m} \lambda_i \partial \varphi_i(\bar{x}) + \sum_{i=m+1}^{m+r} \lambda_i \Big[\partial \varphi_i(\bar{x}) \cup \partial(-\varphi_i)(\bar{x})\Big] + N(\bar{x}; \Omega), \tag{6.17}$$

$$\lambda_i \geq 0, \ i = 0, \ldots, m+r, \ \text{ and } \ \lambda_i \varphi_i(\bar{x}) = 0, \ i = 1, \ldots, m. \tag{6.18}$$

Proof. To justify (i), assume without loss of generality that $\varphi_0(\bar{x}) = 0$. Then it is easy to check that $(\bar{x}, 0)$ is a *locally extremal point* of the following system of locally closed sets in the product space $\mathbb{R}^n \times \mathbb{R}^{m+r+1}$:

$$\Omega_i := \Big\{(x, \alpha_0, \ldots, \alpha_{m+r}) \Big| \ \alpha_i \geq \varphi_i(x)\Big\}, \quad i = 0, \ldots, m,$$

$$\Omega_i := \Big\{(x, \alpha_0, \ldots, \alpha_{m+r}) \Big| \ \alpha_i = \varphi_i(x)\Big\}, \quad i = m+1, \ldots, m+r,$$

$$\Omega_{m+r+1} := \Omega \times \{0\}.$$

Applying the extremal principle of Theorem 2.3 immediately gives us the relationships in (6.12)–(6.15). It follows from Proposition 1.17 that $\lambda_i \geq 0$ for $i = 0, \ldots, m$. To finish the proof of (i), it remains to show that the *complementary slackness* conditions in (6.16) hold for each $i \in \{1, \ldots, m\}$ with $\varphi_i(\bar{x}) < 0$ provided that φ_i is u.s.c. at $\bar{x}$. Indeed, we get from this assumption that $\varphi_i(x) < 0$ for all x around $\bar{x}$, and so $(\bar{x}, 0)$ is an *interior point* of the epigraph of φ_i. Thus $N((\bar{x}, 0); \operatorname{epi} \varphi_i) = \{0\}$ and $(v_i, \lambda_i) = (0, 0)$ for such i.

Assertion (ii) easily follows from (i) due to Theorem 1.22, which shows that the normal cone to the epigraph of a locally Lipschitzian function φ_i is fully determined by the (basic) subdifferential of φ_i. In the case of $\operatorname{gph} \varphi_i$ for the equality constraints, we deal with the epigraph of either φ_i or $-\varphi_i$ scaled by the corresponding *nonnegative* multiplier λ_i due to Proposition 1.17. △

The necessary optimality conditions of Theorem 6.5 are given in the *nonqualified/Fritz John* form, which doesn't ensure that $\lambda_0 \neq 0$ for the multiplier associated with the cost function. However, it is not hard to deduce from them (or from the qualification conditions in the calculus rules employed in the proofs) appropriate *constraint qualifications* of the generalized Mangasarian-Fromovitz and other types, which yield $\lambda_0 = 1$; see, e.g., [523, Chapter 5] with the commentaries and references therein as well as the exercises in Section 6.4.

Observe that for standard nonlinear programs (6.11) with smooth functions φ_i and $\Omega = \mathbb{R}^n$, the necessary optimality conditions of Theorem 6.5(ii) agree with the classical *Lagrange multiplier rule*. However, it is not the case for problems with nonsmooth *equality constraints*. Indeed, in the latter case, the result obtained in Theorem 6.5(ii) involves *nonnegative multipliers* λ_i associated with the the unions $\partial\varphi_i(\bar{x}) \cup \partial(-\varphi_i)(\bar{x})$ as $i = m+1,\ldots,m+r$, which are $\{\nabla\varphi_i(\bar{x}), -\nabla\varphi_i(\bar{x})\}$ for smooth functions. It is not hard to deduce from (6.17) and (6.18) a more conventional form of the generalized Lagrange multiplier rule with no sign condition for the equality multipliers, but in this way we arrive at a weaker necessary optimality condition as shown in Example 6.7 below. To proceed, recall the two-sided version of the basic subdifferential

$$\partial^0\varphi(\bar{x}) = \partial\varphi(x) \cup \partial^+\varphi(\bar{x}),$$

which is the *symmetric subdifferential* (1.75) already used in the book.

Corollary 6.6 (Equality Constraints via Symmetric Subgradients). *Let $\bar{x}$ be a local minimizer of* (6.11) *under the assumptions of Theorem* 6.5(ii). *Then there exists a nonzero collection of multipliers $(\lambda_0,\ldots,\lambda_{m+r}) \in \mathbb{R}^{m+r+1}$ satisfying the sign conditions $\lambda_i \ge 0$ for $i = 0,\ldots,m$, the complementary slackness condition* (6.16), *and the symmetric Lagrangian inclusion*

$$0 \in \sum_{i=0}^{m} \lambda_i\partial\varphi_i(\bar{x}) + \sum_{i=m+1}^{m+r} \lambda_i\partial^0\varphi_i(\bar{x}) + N(\bar{x};\Omega). \tag{6.19}$$

Proof. Follows directly from Theorem 6.5(ii) due to the (proper) inclusion

$$|\lambda|\big[\partial\varphi(\bar{x}) \cup \partial(-\varphi)(\bar{x})\big] \subset \lambda\big[\partial^0\varphi(\bar{x}) \cup \big(-\partial^0\varphi(\bar{x})\big)\big], \quad \lambda \in \mathbb{R},$$

applied to the functions φ_i, $i = m+1,\ldots,m+r$, in (6.17). △

6.1.3 Examples and Discussions on Optimality Conditions

Now we present several examples illustrating the difference between the obtained versions of the generalized Lagrange multiplier rule and compare them with other major versions known in nonsmooth optimization.

Example 6.7 (Nonnegative Sign vs. Symmetric Lagrangian Inclusions). As shown above, inclusion (6.17) with all the nonnegative multipliers always implies the symmetric one (6.19) with $\lambda_i \in \mathbb{R}$ as $i = m+1, \ldots, m+r$. The following two-dimensional problem with a single equality constraint confirms that the converse implication doesn't hold. Consider the problem:

$$\text{minimize } x_1 \text{ subject to } \varphi_1(x_1, x_2) := \varphi(x_1, x_2) + x_1 = 0, \tag{6.20}$$

where φ is taken from Example 1.31(ii). It follows from the subdifferential calculation therein that the set $\partial\varphi_1(0,0) \cup \partial(-\varphi_1)(0,0)$ in (6.17) is

$$\big\{(v_1, v_2) \in \mathbb{R}^2 \big| \, |v_1 - 1| \le v_2 \le 1\big\} \cup \big\{(v_1, -|v_1 - 1|) \big| 0 \le v_1 \le 2\big\}$$
$$\cup\{(v_1, 1)\big| -2 \le v_1 \le 0\big\} \cup \big\{(-2, -1)\big\}$$

as depicted on Fig. 6.1(a). The symmetric subdifferential of φ_1 is

$$\partial^0\varphi_1(0,0) = \partial\varphi(0,0) \cup \big\{(v, -1)\big| -1 \le v \le 1\big\} + (1, 0)$$

with $\partial\varphi(0,0)$ calculated in Example 1.31(ii); see Fig. 6.1(b). It is now easy to check that the nonnegative sign inclusion (6.17) allows us to exclude the feasible solution $\bar{x} = (0,0)$ from the candidates for optimality, while the symmetric one (6.19) is satisfied at the nonoptimal point $\bar{x}$.

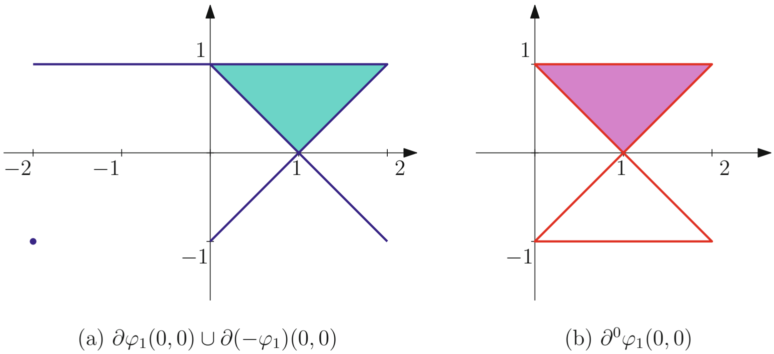

Fig. 6.1 Subdifferentials of $\varphi_1(x_1, x_2) = \big||x_1| + x_2\big| + x_1$ at $(0,0)$.

Example 6.8 (Comparison with the Convexified/Clarke Version of the Lagrange Multiplier Rule). Clarke's version [164, 165] of the Lagrange multiplier rule for nondifferentiable programming (6.11) with Lipschitzian data is given in the form of Corollary 6.6 where the nonconvex subdifferentials $\partial\varphi_i(\bar{x})$ for $i = 0, \ldots, m$ and $\partial^0\varphi_i(\bar{x})$ for $i = m+1, \ldots, m+r$, as well as the normal cone $N(\bar{x}; \Omega)$, are replaced by their convexified counterparts:

$$0 \in \sum_{i=0}^{m+r} \lambda_i \overline{\partial}\varphi_i(\bar{x}) + \overline{N}(\bar{x}; \Omega). \tag{6.21}$$

This version is obviously weaker than (6.6) and doesn't allow us to exclude the nonoptimal solution $\bar{x}$ in problem (6.20) of the preceding Example 6.7. Moreover, Clarke's version (6.21) fails to recognize nonoptimal solutions even in much less sophisticated examples from unconstrained nonsmooth optimization and also for problems with only inequality constraints. One of the reasons for this is that, due to the plus-minus symmetry of $\overline{\partial}\varphi$, condition (6.21) does *not* distinguish between *minima* and *maxima* and also between *inequality* constraints of the "$\le$" and "$\ge$" types. It makes an easy task to construct examples for which (6.21) is satisfied at clearly nonoptimal points.

(i) First consider the simplest *unconstrained* minimization problem:

$$\text{minimize } \ \varphi(x) := -|x| \ \text{ over all } \ x \in \mathbb{R},$$

where $\bar{x} = 0$ is a point of maximum, not minimum. Nevertheless, we have $0 \in \overline{\partial}\varphi(0) = [-1, 1]$ while $0 \notin \partial\varphi(0) = \{-1, 1\}$.

(ii) The second example in this direction concerns the following two-dimensional problem with a single nonsmooth *inequality constraint*:

$$\text{minimize } \ x_1 \ \text{ subject to } \ \varphi(x_1, x_2) := |x_1| - |x_2| \le 0.$$

We have here $\partial\varphi(0,0) = \{(v_1, v_2)| -1 \le v_1 \le 1, v_2 = 1, \text{ or } v_2 = -1\}$ by Example 1.31(i), and hence the point $\bar{x} = (0, 0)$ is ruled out from optimality by Corollary 6.6, while the usage of the generalized gradient $\overline{\partial}\varphi(0,0) = \{(v_1, v_2)| -1 \le v_1 \le 1, -1 \le v_2 \le 1\}$ doesn't allow us to do it by (6.21).

Example 6.9 (Comparison with Warga's Version of the Lagrange Multiplier Rule). Another extension of the Lagrange multiplier rule to problems of nondifferentiable programming (6.11) with $\Omega = \mathbb{R}^n$ and Lipschitzian functions φ_i was obtained by Warga [736, 737] in terms of his *derivate containers* $\Lambda^0\varphi_i(\bar{x})$ in the form of Corollary 6.6 with the Lagrangian inclusion

$$0 \in \sum_{i=0}^{m+r} \lambda_i \Lambda^0 \partial\varphi_i(\bar{x}). \tag{6.22}$$

Note that the set $\Lambda^0\varphi(\bar{x})$ is generally nonconvex, possesses the classical plus-minus symmetry, and may be strictly smaller than Clarke's generalized gradient $\overline{\partial}\varphi(\bar{x})$. As shown in [522, Corollary 2.48], we always have $\partial^0\varphi(\bar{x}) \subset \Lambda^0\varphi(\bar{x})$. Hence the necessary optimality conditions of Theorem 6.5(ii) and Corollary 6.6 definitely yield the result of (6.22). Let us illustrate that the improvement is *strict* in both cases of equality and inequality constraints.

(i) For the case of only *equality* constraints in (6.11), the claimed strict inclusion follows from Example 6.7 with the constraint function φ_1 defined in (6.20). Indeed, condition (6.22) is satisfied at the nonoptimal point $\bar{x} = (0, 0)$, while (6.19) confirms its nonoptimality. Recall that the derivative container $\Lambda^0\varphi(\bar{x})$ for the function φ in this example is depicted on Fig. 1.13.

(ii) To demonstrate the advantage of (6.17) for nondifferentiable programs with *inequality* constraints, consider the problem

$$\text{minimize}\ \ x_2\ \ \text{subject to}\ \ \varphi_1(x_1,x_2) := \varphi(x_1,x_2) + x_2 \le 0,$$

where φ is taken from Example 1.31(ii) and its subdifferential $\partial\varphi(0,0)$ is calculated therein. Hence we have

$$\partial\varphi_1(0,0) = \big\{(v_1,v_2)\big|\ |v_1|+1 \le v_2 \le 2\big\} \cup \big\{(v_1,v_2)\big|\ 0 \le v_2 = -|v_1|+1\big\}$$

as depicted on Fig. 6.2. This shows that the result of Theorem 6.5(ii) (same in Corollary 6.6) allows us to rule out the nonoptimal point $\bar{x} = (0,0)$, while it cannot be done by using Warga's condition (6.22).

Next we derive yet another type of lower subdifferential optimality conditions for problem (6.11) with Lipschitzian data that are expressed in the *condensed form* via the basic subdifferential (1.24) of Lagrangian combinations of the initial data. Consider the standard *Lagrangian*

$$\mathcal{L}(x,\lambda_0,\ldots,\lambda_{m+r}) := \lambda_0\varphi_0(x) + \ldots + \lambda_{m+r}\varphi_{m+r}(x)$$

involving the cost function and all the functional (while not geometric) constraints and also the *extended Lagrangian*

$$\mathcal{L}_\Omega(x;\lambda_0,\ldots,\lambda_{m+r}) := \lambda_0\varphi_0(x) + \ldots + \lambda_{m+r}\varphi_{m+r}(x) + \delta(x;\Omega)$$

involving also the set geometric constraint via its indicator function.

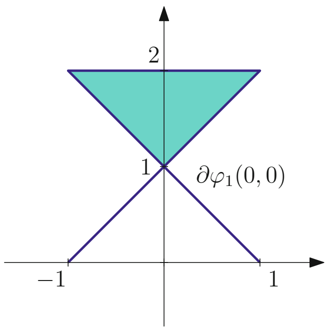

Fig. 6.2 Basic subdifferential of $\varphi_1(x_1,x_2) = \big||x_1|+x_2\big| + x_2$ at $(0,0)$.

Theorem 6.10 (Condensed Lower Subdifferential Optimality Conditions). *Let $\bar{x}$ be a local minimizer of problem* (6.11) *under the assumptions of Theorem* 6.5(ii). *Then there are multipliers $\lambda_0,\ldots,\lambda_{m+r}$, not equal to zero simultaneously, satisfying* (6.16) *and the condensed Lagrangian inclusions*

$$0 \in \partial_x\mathcal{L}_\Omega(\bar{x},\lambda_0,\ldots,\lambda_{m+r}) \subset \partial_x\mathcal{L}(\bar{x},\lambda_0,\ldots,\lambda_{m+r}) + N(\bar{x};\Omega). \quad (6.23)$$

Proof. Note that the second inclusion in (6.23) follows from the first one due to the subdifferential sum rule from Corollary 2.20. To justify the first inclusion therein, consider the set

$$\mathcal{E}(\varphi_0,\ldots,\varphi_{m+r},\Omega):=\Big\{(x,\alpha_0,\ldots,\alpha_{m+r})\in\mathbb{R}^{n+m+r+1}\Big|\,x\in\Omega,\varphi_i(x)\le\alpha_i,\\ i=0,\ldots,m;\ \varphi_i(x)=\alpha_i,\ i=m+1,\ldots,m+r\Big\}$$

and suppose without loss of generality that $\varphi_0(\bar{x}) = 0$. Denoting now by U a neighborhood of the local minimizer $\bar{x}$ in (6.11), we claim that the pair $(\bar{x},0)\in\mathbb{R}^n\times\mathbb{R}^{m+r+1}$ is an *extremal point* of the closed set system

$$\Omega_1:=\mathcal{E}(\varphi_0,\ldots,\varphi_{m+r},\Omega)\ \text{ and }\ \Omega_2:=\mathrm{cl}\,U\times\{0\}. \tag{6.24}$$

Indeed, we obviously have $(\bar{x},0)\in\Omega_1\cap\Omega_2$ and $(\Omega_1-(0,\nu_k,0,\ldots,0))\cap\Omega_2=\emptyset$, $k\in\mathbb{N}$, for any sequence of negative numbers $\nu_k\uparrow 0$ by the local optimality of $\bar{x}$ in (6.11). Applying to this system the basic *extremal principle* from Theorem 2.3 gives us multipliers $(\lambda_0,\ldots,\lambda_{m+r})\ne 0$ satisfying the inclusion

$$(0,-\lambda_0,\ldots,-\lambda_{m+r})\in N\big((\bar{x},0);\mathcal{E}(\varphi_0,\ldots,\varphi_{m+r},\Omega)\big), \tag{6.25}$$

which implies the conditions in (6.16) due to the structure of the set Ω_1 in (6.24). Furthermore, it follows from the scalarization formula of Theorem 1.32 and its proof that (6.25) can be equivalently rewritten as the first inclusion in (6.23) under the assumed local Lipschitz continuity of φ_i. △

If the geometric constraint set Ω is *convex*, the second inclusion in (6.23) can be written in the form of the *abstract maximum principle*.

Corollary 6.11 (Abstract Maximum Principle in Nondifferentiable Programming). *Suppose that the set Ω is convex in the assumptions of Theorem 6.10. Then there are multipliers $(\lambda_0,\ldots,\lambda_{m+r})\ne 0$ such that*

$$\langle v,\bar{x}\rangle=\max_{x\in\Omega}\langle v,x\rangle\ \textit{ for some }\ v\in-\partial_x\mathcal{L}(\bar{x},\lambda_0,\ldots,\lambda_{m_r}).$$

Proof. It follows from Theorem 6.10 by the representation of the normal cone to convex sets given in Proposition 1.7. △

We conclude this section by deriving *upper subdifferential* necessary optimality conditions for (6.11) that are independent of the obtained lower subdifferential conditions; see more discussions in Remark 6.2.

Theorem 6.12 (Upper Subdifferential Optimality Conditions in Nondifferentiable Programming). *Let $\bar{x}$ be a local minimizer of problem (6.11). Assume that the functions φ_i are locally Lipschitzian around $\bar{x}$ for the equality indices $i=m+1,\ldots,m+r$. Then for any $v_i\in\widehat{\partial}^+\varphi_i(\bar{x})$, $i=0,\ldots,m$, there are multipliers $(\lambda_0,\ldots,\lambda_{m+r})\ne 0$ satisfying (6.16) and the inclusion*

$$-\sum_{i=0}^{m}\lambda_i v_i \in \partial\Big(\sum_{i=m+1}^{m+r}\lambda_i\varphi_i\Big)(\bar{x}) + N(\bar{x};\Omega). \tag{6.26}$$

Proof. Suppose without loss of generality that $\widehat{\partial}^+\varphi_i(\bar{x}) \neq \emptyset$ for $i = 0,\ldots,m$. Applying the second part of Theorem 1.27 to $-v_i \in \widehat{\partial}(-\varphi_i)(\bar{x})$ (we can always assume that the functions $-\varphi_i$ are bounded from below, which is actually not needed for the localized version of Theorem 1.27 used in what follows) allows us to find functions $\psi_i\colon \mathbb{R}^n \to \mathbb{R}$ for $i = 0,\ldots,m$ satisfying

$$\psi_i(\bar{x}) = \varphi_i(\bar{x}) \ \text{ and } \ \psi_i(x) \geq \varphi_i(x) \ \text{ around } \ \bar{x}$$

and such that each $\psi_i(x)$ is continuously differentiable around $\bar{x}$ with the gradient $\nabla\psi_i(\bar{x}) = v_i$. It is easy to check that $\bar{x}$ is a local solution to the following optimization problem of type (6.11) but with the cost and inequality constraint functions continuously differentiable around $\bar{x}$:

$$\begin{cases} \text{minimize } \ \psi_0(x) \quad \text{ subject to}\\ \psi_i(x) \leq 0, \quad i = 1,\ldots,m,\\ \varphi_i(x) = 0, \quad i = m+1,\ldots,m+r,\\ x \in \Omega \subset \mathbb{R}^n. \end{cases} \tag{6.27}$$

To arrive finally at (6.26), it remains to apply to the solution $\bar{x}$ of (6.27) the second Lagrangian inclusion in (6.23) of Theorem 6.10 and then to use therein the elementary subdifferential sum rule from Proposition 1.30(ii). △

Employing further in (6.26) the subdifferential sum rule for Lipschitzian functions from Corollary 2.20 and weakening in this way the necessary optimality conditions for the case of equality constraints, we can express them in forms (6.17) and (6.19) via the corresponding subdifferential constructions for the separate functions φ_i, $i = m+1,\ldots,m+r$.

6.2 Problems of Bilevel Programming

In this section we begin considering a remarkable class of problems in hierarchical optimization known as *bilevel programming* and also as *Stackelberg games*. Such problems are highly interesting and challenging in optimization theory and important for numerous applications. There is an enormous bibliography on bilevel programming and related topics; see commentaries and references in Section 6.5 for more discussions on major approaches and results.

Our primary goal here is to reduce bilevel programs to those in nondifferentiable programming considered above and derive in this way several types of necessary optimality conditions in terms of the initial bilevel data by using the results of Section 6.1 together with subdifferentiation of marginal functions and other machinery of variational analysis.

6.2.1 Optimistic and Pessimistic Versions

Bilevel programming deals with problems of hierarchical optimization that address minimizing a given *upper-level/leader's* objective function $f(x, y)$ from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}$ subject to the upper-level constraints $x \in \Omega \subset \mathbb{R}^n$ along an optimal solution $y = y(x)$ to the parametric *lower-level/follower's* problem

$$\text{minimize}_y \ \varphi(x, y) \ \text{ subject to } \ y \in G(x) \tag{6.28}$$

with the objective/cost $\varphi\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the constraint set-valued mapping $G\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. For simplicity we confine ourselves to the case where the lower-level constraints are given by the parameterized inequality systems

$$G(x) := \big\{y \in \mathbb{R}^m \big|\ g(x, y) \le 0\big\}, \tag{6.29}$$

where $g = (g_1, \ldots, g_p)\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ and the vector inequality for g are understood componentwise. As follows from the proofs below, appropriately modified similar results can be derived for other types of constraints in (6.28).

Note that the bilevel optimization problem formulated above is not fully determined when the *solution/argminimum map*

$$S(x) := \text{argmin}\big\{\varphi(x, y)\big|\ y \in G(x)\big\}, \quad x \in \mathbb{R}^n, \tag{6.30}$$

for the lower-level problem is set-valued, since in this case we did not specify how to choose a single-valued decision function $y(x)$. To deal with such a typical situation, the two major versions, known as optimistic and pessimistic models, have been designated in bilevel programming. We always suppose that the argminimum sets $S(x)$ are nonempty around the reference point.

The *optimistic* version in bilevel programming is formulated as follows:

$$\begin{aligned} &\text{minimize } f_{opt}(x) \ \text{ subject to } \ x \in \Omega,\\ &\text{where } \ f_{opt}(x) := \inf\big\{f(x, y)\big|\ y \in S(x)\big\}, \end{aligned} \tag{6.31}$$

which means that the decision $y(x)$ is chosen in $S(x)$ to benefit the objective f_{opt}. As usual, a point $\bar{x} \in \Omega$ is called a global (local) optimistic solution to (6.31) if $f_{opt}(\bar{x}) \le f_{opt}(x)$ for all $x \in \Omega$ (sufficiently close to $\bar{x}$). From the economics viewpoint , this corresponds to a situation where the follower participates in the profit of the leader, i.e., there exists some cooperation between both players on the upper and lower levels.

However, it would not always be possible for the leader to convince the follower to make choices that are favorable for him or her. Hence it is necessary for the upper-level player to reduce damages resulting from undesirable selections on the lower level. This brings us to the *pessimistic* version in bilevel programming formulated in the following way:

$$\begin{aligned}&\text{minimize } f_{pes}(x) \text{ subject to } x \in \Omega,\\&\text{where } f_{pes}(x) := \sup\{f(x,y) \mid y \in S(x)\}.\end{aligned} \tag{6.32}$$

We can see that (6.32) is a special type of *minimax* problems, which challenges come from the complicated structure of the moving set $S(x)$ as the solution set to the lower-level optimization problem.

Our main attention in this chapter is paid to the *optimistic version*, although we'll present some comments on the pessimistic version as well. Further, we'll discuss in the exercise and commentary sections of this chapter a *multiobjective* approach to problems of bilevel programming that can be applied to both optimistic and pessimistic versions by reducing them to constrained multiobjective optimization problems studied in Chapter 9.

6.2.2 Value Function Approach

There are several approaches to optimistic bilevel programs known in the literature; see Section 6.5 for more discussions and references. We concentrate here on the so-called *value function approach*, which involves the optimal value function of the lower-level problem (6.28) defined by

$$\mu(x) := \inf\{\varphi(x,y) \mid y \in G(x)\}, \quad x \in \mathbb{R}^n, \tag{6.33}$$

and provides a reformulation of the bilevel problem (6.31) in the form

$$\begin{aligned}&\text{minimize } f(x,y) \text{ subject to } x \in \Omega,\\&g(x,y) \le 0, \text{ and } \varphi(x,y) \le \mu(x).\end{aligned} \tag{6.34}$$

It is easy to see that problem (6.34) is *globally* equivalent to the original optimistic bilevel program (6.31). The next proposition reveals relationships between *local* solutions to these problems. To give its exact formulation and proof, we introduce the *two-level value function*

$$\eta(x) := \inf\{f(x,y) \mid g(x,y) \le 0,\ \varphi(x,y) \le \mu(x)\}, \quad x \in \mathbb{R}^n, \tag{6.35}$$

and then define the corresponding modification of the solution map (6.30) by

$$\widetilde{S}(x) := \operatorname{argmin}\{\varphi(x,y) \mid g(x,y) \le 0,\ f(x,y) \le \eta(x)\}. \tag{6.36}$$

We obviously have $\widetilde{S}(x) \subset S(x)$ for all $x \in \mathbb{R}^n$.

Proposition 6.13 (Local Optimal Solutions to Optimistic Bilevel Programs). *Let $\widetilde{S}(x)$ be defined in* (6.36). *The following assertions hold:*

(i) *If $\bar{x}$ is a local optimal solution to* (6.31), *then for any $\bar{y} \in \widetilde{S}(\bar{x})$, the pair $(\bar{x}, \bar{y})$ is a local optimal solution to problem* (6.34).

(ii) *Conversely, let* $(\bar{x}, \bar{y})$ *be a local optimal solution to* (6.34) *for some* $\bar{y} \in \widetilde{S}(\bar{x})$, *and let the mapping* $\widetilde{S}$ *be inner semicontinuous at* $(\bar{x}, \bar{y})$. *Then* $\bar{x}$ *is a local optimal solution to the original optimistic bilevel problem* (6.31).

Proof. We verify (i) arguing by contradiction. Suppose that $(\bar{x}, \bar{y})$ with some $\bar{y} \in \widetilde{S}(\bar{x})$ is not a local optimal solution to (6.34). Then we find a sequence of (x_k, y_k) with $x_k \to \bar{x}$, $y_k \to \bar{y}$ so that $x_k \in \Omega$, $g(x_k, y_k) \le 0$, $\varphi(x_k, y_k) \le \mu(x_k)$, and $f(x_k, y_k) < f(\bar{x}, \bar{y}) = \eta(\bar{x})$ for all $k \in \mathbb{N}$. It follows from the construction of $\eta(\cdot)$ in (6.35) that $\eta(x_k) \le f(x_k, y_k)$. This shows that $f_{opt}(x_k) < f_{opt}(\bar{x})$, which contradicts the local optimality of $\bar{x}$ in (6.31).

To justify (ii), suppose that $\bar{x}$ is not a local optimal solution to (6.31) while the assumptions in (ii) are satisfied. Then we find a sequence $x_k \to \bar{x}$ with $x_k \in \Omega$ such that $f_{opt}(x_k) < f_{opt}(\bar{x})$ for all k. Since $\widetilde{S}$ is inner semicontinuous at $(\bar{x}, \bar{y})$, there is a sequence of $y_k \in \widetilde{S}(x_k)$ with $y_k \to \bar{y}$. This implies by (6.36) that $\varphi(x_k, y_k) = \mu(x_k)$, $g(x_k, y_k) \le 0$, and $f(x_k, y_k) < f(\bar{x}, \bar{y})$, which contradicts the local optimality of $(\bar{x}, \bar{y})$ in (6.34). △

The obtained results (see also Exercise 6.36) allow us to adequately replace the original optimistic bilevel problem (6.31) by the problem of constrained optimization (6.34) of the type considered in Section 6.1 and derive necessary optimality conditions for (6.31) from those for (6.34). Observe to this end that problem (6.34) is written in form (6.11) of nonlinear programming without equality constraints, where the inequality constraint

$$\varphi(x, y) - \mu(x) \le 0 \tag{6.37}$$

unavoidably involves the *nondifferentiable* function $\mu(x)$ of the marginal type (4.1) the generalized differential properties of which were studied in Section 4.1. Note however that the designated constraint (6.37) contains the term $-\mu(x)$, different from $\mu(x)$ in generalized differentiation, and that the constraint mapping G in (6.33) is given in the particular form (6.29).

It turns out that, even in the case where the upper-level constraint set Ω reduces to the whole space $\mathbb{R}^n$ or it is described by smooth inequalities, the usual Mangasarian-Fromovitz and other standard constraint qualifications as well as their natural extensions are *violated*; see more in Section 6.5.

6.2.3 Partial Calmness and Weak Sharp Minima

To overcome these difficulties, we present a qualification condition of another type that allows us to incorporate the troublesome constraint (6.37) into a *penalized* cost function and deal with it by using appropriate calculus rules of generalized differentiation. Consider a perturbed version of (6.34) with the linear parameterization of constraint (6.37) defined as follows:

$$\begin{aligned} &\text{minimize } f(x, y) \text{ subject to } x \in \Omega,\ g(x, y) \le 0, \\ &\text{and } \varphi(x, y) - \mu(x) + \vartheta = 0, \quad \vartheta \in \mathbb{R}. \end{aligned} \tag{6.38}$$

Definition 6.14 (Partial Calmness). *The unperturbed problem* (6.34) *is* PARTIALLY CALM *at its feasible solution* $(\bar{x},\bar{y})$ *if there exist a constant* $\kappa > 0$ *and a neighborhood* U *of the triple* $(\bar{x},\bar{y},0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$ *such that*

$$f(x,y) - f(\bar{x},\bar{y}) + \kappa|\vartheta| \ge 0 \tag{6.39}$$

for all $(x,y,\vartheta) \in U$ *feasible to* (6.38).

The next result reveals the role of partial calmness in bilevel programming.

Proposition 6.15 (Penalization via Partial Calmness). *Let* $(\bar{x},\bar{y})$ *be a partially calm feasible solution to problem* (6.34)*, and let* f *be continuous at this point. Then* $(\bar{x},\bar{y})$ *is a local optimal solution to the penalized problem*

$$\begin{aligned} &\text{minimize} \;\; f(x,y) + \kappa\big(\varphi(x,y) - \mu(x)\big) \\ &\text{subject to} \;\; x \in \Omega \;\; \text{and} \;\; g(x,y) \le 0, \end{aligned} \tag{6.40}$$

where the constant κ *is taken from* (6.39)*. Conversely, any local optimal solution* $(\bar{x},\bar{y})$ *to* (6.40) *with some number* $\kappa > 0$ *is partially calm in* (6.34).

Proof. By the assumed partial calmness, we get κ and U for which (6.39) holds. It follows from the continuity of f at $(\bar{x},\bar{y})$ that there are numbers $\gamma > 0$ and $\eta > 0$ such that $V := [(\bar{x},\bar{y}) + \eta\mathbb{B}] \times (-\gamma,\gamma) \subset U$ and that

$$|f(x,y) - f(\bar{x},\bar{y})| \le \kappa\gamma \;\; \text{whenever} \;\; (x,y) - (\bar{x},\bar{y}) \in \eta\mathbb{B}.$$

This allows us to establish the relationship

$$f(x,y) - f(\bar{x},\bar{y}) + \kappa\big(\varphi(x,y) - \mu(x)\big) \ge 0 \tag{6.41}$$

whenever $(x,y) \in [(\bar{x},\bar{y}) + \eta\mathbb{B}] \cap \operatorname{gph} G$ with G defined in (6.29) and $x \in \Omega$. Indeed, for $(x,y,\mu(x) - \varphi(x,y)) \in V$, we deduce (6.41) directly from (6.39). If otherwise $(x,y,\mu(x) - \varphi(x,y)) \notin V$, it follows that

$$\varphi(x,y) - \mu(x) \ge \gamma \;\; \text{and so} \;\; \kappa\big(\varphi(x,y) - \mu(x)\big) \ge \kappa\gamma.$$

This also implies (6.41) due to $f(x,y) - f(\bar{x},\bar{y}) \ge -\kappa\gamma$. To complete the proof of the first assertion of the proposition, observe that $\varphi(\bar{x},\bar{y}) - \mu(\bar{x}) = 0$ since $(\bar{x},\bar{y})$ is a feasible solution to (6.34). The converse statement follows directly from the definitions while arguing by contradiction. △

It is easy to see that a verifiable sufficient condition for the desired partial calmness is provided by the following notion of local weak sharp minima, which has been well recognized in qualitative and numerical aspects of optimization.

Definition 6.16 (Local Weak Sharp Minima). *Given* $Q \subset \mathbb{R}^s$*, we say that* $P \subset Q$ *is a set of* (LOCAL) WEAK SHARP MINIMA *for a function* $\phi\colon \mathbb{R}^s \to \mathbb{R}$ *over* Q *at* $\bar{z} \in P$ *with modulus* $\alpha > 0$ *if*

$$\phi(z) \ge \phi(\bar{z}) + \alpha\,\operatorname{dist}(z;P) \;\; \textit{for all} \;\; z \in Q \;\; \textit{near} \;\; \bar{z}. \tag{6.42}$$

The next proposition presents the precise formulation and provides a simple proof of the result needed in what follows with some *uniformity* in (6.42).

Proposition 6.17 (Partial Calmness from Uniform Weak Sharp Minima). *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution to the bilevel program* (6.34) *such that we have the* UNIFORM WEAK SHARP MINIMUM *condition:*

$$\varphi(x, y) - \mu(x) \geq \alpha \operatorname{dist}\big(y; S(x)\big) \ \text{ with some } \ \alpha > 0 \tag{6.43}$$

for all (x, y) *near* $(\bar{x}, \bar{y})$ *with* $x \in \Omega$ *and* $y \in G(x)$. *Assume that* f *is locally Lipschitzian around* $(\bar{x}, \bar{y})$. *Then problem* (6.34) *is partially calm at* $(\bar{x}, \bar{y})$.

Proof. Picking any triple (x, y, ϑ) feasible to problem (6.38) and sufficiently close to $(\bar{x}, \bar{y}, 0)$, we have $x \in \Omega$, $y \in G(x)$, and $\varphi(x, y) - \mu(x) + \vartheta = 0$, where $|\vartheta|$ is small enough. Using assumption (6.43) gives us some $v \in S(x)$ with

$$\varphi(x, y) - \mu(x) \geq \frac{\alpha}{2}\|y - v\| \geq 0.$$

Since $(\bar{x}, \bar{y})$ is a local optimal solution to (6.34), we get

$$\begin{aligned} & f(x, y) - f(\bar{x}, \bar{y}) \geq f(x, y) - f(x, v) \geq -\ell\|y - v\| \\ & \geq -\frac{2\ell}{\alpha}\Big(\varphi(x, y) - \mu(x)\Big) = -\kappa|\vartheta| \end{aligned}$$

with $\kappa := 2\ell/\alpha$, where $\ell > 0$ is a Lipschitz constant of f around $(\bar{x}, \bar{y})$. This justifies the partial calmness condition (6.39). △

Note that assumption (6.43) corresponds to the local weak sharp minimum condition of Definition 6.16 at $\bar{z} = (\bar{x}, \bar{y})$ with respect to y for any fixed feasible vector x with the following data:

$$z := (x, y), \ \ \phi(z) := \varphi(x, y), \ \ P := S(x), \ \text{ and } \ Q := G(x). \tag{6.44}$$

Observe also that the uniform weak sharpness in (6.43) requires that the constant $\alpha > 0$ therein can be selected *uniformly* in x. Proceeding in this way and deriving, in particular, sufficient conditions for (6.42) that being applied to (6.43) are independent of x, would allow us to decrease serious difficulties in dealing with nonsmooth marginal function (6.33) in the value function approach to optimistic bilevel programs.

Let us now present an easily verifiable condition of this type for weak sharp minimizers in nonlinear programming, which is of its own interest while being useful bilevel optimization; see below more discussions in this vein.

Proposition 6.18 (Sufficient Conditions for Weak Sharp Minima). *Let* $\bar{z}$ *be a local optimal solution to the nonlinear program:*

$$\text{minimize } \ \phi(z) \ \text{ subject to } \ \psi_i(z) \leq 0 \ \text{ for } i = 1, \ldots, p, \tag{6.45}$$

where the functions $\phi, \psi_i : \mathbb{R}^s \to \mathbb{R}$ *as* $i \in I(\bar{z}) := \{i \mid \psi_i(\bar{z}) = 0\}$ *are Fréchet differentiable at* $\bar{z}$. *Suppose that necessary optimality conditions for* $\bar{z}$ *hold in the qualified Karush-Kuhn-Tucker form*

$$\nabla\phi(\bar{z}) + \sum_{i \in I(\bar{z})} \lambda_i \nabla\psi_i(\bar{z}) = 0 \ \textit{for some} \ \lambda_i \geq 0$$

and that the following kernel condition

$$\bigcap_{i \in J} \ker \nabla\psi_i(\bar{z}) = \{0\} \ \textit{with} \ J := \left\{i \,\middle|\, \lambda_i > 0\right\} \tag{6.46}$$

is satisfied. Then there exists a positive constant α *such that*

$$\phi(z) - \phi(\bar{z}) \geq \alpha\|z - \bar{z}\| \ \textit{if} \ \psi_i(z) \leq 0 \ \textit{for} \ i = 1, \ldots, p \tag{6.47}$$

whenever z *is sufficiently close to* $\bar{z}$. *Consequently,* ϕ *admits a set of local weak sharp minima over* $Q := \{z \in \mathbb{R}^s \mid \psi_i(z) \leq 0, \ i = 1, \ldots, p\big\}$ *at* $\bar{z}$.

Proof. To justify (6.47) with some $\alpha > 0$, suppose on the contrary that there exists a sequence $\{z_k\} \subset Q$ with $z_k \neq \bar{z}$ and $z_k \to \bar{z}$ such that

$$\phi(z_k) - \phi(\bar{z}) \leq \frac{1}{k}\|z_k - \bar{z}\| \ \text{ for all } \ k \in \mathbb{N}. \tag{6.48}$$

Let $d_k := \frac{z_k - \bar{z}}{\|z_k - \bar{z}\|}$ and without loss of generality assume that $d_k \to d$ as $k \to \infty$ with $\|d\| = 1$. It follows from (6.48) by the (Fréchet) differentiability of ϕ at $\bar{z}$ that $\langle \nabla\phi(\bar{z}), d\rangle \leq 0$. On the other hand, the assumed differentiability of the active constraint functions at $\bar{z}$ ensures that

$$\langle \nabla\psi_i(\bar{z}), d\rangle \leq 0 \ \text{ for all } \ i \in I(\bar{z}).$$

Using the last two inequalities and the imposed KKT condition tells us that

$$0 \leq -\langle \nabla\phi(\bar{z}), d\rangle = \sum_{i \in J} \lambda_i \langle \nabla\psi_i(\bar{z}), d\rangle \leq 0,$$

which yields $\langle \nabla\psi_i(\bar{z}), d\rangle = 0$ for all $i \in J$. Thus we have $d = 0$ by the kernel condition (6.46), a contradiction that completes the proof. △

Observe that the *kernel condition* (6.46) is *essential* for Proposition 6.18 to hold. Indeed, consider problem (6.45) with $\phi, \psi : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\phi(z_1, z_2) := z_1^2 - z_2 \ \text{ and } \ \psi(z_1, z_2) := z_2.$$

Then $Q = \mathbb{R} \times \mathbb{R}_-$ and $\bar{z} := (0, 0)$ is the only solution to this problem. Since

$$\ker \nabla\psi(\bar{z}) = \mathbb{R} \times \{0\},$$

the kernel condition (6.46) is violated. It is easy to see that for any vector $z = (\gamma, 0) \in Q$ with $\gamma \neq 0$, we have the equalities

$$\phi(z) - \phi(\bar{z}) = \gamma^2 \text{ and } \|z - \bar{z}\| = \gamma,$$

which immediately imply that the conclusion in (6.47) doesn't hold, since the number $\gamma > 0$ can be chosen arbitrarily small.

Besides the presented conditions for weak sharp minima and their uniform counterparts, there are other sufficient conditions for these properties with various applications to partial calmness in bilevel programming and related topics; see more details in Sections 6.4 and 6.5. In particular, partial calmness is always satisfied for bilevel programs where lower-level problems are *linear* with respect to their *lower-level decision variables*; see Exercise 6.37(i).

The following examples illustrate some possibilities of verifying partial calmness in bilevel program via the results established above.

Example 6.19 (Verification of Partial Calmness via Penalization). Let us show that the penalty function characterization of partial calmness in Proposition 6.15 is a convenient tool to verify the validity or failure of partial calmness in bilevel programming. Consider first the *fully nonlinear*, at both lower and upper levels, bilevel program (6.34) with $(x, y) \in \mathbb{R}^2$, $\Omega = \mathbb{R}$, and

$$f(x, y) := \frac{(x-1)^2}{2} + \frac{y^2}{2}, \quad S(x) = \text{argmin}\left\{\frac{x^2}{2} + \frac{y^2}{2}\right\}.$$

It is easy to see that $S(x) = \{0\}$ for all $x \in \mathbb{R}$ and that $\mu(x) = x^2/2$ for the lower-level value function in (6.33). Furthermore, the pair $(\bar{x}, \bar{y}) = (1, 0)$ is the only solution to the upper-level problem, and so it is an optimal solution to the bilevel program under consideration. We have $\varphi(x, y) - \mu(x) = y^2/2$, and hence the corresponding unconstrained penalized problem (6.40) is

$$\text{minimize } \frac{(x-1)^2}{2} + \frac{y^2}{2} + \kappa\frac{y^2}{2}$$

with no constraints on (x, y). Observe that for any $\kappa > 0$, the latter problem is smooth and strictly convex and has the unique optimal solution $(\bar{x}, \bar{y}) = (1, 0)$. Thus the initial bilevel program is partially calm at this point.

On the other hand, replacing the upper-level cost function $f(x, y)$ by

$$\frac{(x-1)^2}{2} + \frac{(y-1)^2}{2}$$

and keeping the same lower-level problem gives us the bilevel program (6.34) with the optimal solution $(\bar{x}, \bar{y}) = (1, 1)$, which fails to satisfy the partial calmness condition. Indeed, it is easy to see that the corresponding penalized problem (6.40) has the only optimal solution

$$\left(1, \frac{1}{1+\kappa}\right) \neq (1,1) \text{ whenever } \kappa > 0.$$

Example 6.20 (Verification of Partial Calmness via Uniform Weak Sharp Minima). Consider the constrained optimization problem in $\mathbb{R}^3$:

$$\text{minimize } \frac{x_1^2}{2} + \frac{x_2^2}{2} \text{ subject to } a_i \le x_i \le b_i, \quad i = 1,2,3. \tag{6.49}$$

It is not hard to check that optimal solutions to this problem constitute the set of weak sharp minima if either $a_i > 0$ or $b_i < 0$ for $i = 1,2$; see Exercise 6.38(ii). Thus Proposition 6.17 tells us that any bilevel program having (6.49) as its lower-level problem with the above parameters a_i, b_i is partially calm at each of its local optimal solutions.

Note that Example 6.19 shows that partial calmness in bilevel programs may significantly *depend* on the structure of upper-level objectives. On the contrary, Example 6.20 describes a class of multidimensional bilevel programs where partial calmness holds *independently* of the upper level.

6.3 Bilevel Programs with Smooth and Lipschitzian Data

In this section we develop the value function approach to bilevel programming discussed above to obtain necessary optimality conditions in optimistic bilevel programs first with smooth and then with Lipschitzian initial data. For simplicity, consider here the bilevel program (6.31) in the value/marginal function form (6.34) with the upper-level constraint set Ω given by the inequalities

$$\Omega := \left\{x \in \mathbb{R}^n \,\middle|\, h(x) \le 0\right\} \text{ with } h(x) = \big(h_1(x), \ldots, h_q(x)\big), \tag{6.50}$$

which are described by the real-valued functions h_j. Our major results are derived under the inner semicontinuity of the argminimum map S in (6.30) at the reference local optimal solution $(\bar{x}, \bar{y})$ by passing to problem (6.34) via Proposition 6.13. Imposing further the partial calmness of (6.34) at the given local solution $(\bar{x}, \bar{y})$ and using Proposition 6.15, we reduce (6.34) to the single-level programming form (6.40), which is essentially used in our proofs.

Observe that problem (6.40) with constraints (6.50) is written as

$$\begin{aligned} &\text{minimize } f(x,y) + \kappa\big(\varphi(x,y) - \mu(x)\big) \\ &\text{subject to } g(x,y) \le 0 \text{ and } h(x) \le 0 \end{aligned} \tag{6.51}$$

for some $\kappa > 0$. Hence it can be treated as a particular case of the mathematical program (6.11) with inequality constraints. The most essential specific features of (6.51) are *intrinsic nonsmoothness* of the marginal function $\mu(x)$ from (6.33), regardless of smoothness of the initial data, and the presence of function (6.33) in the objective of (6.51) with the *negative sign*. Nevertheless, the above subdifferential results for marginal functions and explicit representations of normals to sets described by inequality constraints allow us to efficiently proceed in deriving necessary optimality conditions for (6.34).

6.3.1 Optimality Conditions for Smooth Bilevel Programs

Given a feasible solution $(\bar{x}, \bar{y})$ to the original optimistic bilevel program (6.34) with the constraint set Ω defined in (6.50), denote by

$$I(\bar{x}, \bar{y}) := \big\{i \in \{1, \ldots, p\}\big|\ g_i(\bar{x}, \bar{y}) = 0\big\},\ \ J(\bar{x}) := \big\{j \in \{1, \ldots, q\}\big|\ h_j(\bar{x}) = 0\big\}$$

the collections of the corresponding active constraint indices. Considering first problem (6.34) with smooth initial data and following the traditional terminology in bilevel programming, we say that $(\bar{x}, \bar{y})$ is *lower-level regular* if for any nonnegative numbers λ_i the implication

$$\Big[\sum_{i \in I(\bar{x}, \bar{y})} \lambda_i \nabla_y g_i(\bar{x}, \bar{y}) = 0\Big] \Longrightarrow \Big[\lambda_i = 0 \ \text{ whenever } \ i \in I(\bar{x}, \bar{y})\Big] \qquad (6.52)$$

holds. Similarly, $\bar{x}$ is *upper-level regular* if

$$\Big[\lambda_j \geq 0,\ \sum_{j \in J(\bar{x})} \lambda_j \nabla h_j(\bar{x}) = 0\Big] \Longrightarrow \Big[\lambda_j = 0 \ \text{ whenever } \ j \in J(\bar{x})\Big]. \qquad (6.53)$$

Now we are ready to present our first result on necessary optimality conditions for the original optimistic version of bilevel programming (6.31) with the upper-level constraint set Ω given in (6.50).

Theorem 6.21 (Optimality Conditions for Smooth Bilevel Programs, I). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to the bilevel program* (6.31) *with Ω from* (6.50)*. Assume that all the functions therein are smooth around $(\bar{x}, \bar{y})$ and $\bar{x}$, respectively, and that the bilevel program is partially calm at $(\bar{x}, \bar{y})$. Suppose further that $(\bar{x}, \bar{y})$ is lower-level regular, that $\bar{x}$ is upper-level regular, and that the solution map S in* (6.30) *is inner semicontinuous at $(\bar{x}, \bar{y})$. Then there are numbers $\kappa > 0$, $\lambda_1, \ldots, \lambda_p$, $\beta_1, \ldots, \beta_p$, and $\alpha_1, \ldots, \alpha_q$ such that*

$$\nabla_x f(\bar{x}, \bar{y}) + \sum_{i=1}^{p}(\beta_i - \kappa\lambda_i)\nabla_x g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{q} \alpha_j \nabla h_j(\bar{x}) = 0, \qquad (6.54)$$

$$\nabla_y f(\bar{x}, \bar{y}) + \kappa \nabla_y \varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \beta_i \nabla_y g_i(\bar{x}, \bar{y}) = 0, \qquad (6.55)$$

$$\nabla_y \varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_i \nabla_y g_i(\bar{x}, \bar{y}) = 0 \tag{6.56}$$

with the following sign and complementary slackness conditions:

$$\lambda_i \geq 0, \quad \lambda_i g_i(\bar{x}, \bar{y}) = 0 \;\; \text{for all} \;\; i = 1, \ldots, p, \tag{6.57}$$

$$\beta_i \geq 0, \quad \beta_i g_i(\bar{x}, \bar{y}) = 0 \;\; \text{for all} \;\; i = 1, \ldots, p, \tag{6.58}$$

$$\alpha_j \geq 0, \quad \alpha_j h_j(\bar{x}) = 0 \;\; \text{for all} \;\; j = 1, \ldots, q. \tag{6.59}$$

Proof. Proposition 6.13(i) tells us that $(\bar{x}, \bar{y})$ is a local optimal solution to (6.34), even without the lower semicontinuity of S at $(\bar{x}, \bar{y})$. Furthermore, the imposed partial calmness ensures that $(\bar{x}, \bar{y})$ is a local minimizer of the penalized problem (6.51) with some fixed $\kappa > 0$. As mentioned above, the latter problem is a particular case of the nondifferentiable program (6.11) with only the inequality constraints therein. To apply to it the results of Theorem 6.5(ii), we need to check first that the marginal function $\mu(x)$ from (6.33), where $G(x)$ defined in (6.29) is locally Lipschitzian around $\bar{x}$ under the assumed lower-level regularity of $(\bar{x}, \bar{y})$ in the bilevel program under consideration.

Indeed, it is easy to see that the function $\mu(x)$ is l.s.c. around $\bar{x}$. Since $\bar{y} \in S(\bar{x})$, the mapping M in (4.2) obviously reduces in this case to S that is assumed to be inner semicontinuous at $(\bar{x}, \bar{y})$, we deduce from formula (4.5) of Theorem 4.1(i) the following inclusion:

$$\partial^\infty \mu(\bar{x}) \subset D^* G(\bar{x}, \bar{y})(0) \;\; \text{with} \;\; G(x) = \big\{ y \in \mathbb{R}^m \big| \; g(x, y) \leq 0 \big\}. \tag{6.60}$$

The result of Exercise 2.51(ii) on representing the normal cone to the set

$$\operatorname{gph} G = \big\{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m \big| \; g_i(x, y) \leq 0, \quad i = 1, \ldots, p \big\}$$

at $(\bar{x}, \bar{y})$ tells us that $D^* G(\bar{x}, \bar{y})(0) = \{0\}$ under the imposed lower-level regularity. Thus we have $\partial^\infty \mu(\bar{x}) = \{0\}$ from (6.60), which ensures that $\mu(\cdot)$ is locally Lipschitzian around $\bar{x}$ due to Theorem 1.22; see also Exercise 4.25(iv).

Applying now the necessary optimality conditions of Theorem 6.5(ii) to problem (6.51) at $(\bar{x}, \bar{y})$ and then using the subdifferential sum rule from Proposition 1.30(ii) give us multipliers $\lambda \geq 0$, $\beta_1, \ldots, \beta_p$, and $\alpha_1, \ldots, \alpha_q$, not all zero, satisfying the sign and complementary slackness conditions in (6.58) and (6.59) and the generalized Lagrangian inclusion

$$\begin{aligned} 0 \in \lambda \nabla f(\bar{x}, \bar{y}) + \kappa \lambda \nabla \varphi(\bar{x}, \bar{y}) + \big(\kappa \lambda \partial(-\mu)(\bar{x}), 0 \big) \\ + \sum_{i=1}^{p} \beta_i \nabla g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{q} \alpha_j \big(\nabla h_j(\bar{x}), 0 \big). \end{aligned} \tag{6.61}$$

It follows from the assumed lower-level regularity of $(\bar{x}, \bar{y})$ and upper-level regularity of $\bar{x}$, combined with the sign and complementarity slackness conditions, that $\lambda \neq 0$ and hence $\lambda = 1$ without loss of generality. Since

$$\partial(-\mu)(\bar{x}) \subset \overline{\partial}(-\mu)(\bar{x}) = -\overline{\partial}\mu(\bar{x}) = -\text{co}\,\partial\mu(\bar{x})$$

by (1.83) and (1.79) due to the Lipschitz continuity of $\mu(x)$, we can incorporate into (6.61) with $\lambda = 1$ the basic subdifferential estimate (4.4) for the marginal function with the smooth constraints (6.29) under the imposed inner semicontinuity assumption on S at $(\bar{x}, \bar{y})$. This gives us multipliers $\lambda_1, \ldots, \lambda_p$ satisfying (6.56) and (6.57) such that the conditions in (6.55) and

$$\nabla_x f(\bar{x}, \bar{y}) + \kappa\nabla_x\varphi(\bar{x}, \bar{y}) - \kappa\Big[\nabla_x\varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p}\lambda_i\nabla_x g_i(\bar{x}, \bar{y})\Big]$$
$$+\sum_{i=1}^{p}\beta_i\nabla_x g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{q}\alpha_j\nabla h_j(\bar{x}) = 0$$

hold. Collecting the like terms in the latter equation, we arrive at the remaining equality (6.54) and thus complete the proof of the theorem. △

Now we develop a different device of necessary optimality conditions for bilevel programs, which brings us to results significantly different from Theorem 6.21 in both assumptions and conclusions. To proceed, let us first present a lemma of its own interest that is crucial in the device below. It concerns calculus of regular subgradients, which is pretty limited in general (e.g., no sum rule, etc.) while happens to contain a nice *difference rule* particularly important in applications to bilevel programs via the value function approach. Note that the proof of the following lemma is based on the *smooth variational* description of regular subgradients taken from Theorem 1.27. Observe also that the necessary optimality condition in this lemma has been already deduced in Proposition 6.3 from the upper subdifferential one.

Lemma 6.22 (Difference Rule for Regular Subgradients). *Let both functions $\varphi_1, \varphi_2\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be finite at $\bar{x}$, and let $\widehat{\partial}\varphi_2(\bar{x}) \neq \emptyset$. Then we have*

$$\widehat{\partial}(\varphi_1 - \varphi_2)(\bar{x}) \subset \bigcap_{v\in\widehat{\partial}\varphi_2(\bar{x})}\Big[\widehat{\partial}\varphi_1(\bar{x}) - v\Big] \subset \widehat{\partial}\varphi_1(\bar{x}) - \widehat{\partial}\varphi_2(\bar{x}). \tag{6.62}$$

This implies that any local minimizer $\bar{x}$ of the difference function $\varphi_1 - \varphi_2$ satisfies the necessary optimality condition

$$\widehat{\partial}\varphi_2(\bar{x}) \subset \widehat{\partial}\varphi_1(\bar{x}). \tag{6.63}$$

Proof. To verify the first inclusion in (6.62), fix any $u \in \widehat{\partial}(\varphi_1 - \varphi_2)(\bar{x})$ and $v \in \widehat{\partial}\varphi_2(\bar{x})$. Employing the first assertion of Theorem 1.27, find a real-valued function $s(\cdot)$ defined on a neighborhood U of $\bar{x}$ such that it is (Fréchet) differentiable at $\bar{x}$ satisfying the relationships

$$s(\bar{x}) = \varphi_2(\bar{x}), \quad \nabla s(\bar{x}) = v, \quad \text{and} \quad s(x) \le \varphi_2(x) \quad \text{for all} \quad x \in U.$$

This yields due to definition (1.33) of the regular subgradient $u \in \widehat{\partial}(\varphi_1 - \varphi_2)(\bar{x})$ that for any $\varepsilon > 0$ there exists $\gamma > 0$ such that

$$\begin{aligned}\langle u, x - \bar{x}\rangle &\le \varphi_1(x) - \varphi_2(x) - \big(\varphi_1(\bar{x}) - \varphi_2(\bar{x})\big) + \varepsilon\|x - \bar{x}\| \\ &\le \varphi_1(x) - s(x) - \big(\varphi_1(\bar{x}) - s(\bar{x})\big) + \varepsilon\|x - \bar{x}\|\end{aligned}$$

whenever $\|x - \bar{x}\| \le \gamma$. The latter ensures by Proposition 1.30(ii) that

$$u \in \widehat{\partial}(\varphi_1 - s)(\bar{x}) = \widehat{\partial}\varphi_1(\bar{x}) - \nabla s(\bar{x}) = \widehat{\partial}\varphi_1(\bar{x}) - v,$$

which justifies the first inclusion in (6.62) and obviously yields the second one.

To verify (6.63), observe that it is trivial if $\widehat{\partial}\varphi_2(\bar{x}) = \emptyset$. Otherwise, pick $v \in \widehat{\partial}\varphi_2(\bar{x})$ and deduce from (6.62) by the generalized Fermat rule that

$$0 \in \widehat{\partial}(\varphi_1 - \varphi_2)(\bar{x}) \subset \widehat{\partial}\varphi_1(\bar{x}) - v,$$

which shows that $v \in \widehat{\partial}\varphi_1(\bar{x})$ and thus justifies the set inclusion (6.63). △

For simplicity we consider in the next theorem the optimistic bilevel problem (6.31) without upper-level constraints.

Theorem 6.23 (Optimality Conditions for Smooth Bilevel Programs, II). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to problem (6.31) with $\Omega = \mathbb{R}^n$ and with the functions $f, g_1, \ldots, g_p, \varphi$ continuously differentiable around $(\bar{x}, \bar{y})$. Assume that this problem is partially calm at the point $(\bar{x}, \bar{y})$, which is lower-level regular for (6.34), and also that $\widehat{\partial}\mu(\bar{x}) \neq \emptyset$ for lower-level value function (6.33). Then there exist multipliers ν_i and β_i as $i = 1, \ldots, p$ such that β_i satisfy the sign and complementarity slackness conditions in (6.58), that ν_i satisfy the complementarity slackness conditions*

$$\nu_i g_i(\bar{x}, \bar{y}) = 0 \ \textit{for all} \ \ i = 1, \ldots, p,$$

and that the following equalities hold:

$$\begin{aligned}&\nabla f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \nu_i \nabla g_i(\bar{x}, \bar{y}) = 0, \\ &\nabla_y \varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \beta_i \nabla_y g_i(\bar{x}, \bar{y}) = 0.\end{aligned}$$

Proof. By Proposition 6.13(i) we get that $(\bar{x}, \bar{y})$ is a local optimal solution to the nondifferentiable program (6.34). It follows from Proposition 6.15 and the infinite constraint penalization via the indicator function $\delta(\cdot; \operatorname{gph} G)$ that $(\bar{x}, \bar{y})$ a local optimal solution to the unconstrained problem:

$$\text{minimize} \ \ f(x, y) + \kappa\big(\varphi(x, y) - \mu(x)\big) + \delta\big((x, y); \operatorname{gph} G\big), \qquad (6.64)$$

where the constant $\kappa > 0$ is taken from the definition of partial calmness. Applying now the necessary optimality condition (6.63) from Lemma 6.22 to the difference function in (6.64), we get

$$\big(\kappa\widehat{\partial}\mu(\bar{x}),0\big)\subset\widehat{\partial}\big(f(\cdot)+\kappa\varphi(\cdot)+\delta(\cdot;\operatorname{gph}G)\big)(\bar{x},\bar{y}). \tag{6.65}$$

It is not hard to observe (cf. the proof of Theorem 4.1) that

$$\big(\widehat{\partial}\mu(\bar{x}),0\big)\subset\widehat{\partial}\big(\varphi(\cdot)+\delta(\cdot;\operatorname{gph}G)\big)(\bar{x},\bar{y}). \tag{6.66}$$

Passing to the larger limiting subdifferential on the right-hand sides of (6.65) and (6.66) and employing the elementary subdifferential sum rule, we have

$$\big(\kappa\widehat{\partial}\mu(\bar{x}),0\big)\subset\nabla f(\bar{x},\bar{y})+\kappa\nabla\varphi(\bar{x},\bar{y})+N\big((\bar{x},\bar{y});\operatorname{gph}G\big),$$
$$\big(\widehat{\partial}\mu(\bar{x}),0\big)\subset\nabla\varphi(\bar{x},\bar{y})+N\big((\bar{x},\bar{y});\operatorname{gph}G\big).$$

Then the description of basic normals from Exercise 2.51 for sets given by inequality constraints under the imposed lower-level regularity ensures the existence of multipliers λ_i and β_i satisfying the sign and complementarity slackness conditions in (6.57) and (6.58) as well as a vector $v\in\widehat{\partial}\mu(\bar{x})$ with

$$(v,0)=\nabla\varphi(\bar{x},\bar{y})+\sum_{i=1}^{p}\lambda_i\nabla g_i(\bar{x},\bar{y})\ \text{ and}$$
$$\kappa(v,0)=\nabla f(\bar{x},\bar{y})+\kappa\nabla\varphi(\bar{x},\bar{y})+\sum_{i=1}^{p}\beta_i\nabla g_i(\bar{x},\bar{y}).$$

Dividing the latter inclusion by $\kappa > 0$ and denoting $\nu := \kappa^{-1}$ while keeping the same notation for the modified multipliers β_i and collecting the like terms, we arrive at the equalities claimed in the theorem. △

The following example, consisting of two parts, illustrates the possibility to solve bilevel programs by using necessary optimality conditions obtained in Theorem 6.21 and Theorem 6.23, respectively.

Example 6.24 (Solving Bilevel Programs via Optimality Conditions).
(i) *Applying the conditions of Theorem* 6.21. Consider the bilevel program:

$$\text{minimize }\ f(x,y):=-y\ \text{ subject to }\ y\in S(x),$$

where $S\colon\mathbb{R}\rightrightarrows\mathbb{R}$ is the solution map of the nonlinear lower-level problem:

$$\text{minimize }\ \varphi(x,y):=-y^2+x^4-3x^2+1\ \text{ subject to}$$
$$y\in G(x):=\big\{y\in\mathbb{R}\big|\ y+x^2-1\le 0,\ -y+x^2-1\le 0\big\}.$$

It is easy to check that the bilevel program in this example admits an optimal solution with x belonging to the interval $[-1,1]$. Furthermore, we have

$$S(x) = \big\{ -x^2+1, x^2-1\big\} \ \text{and} \ \mu(x) = -x^2 \ \text{for} \ x \in [-1, 1].$$

This shows that S is inner semicontinuous at any point $(x, y) \in \operatorname{gph} S$ and the lower-regularity assumption (6.52) is satisfied everywhere but $(-1, 0)$ and $(1, 0)$; the upper regularity is automatic due to the absence of inequality constraints on the upper level. Applying Theorem 6.21, we calculate

$$\begin{aligned} &\nabla f(x, y) = (0, -1), \quad \nabla\varphi(x, y) = (4x^3 - 6x, -2y),\\ &\nabla g_1(x, y) = (2x, 1), \quad \nabla g_2(x, y) = (2x, -1) \end{aligned}$$

and hence obtain the following relationships:

$$\begin{aligned} &0 = (\beta_1 - \kappa\lambda_1)2x + (\beta_2 - \kappa\lambda_2)2x, \quad 0 = -1 + \kappa(-2y) + \beta_1(1) + \beta_2(-1),\\ &0 = -2y + \lambda_1(1) + \lambda_2(-1), \quad 0 = \lambda_1(y + x^2 - 1) = \lambda_2(-y + x^2 - 1),\\ &0 = \beta_1(y + x^2 - 1) = \beta_2(-y + x^2 - 1) \end{aligned}$$

with $\kappa > 0$ and all the nonnegative multipliers. Solving the above system gives us the points $(x, y) \in \{(0, 1), (0, -1), (1, 0), (-1, 0)\}$ suspicious for optimality. Comparing the value of the upper-level objective at these points, we arrive at the pair $(\bar{x}, \bar{y}) = (0, 1)$ and check finally that the given bilevel program is partially calm at $(0, 1)$. Thus this pair is the unique optimal solution to the bilevel program under consideration by Theorem 6.21.

(ii) *Applying the conditions of Theorem* 6.23. Consider the program:

$$\text{minimize} \ \ f(x, y) := -y \ \ \text{subject to} \ y \in S(x),$$

where $S\colon \mathbb{R} \rightrightarrows \mathbb{R}$ is the solution map for the lower-level problem:

$$\begin{aligned} &\text{minimize} \ \ \varphi(x, y) := -y^2 \ \ \text{subject to}\\ &y \in G(x) := \big\{y \in \mathbb{R}\big| \ -x + y^4 - 1 \le 0, \ x + y^4 - 1 \le 0\big\}. \end{aligned}$$

It is easy to see that this bilevel program admits an optimal solution. Then we calculate the lower-level solution map by $S(x) = \{\pm\sqrt{1 - |x|}\}$ and the marginal function by $\mu(x) = -\sqrt{1 - |x|}$ for which $\widehat{\partial}\mu(x) \neq \emptyset$ on $\mathbb{R}$. Applying the necessary optimality conditions of Theorem 6.23 gives us the relationships

$$\begin{gathered} -\nu_1 + \nu_2 = 0, \ -1 + 4y^3\nu_1 + 4y^3\nu_2 = 0,\\ -2y + 4y^3\beta_1 + 4y^3\beta_2 = 0, \ \nu_1(-x + y^4 - 1) = \nu_2(x + y^4 - 1) = 0,\\ \beta_1(-x + y^4 - 1) = \beta_1(x + y^4 - 1) = 0, \ \beta_1 \ge 0, \ \beta_2 \ge 0. \end{gathered}$$

Solving this system of equations, we obtain the points $(x, y) = (0, \pm 1)$. Comparing the upper-level objective selects the point $(0, 1)$. Since the bilevel program under consideration is partially calm at $(0, 1)$, we conclude that $(0, 1)$ is the unique optimal solution to this problem.

6.3.2 Optimality Conditions for Lipschitzian Problems

Analyzing the proofs of Theorem 6.21 and Theorem 6.23, it is not difficult to observe that these proofs and the results used therein lead us to necessary optimality conditions for bilevel programs with *Lipschitzian data*. In the following Lipschitzian versions of necessary optimality conditions, we replace the gradients of the Lipschitzian functions involved by their basic subgradients and reformulate the upper-level regularity condition (6.53) as satisfied for all the subgradients of h_j at $\bar{x}$ and the lower-level regularity condition (6.52) as satisfied for all (u_i, v_i) with $(u_i, v_i) \in \partial g_i(\bar{x}, \bar{y})$. In this way we have:

Theorem 6.25 (Optimality Conditions for Lipschitzian Bilevel Programs, I). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to the bilevel program (6.31) with Ω from (6.50). Suppose that all the functions therein are locally Lipschitzian around $(\bar{x}, \bar{y})$ and $\bar{x}$, respectively, under the validity of the other assumptions of Theorem 6.21. Then there exist a number $\nu > 0$, multipliers $\lambda_1, \ldots, \lambda_p$, $\beta_1, \ldots, \beta_p$, and $\alpha_1, \ldots, \alpha_q$ as well as a vector $u \in \mathbb{R}^n$ such that conditions (6.57)–(6.59) are satisfied together with*

$$(u, 0) \in \operatorname{co} \partial\varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_i \operatorname{co} \partial g_i(\bar{x}, \bar{y}) \;\; \textit{and}$$

$$(u, 0) \in \partial\varphi(\bar{x}, \bar{y}) + \nu\partial f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \beta_i \partial g_i(\bar{x}, \bar{y}) + \sum_{j=1}^{q} \alpha_j \Big(\partial h_j(\bar{x}), 0\Big).$$

Theorem 6.26 (Optimality Conditions for Lipschitzian Bilevel Programs, II). *Let $(\bar{x}, \bar{y})$ be a local optimal solution to problem (6.31) with $\Omega = \mathbb{R}^n$. Suppose that all the functions therein are locally Lipschitzian around $(\bar{x}, \bar{y})$ under the validity of the other assumptions of Theorem 6.23. Then there exist a number $\nu > 0$, nonnegative multipliers λ_i and β_i satisfying the complementary slackness condition (6.57) and (6.58) as $i = 1, \ldots, p$, and a vector $u \in \mathbb{R}^n$ such that we have the inclusions*

$$(u, 0) \in \partial\varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_i \partial g_i(\bar{x}, \bar{y}) \;\; \textit{and}$$

$$(u, 0) \in \partial\varphi(\bar{x}, \bar{y}) + \nu\partial f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \beta_i \partial g_i(\bar{x}, \bar{y}).$$

The proofs of these results as well as their several extensions are assigned in the exercises of Section 3.4.

Remark 6.27 (Inner Semicompactness vs. Inner Semicontinuity of Solution Maps). Observe that the necessary optimality conditions of Theorems 6.23 and 6.26 hold, in contrast to those in Theorems 6.21 and 6.25, without the inner semicontinuity assumption on the solution map S (6.30). While the latter assumption is satisfied in rather broad settings (e.g., when S is Lipschitz-like around $(\bar{x}, \bar{y})$ and also when

$S(\bar{x})$ is a singleton but $S(x)$ may not be for x close to $\bar{x}$), it definitely doesn't hold in generality.

A significantly less restrictive assumption in the frameworks of Theorems 6.21 and 6.25 is provided by the *inner semicompactness* property of S at the domain point $\bar{x}$ defined in Exercise 2.46. In finite dimensions this property is rather close to the *local boundedness* of S around $\bar{x}$. The results obtained under the lower semicompactness of S are different from their inner semicontinuity counterparts in that they require considering all the vectors $\bar{y}$ from the set $S(\bar{x})$. The proofs go in the same direction with replacing the results on the subdifferentiation of marginal functions from Theorem 4.1(i) by their "union" versions from assertion (ii) therein.

Some consequences and specifications of the necessary optimality conditions for bilevel programs with fully and partially convex (smooth and nonsmooth) structures can be derived from Theorems 6.25 and 6.26. However, significantly stronger results for problems of these type will be obtained as consequences of those given in Subsection 7.5.4. Hence we omit here formulating the corresponding consequences of Theorems 6.25 and 6.26 while leaving this as exercises for the reader; see more hints in Exercise 6.46.

6.4 Exercises for Chapter 6

Exercise 6.28 (Optimization Problems with Geometric Constraints).

(i) Derive necessary optimality conditions of Theorem 6.1(ii) and Proposition 6.4(ii) directly from the extremal principle.

(ii) Extend the necessary optimality conditions of Theorem 6.1 and Proposition 6.4 to appropriate Banach spaces. Which assumption should be added to (6.4) to ensure the validity of Theorem 6.1(ii) in infinite dimensions? *Hint:* Compare this with [523, Propositions 5.2 and 5.3 and Theorem 5.5].

(iii) Construct an example of the optimization problem (6.1) with a Lipschitz continuous objective function φ defined on a Banach space X such that condition (6.5) doesn't hold at a local minimizer $\bar{x}$ of this problem.

Exercise 6.29 (Problems of DC Programming).

(i) Extend the results of Proposition 6.3 to problems with convex geometric constraints of the type $x \in \Omega$.

(ii) Show that the convexity of the function φ_1 in Proposition 6.3 can be replaced by the more general property of *quasiconvexity* in the sense that

$$\varphi\big(\lambda x_1 + (1-\lambda)x_2\big) \le \max\big\{\varphi(x_1), \varphi(x_2)\big\} \text{ for all } x_1, x_2 \in \mathbb{R}^n,\ \lambda \in [0,1].$$

(iii) Do all the results of Proposition 6.3 and the parts (i) and (ii) of this exercise hold in arbitrary Banach spaces?

Exercise 6.30 (Necessary Conditions in Nondifferentiable Programming).

(i) Extend necessary optimality conditions of Theorem 6.5 for problems of nondifferentiable programming of type (6.11) with finitely many geometric constraints.

(ii) Derive appropriate versions of Theorem 6.5 in Asplund spaces. *Hint:* Proceed as in the proof of Theorem 6.5 with applying the corresponding results in infinite dimensions taken from Exercises 2.31 and 1.69; compare with [523, Theorem 5.5].

Exercise 6.31 (Extended Lagrangian Conditions for Lipschitzian Nondifferentiable Programs in Asplund Spaces). Consider the nondifferentiable program (6.11) described by locally Lipschitzian functions on an Asplund space. Show that the necessary optimality conditions of Theorem 6.10 hold true in this case.

Hint: Proceed by using the exact extremal principle from Exercise 2.31 and the subdifferential sum from Exercise 2.54 in Asplund spaces; cf. [523, Theorem 5.24].

Exercise 6.32 (Constraint Qualifications in Nondifferentiable Programming).

(i) Based on Theorem 6.5(ii) and the normal cone representations from Exercises 2.51 and 2.52 in the case of locally Lipschitzian functions, derive constraint qualifications ensuring that $\lambda_0 = 1$ in the optimality conditions of Theorem 6.5.

(ii) Which constraint qualifications correspond to those obtained in (i) in the case of smooth constraint functions φ_i and $\Omega = \mathbb{R}^n$?

(iii) Derive extensions of the results in (i) to problems in Asplund spaces.

Exercise 6.33 (Necessary Optimality Conditions for Problems with Inclusion/Operator Constraints). Given $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, $f\colon \mathbb{R}^n \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$, and $\Theta \subset \mathbb{R}^m$, consider the optimization problem:

$$\text{minimize } \varphi(x) \text{ subject to } f(x) \in \Theta,\ x \in \Omega, \tag{6.67}$$

where f is strictly differentiable at the reference local minimizer $\bar{x} \in f^{-1}(\Theta) \cap \Omega$ and its Jacobian matrix $\nabla f(\bar{x})$ has full row rank.

(i) Prove that the upper subdifferential optimality condition

$$-\widehat{\partial}^+\varphi(\bar{x}) \subset \nabla f(\bar{x})^* N\big(f(\bar{x}); \Theta\big) + N(\bar{x}; \Omega)$$

holds under the validity of the constraint qualification

$$\nabla f(\bar{x})^* N\big(f(\bar{x}); \Theta\big) \cap \big(-N(\bar{x}; \Omega)\big) = \{0\}.$$

(ii) Derive lower subdifferential optimality conditions for $\bar{x}$ in both qualified/KKT and non-qualified/Fritz John forms.

(iii) Extend the results of (i) and (ii) to appropriate infinite-dimensional spaces and specify the results for the operator constraints $f(x) = 0 \in Y$ with $\dim Y = \infty$.

Hint: Employ the corresponding calculus rules in the framework of Proposition 6.4 with $\Omega_1 := f^{-1}(\Theta)$, $\Omega_2 := \Omega$; compare it with [523, Theorems 5.7,5.8,5.11].

Exercise 6.34 (Optimization Problems with Inverse Image Constraints via the Extremal Principle). Given $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$, and $\Theta \subset \mathbb{R}^m$, consider the optimization problem:

$$\text{minimize } \varphi(x) \text{ subject to } F^{-1}(\Theta) \cap \Omega. \tag{6.68}$$

(i) Let $\bar{x}$ be a local minimizer for problem (6.68). Show that the point $(\bar{x}, \varphi(\bar{x}))$ is locally extremal for the system of three sets in $\mathbb{R}^n \times \mathbb{R}$:

$$\Omega_1 := \operatorname{epi}\varphi, \quad \Omega_2 := F^{-1}(\Theta), \quad \Omega_3 := \Omega \times \mathbb{R}.$$

(ii) Derive necessary optimality conditions for problem (6.68) by applying the extremal principle to the set system in (i).

(iii) Extend the result of (ii) to problem (6.68) in Asplund spaces.

Exercise 6.35 (Suboptimality Conditions in Nonlinear and Nondifferentiable Programming). Consider problem (6.11), fix $\varepsilon > 0$, and recall that x_ε is an ε-*optimal* (suboptimal) solution to this problem if it is feasible to (6.11) and satisfies the inequality $\varphi_0(x_\varepsilon) \le \inf \varphi_0(x) + \varepsilon$, where the infimum of φ_0 is taken over all the feasible solutions to problem (6.11).

(i) Assume that $\Omega = \mathbb{R}^n$ and that the functions $\varphi_0, \ldots, \varphi_{m+r}$ are strictly differentiable on the set of ε-optimal solutions to (6.11) while $\varphi_1, \ldots, \varphi_{m+r}$ satisfy the Mangasarian-Fromovitz constraint qualifications (see Exercise 2.53) on this set. Then for any ε-optimal solution x_ε to (6.11) and any $\gamma > 0$, there exist an ε-optimal solution $\bar{x}$ to this problem and multipliers $\lambda_1, \ldots, \lambda_{m+r}$ such that

$$\|\bar{x} - x_\varepsilon\| \le \gamma, \ \lambda_i \ge 0, \ \lambda_i \varphi_i(\bar{x}) = 0 \ \text{ for } \ i = 1, \ldots, m,$$
$$\Big\|\nabla\varphi_0(\bar{x}) + \sum_{i=1}^{m+r} \lambda_i \nabla\varphi_i(\bar{x})\Big\| \le \frac{\varepsilon}{\gamma}.$$

(ii) Assume that φ_i, $i = 0, \ldots, m + r$, are locally Lipschitzian on the set of ε-optimal solutions to (6.11) and that Ω is closed therein. Then for any ε-optimal solution x_ε to (6.11) and any $\gamma > 0$, there exist an ε-optimal solution $\bar{x}$ to this problem and multipliers $\lambda_0, \ldots, \lambda_{m+r}$ such that $\|\bar{x} - x_\varepsilon\| \le \gamma$,

$$\Big\|\sum_{i \in I(\bar{x}) \cup \{0\}} \lambda_i x_i^* + x^*\Big\| \le \frac{\varepsilon}{\gamma}, \qquad \sum_{i \in I(\bar{x}) \cup \{0\}} \lambda_i = 1$$

with some $\lambda_i \ge 0$ for $i \in I(\bar{x}) \cup \{0\}$, $x^* \in N(\bar{x}; \Omega)$, $x_0^* \in \partial\varphi_0(\bar{x})$,

$$x_i^* \in \partial\varphi_i(\bar{x}) \ \text{ for } \ i \in \big\{1, \ldots, m\big\} \cap I(\bar{x}), \quad \text{ and}$$
$$x_i^* \in \partial\varphi_i(\bar{x}) \cup \partial(-\varphi_i)(\bar{x}) \ \text{ for } \ i = m + 1, \ldots, m + r.$$

(iii) Extend the results in (i) and (ii) to problems (6.11) in Asplund spaces.

Hint: Employ the subdifferential variational principle from Exercise 2.39 and then the subdifferential sum rule from Corollary 2.20; cf. [523, Theorem 5.30].

Exercise 6.36 (Single-Level Reduction of Optimistic Bilevel Programs)

(i) Show that standard constraint qualifications (of the Mangasarian-Fromovitz type, etc.) fail for the nondifferentiable program (6.34). *Hint:* Compare it with the results and proofs in [194, 745, 748].

(ii) Show that the inner semicontinuity assumption on the mapping $\widetilde{S}$ from (6.36) at $(\bar{x}, \bar{y})$ is essential for the validity of Proposition 6.13(ii).

(iii) Prove that assertion (i) of Proposition 6.13 holds for *some* $\bar{y} \in S(\bar{x})$ if the latter set from (6.30) is assumed to be bounded and the upper-level cost function $f(\bar{x}, \cdot)$ is assumed to be l.s.c. on $S(\bar{x})$. Give examples showing that both of these assumptions are essential for the validity of the result in question. Does it follow from the presented version of Proposition 6.13(i)?

Exercise 6.37 (Partial Calmness and Uniform Weak Sharp Minima in Bilevel Programming). Consider the class of optimistic bilevel programs in form (6.34).

(i) Let $\Omega = \mathbb{R}^n$, and let g_i in (6.34) be *linear* with respect y with $\operatorname{dom} G = \mathbb{R}^n$. Prove that any local optimal solution $(\bar{x}, \bar{y})$ to (6.34) is partially calm provided that f is locally Lipschitzian around this point; compare with the results in [201, 748].

(ii) Construct an example of a bilevel program partially calm at its local optimal solution without the validity of the uniform weak sharp minimum condition (6.43).

(iii) Construct an example of a bilevel program where the partial calmness condition fails at a local optimal solution.

Exercise 6.38 (Sufficient Conditions for Uniform Weak Sharp Minima).

(i) Under which assumptions on problem (6.34) the pointwise local weak sharp minima as in (6.42) with data (6.44) yields the uniform one as in (6.43)?

(ii) Show that the set of optimal solutions to problem (6.49) consists of weak sharp minimizers if either $a_i > 0$ or $b_i < 0$ for $i = 1, 2$. *Hint:* Compare it with [133].

(iii) Derive sufficient conditions for uniform sharp minima in the case of quadratic lower-level problems. *Hint:* Compare it with [748, 749].

Exercise 6.39 (Kernel Condition for Weak Sharp Minima)

(i) Is the kernel condition (6.46) equivalent to a full rank property of a matrix?

(ii) Apply the kernel condition (6.46) in the parametric version of (6.45) with data (6.44) to ensure the uniform weak sharp minima in bilevel programming.

(iii) Extend the result of Proposition 6.18 to Lipschitzian nonlinear programs, and apply it to bilevel programs with nonsmooth data.

Exercise 6.40 (Inf-Differentiability and Dual Characterizations of Weak Sharp Minimizers). Considering a function $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ and a set $\Omega \subset \mathbb{R}^n$, we say as in [785] that φ is *inf-differentiable* at $\bar{x} \in \operatorname{dom}\varphi$ *relative* to Ω if

$$\liminf_{x \xrightarrow{\Omega} \bar{x},\, u \to \bar{x}} \frac{\varphi(u) - \varphi(x) - d\varphi(x; u - x)}{\|u - x\|} = 0, \tag{6.69}$$

where the contingent directional derivative $d\varphi$ is taken from (1.42). In particular, if (6.69) holds with $\Omega = \mathbb{R}^n$ and with $\Omega = \{\bar{x}\}$, then φ is called to be *inf-differentiable at* $\bar{x}$ and *single inf-differentiable* at $\bar{x}$, respectively.

(i) Verify that if φ is locally Lipschitzian around $\bar{x}$, then it is single inf-differentiable at this point. Could the later property hold for non-Lipschitzian functions?

(ii) Show that every convex function is inf-differentiable on any closed and bounded subset of the interior of its domain.

(iii) Let φ be locally Lipschitzian around $\bar{x}$, subdifferentially regular on the set $L_\varphi(\bar{x}) := \{x \in \mathbb{R}^n \mid \varphi(x) = \varphi(\bar{x})\}$ and inf-differentiable at $\bar{x}$ relative to $L_\varphi(\bar{x})$. Prove that the existence of $\eta, r > 0$ such that the inclusion

$$N\big(x; L_\varphi(\bar{x})\big) \cap \eta\mathbb{B} \subset \partial\varphi(x)$$

holds for any $x \in L_\varphi(\bar{x}) \cap B_r(\bar{x})$ is necessary and sufficient for the following specification of local weak sharp minima in Definition 6.16:

$$\eta \operatorname{dist}\big(x; L_\varphi(\bar{x})\big) \le \varphi(x) - \varphi(\bar{x}) \quad \text{whenever} \quad x \in B_r(\bar{x}).$$

Hint: Compare (i)–(iii) with the corresponding statements and proofs in [785].

(iv) Clarify possible counterparts of (iii) for the study of uniform weak sharp minima in parametric optimization and bilevel programs.

Exercise 6.41 (Regular Subgradients of Value Functions in Lower-Level Problems). Let $\mu(x)$ be the optimal value function of the lower-level problem in (6.34).

(i) Give a detailed proof of inclusion (6.65) in general Banach spaces.

(ii) Show that the equalities don't hold in (6.64) and (6.65) for problems with smooth data in finite dimensions.

Exercise 6.42 (Comparing Necessary Optimality Conditions for Bilevel Programs with Smooth Data). Construct examples in which all the assumptions of both Theorem 6.21 and Theorem 6.23 are satisfied while the necessary optimality conditions obtained in these theorems are independent.

Exercise 6.43 (Necessary Optimality Conditions in Lipschitzian Bilevel Programming). Consider local optimal solutions to the optimistic model (6.31).

(i) Give a detailed proof of Theorem 6.25. *Hint:* Compare it with [195, 540].

(ii) Give a detailed proof of Theorem 6.26. *Hint:* Compare it with [540].

(iii) Extend these theorems to bilevel programs with Lipschitzian (and smooth) data in the presence of equality constraints.

(iv) Derive versions of these results for bilevel problems in Asplund spaces. *Hint:* Apply the calculus rules used in the proofs of Theorems 6.21 and 6.23, their equality constraint versions presented in Chapters 2 and 4, and their infinite-dimensional extensions discussed therein in the commentaries and exercises.

(v) Investigate the possibility to improve the necessary optimality conditions in Theorems 6.21 and 6.25 by using the symmetric subdifferential $\partial^0\mu(\bar{x})$ of the value function (6.33) instead of the convexified one in their proofs.

Exercise 6.44 (Extended Inner Semicontinuity in Bilevel Programming). Obtain finite-dimensional and Asplund space extensions of Theorems 6.21 and 6.25 with replacing the inner semicontinuity of the solution map $S(x)$ by its μ-inner semicontinuity defined in Exercise 4.21. *Hint:* Proceed similarly the proofs of these theorems and compare it with [540].

Exercise 6.45 (Bilevel Programs with Inner Semicompact Solution Maps for Lower-level Problems). Considering the solution map $S(x)$ of the lower-level problem in (6.34), verify the following assertions:

(i) $S(x)$ may not be inner semicontinuous at $(\bar{x}, \bar{y})$ as in Theorems 6.21 and 6.25.

(ii) Show that the necessary optimality conditions of Theorems 6.21 and 6.25 may fail without the inner semicontinuity requirement imposed on $S(x)$ at $(\bar{x}, \bar{y})$ under the validity of the other assumptions therein.

(iii) Derive the corresponding version of Theorems 6.21 and 6.25 with replacing the inner semicontinuity of $S(x)$ by its inner semicompactness as well as the more general μ-semicompactness property. *Hint:* Proceed as discussed in Remark 6.27 in the case of finite-dimensional and Asplund spaces.

Exercise 6.46 (Convex Bilevel Programs). Consider the bilevel program (6.31) and its partially calm local optimal solution. Suppose that the lower-level cost and constraint functions are convex jointly with respect to all their variables.

(i) Prove the convexity of the optimal value function (6.33).

(ii) Assuming that the upper-level cost and constraint functions are also fully convex, derive a specification of Theorem 6.21 by using the decomposition property

$$\partial\psi(\bar{x}, \bar{y}) \subset \partial_x\psi(\bar{x}, \bar{y}) \times \partial_y\psi(\bar{x}, \bar{y})$$

valued for full and partial subdifferentials of convex continuous functions ψ and the symmetric property $\partial(-\varphi)(\bar{x}) \subset -\partial\varphi(\bar{x})$. *Hint:* Compare this with [195].

(iii) Assuming that all the functions involved in (6.31) are continuously differentiable in addition to full convexity at the lower level, derive a further specification of Theorem 6.21 by using the equality formula

$$\partial\mu(\bar{x}) = \bigcup_{(\lambda_1,\ldots,\lambda_p)\in\Lambda(\bar{x},\bar{y})} \Big\{\nabla_x\varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \nabla_x g_i(\bar{x}, \bar{y})\Big\} \tag{6.70}$$

for the subdifferential of the optimal value function, where

$$\Lambda(\bar{x}, \bar{y}) := \Big\{(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p \Big|\ \nabla_y\varphi(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \lambda_i\nabla_y g_i(\bar{x}, \bar{y}) = 0,$$
$$\lambda_i \geq 0,\ \lambda_i g_i(\bar{x}, \bar{y}) = 0,\ i = 1, \ldots, p\Big\}.$$

Hint: Deduce (6.70) from the equality representation

$$\partial\mu(\bar{x}) = \bigcup_{(u,v)\in\partial\varphi(\bar{x},\bar{y})} \Big\{u + D^*G(\bar{x}, \bar{y})(v)\Big\}$$

for the subdifferential of the marginal function (6.33) with a convex function φ and a convex-graph mapping G given in [537, Theorem 2.61] and the normal cone representations from Exercises 2.51

and 2.52 with the equalities therein for convex functions due to the equality statement of Theorem 2.26. Compare this with another approach to justify (6.70) with convex differentiable data originated in [703].

(iv) Derive the corresponding consequences of Theorem 6.23 for convex bilevel programs with continuous data cost functions and inequality constraints.

Exercise 6.47 (Hölder Subgradients in Bilevel Programming). Given a Banach space X, we say as in [108] that $x^* \in X^*$ is a *Hölder subgradient* of order $s \geq 0$ for $\varphi\colon X \to \overline{\mathbb{R}}$ at $\bar{x} \in \operatorname{dom}\varphi$ if there are constants $C \geq 0$ and $r > 0$ such that

$$\langle x^*, x - \bar{x}\rangle \leq \varphi(x) - \varphi(\bar{x}) + C\|x - \bar{x}\|^{1+s} \quad \text{for all} \quad x \in \bar{x} + r\mathbb{B}. \tag{6.71}$$

The collection of all x^* satisfying (6.71) is called the *s-Hölder subdifferential* of φ at $\bar{x}$ and is denoted by $\widehat{\partial}_{H(s)}(\bar{x})$. The case of $s = 0$ in (6.71) reduces to the regular/Fréchet subdifferential, while the case of $s = 1$ corresponds to the proximal subdifferential $\partial_p\varphi(\bar{x})$ defined above. We also consider the *upper s-Hölder subdifferentials* of φ at $\bar{x} \in \operatorname{dom}\varphi$ defined symmetrically by

$$\widehat{\partial}^+_{H(s)}\varphi(\bar{x}) := -\widehat{\partial}_{H(s)}(-\varphi)(\bar{x}).$$

Similarly to our basic subdifferential, let us introduce the *limiting s-Hölder subdifferential* $\partial_{H(s)}(x)$ of φ at $\bar{x}$ by taking the outer limit of $\widehat{\partial}_{H(s)}(x)$ as $x \xrightarrow{\varphi} \bar{x}$.

(i) Show that the regular subgradient difference rule given in Lemma 6.22 can be extended to the s-Hölder subdifferentials of any real order $s \geq 0$. *Hint:* Proceed as in the proof of Lemma 6.22 and compare it with [540].

(ii) For each $s \geq 0$, determine the classes of Banach spaces, where the limiting s-Hölder subdifferential $\partial_{H(s)}(\bar{x})$ agrees with our basic limiting construction $\partial\varphi(\bar{x})$, and where these constructions may be different.

(iii) Derive counterparts of the necessary optimality conditions from Theorem 6.26 in terms of the corresponding s-Hölder subdifferentials, and clarify whether they are different, in appropriate Banach spaces, from those given in the theorem. *Hint:* For the latter part, apply the tools of analysis developed in [108].

Exercise 6.48 (Mathematical Programs with Equilibrium Constraints). This class of optimization problems (abbr. *MPECs*) is written in the form:

$$\text{minimize} \quad f(x, y) \quad \text{subject to} \quad y \in S(x), \ \ x \in \Omega, \tag{6.72}$$

where $f\colon X \times Y \to \overline{\mathbb{R}}$ is defined on finite-dimensional or infinite-dimensional spaces and where $S\colon X \rightrightarrows Y$ is given, with $q\colon X \times Y \to Z$ and $Q\colon X \times Y \rightrightarrows Z$, by

$$S(x) := \big\{y \in Y \big|\ 0 \in q(x, y) + Q(x, y)\big\}, \tag{6.73}$$

i.e., $x \mapsto S(x)$ is the solution map to the parametric variational system in (6.73). The latter is often labeled as the parameterized generalized equation (GE) if $Q(x, y) = N(y; G(x))$ for some $G\colon X \rightrightarrows Y$; cf. Section 3.3 with a bit different notation.

(i) Derive necessary optimality conditions for abstract MPECs given in form (6.72) under the most general assumptions on $f(x, y)$ and $S(x)$, and then deduce from them optimality conditions for (6.73) entirely via the initial data q, Q, G. Provide specifications of the obtained results in the case of finite-dimensional spaces. *Hint:* Reduce the models under consideration to those studied in Section 6.1 and , then apply to the necessary optimality conditions therein the corresponding results of generalized differential calculus. Compare it with [523, Section 5.2].

(ii) Under which assumptions the solution map $S(x)$ for the lower-level problem (6.30) can be equivalently written in the MPEC form (6.73)?

(iii) Investigate relationships between global and local solutions to optimistic bilevel programs and to MPECs in (6.72), (6.73) for the case where the lower-level program in (6.28) is convex. *Hint:* Consult [194] for problems with smooth data.

Exercise 6.49 (Value Function Constraint Qualification). Consider the class of optimistic bilevel programs defined by (6.72) with $\Omega = \mathbb{R}^n$ and the solution map to the lower-level problem given as

$$S(x) := \text{argmin}\big\{\varphi(x, y)\big|\ y \in G\big\} \ \text{ for } \ G := \big\{y \in \mathbb{R}^m\big|\ g_i(y) \le 0,\ i = 1, \ldots, p\big\},$$

where $\varphi\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is convex and continuously differentiable in y together with the functions $g_i : \mathbb{R}^m \to \mathbb{R}$. Following [341], introduce the parameterized sets

$$C(\nu) := \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m\big|\ \varphi(x, y) - \mu(x) \le \nu\big\}, \quad \nu \in \mathbb{R},$$

involving the value function $\mu(x)$ from (6.33) for $G(x) = G$, and say that the *value function constraint qualification* (VFCQ) is satisfied at $(\bar{x}, \bar{y}) \in \text{gph}\, S$ if the mapping $C\colon \mathbb{R} \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ is calm at $(0, \bar{x}, \bar{y})$ as defined in Exercise 3.51.

(i) Verify that $S(x)$ can be equivalently written, under the assumptions made, in the MPEC form (6.73) with $q(x, y) = \nabla_y\varphi(x, y)$ and $Q(x, y) = N(y; G)$. *Hint:* Use the classical necessary and sufficient conditions in convex programming.

(ii) Show that if the bilevel program defined in this way has the uniform weak sharp minima (6.43) around the local solution pair $(\bar{x}, \bar{y})$, then VFCQ is satisfied at $(\bar{x}, \bar{y})$. Give an example that the reverse implication fails.

(iii) Verify that the validity of VFCQ at $(\bar{x}, \bar{y})$ ensures that the partial calmness property holds at this point, but not vice versa.

(iv) Assuming that the set G is bounded and that VFCQ is satisfied at $(\bar{x}, \bar{y})$, prove that the perturbation mapping

$$M(\nu) := \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m\big|\ \nu \in \nabla_y\varphi(x, y) + N(y; G)\big\}, \quad \nu \in \mathbb{R}, \tag{6.74}$$

is *calm* at $(0, \bar{x}, \bar{y})$ as defined in Exercise 3.51, while the latter property is strictly weaker than VFCQ in the setting under consideration.

Hint: Consult [341] for the proofs of the results stated in (ii)–(iv).

Exercise 6.50 (Necessary Optimality Conditions for Optimistic Bilevel Programs Without Imposing Partial Calmness).

(i) Investigate the possibility of deriving necessary optimality conditions for the optimistic bilevel program (6.31) by applying the corresponding results of the Fritz John type from Section 6.1 to the equivalent nondifferentiable program (6.34).

(ii) With the usage of the necessary optimality conditions for problems in the general form (6.72) obtained in [523, Subsection 5.2.1] and expressed via the basic coderivative $S(x)$, while normal and mixed versions of it in Asplund spaces, derive their specifications for bilevel programming by evaluating the coderivatives of the solution map (6.30) to the lower-level problem. *Hint:* Consult [198] and the references therein for evaluating the basic coderivative of $S(x)$ in finite dimensions.

(iii) Following the approach of [341] developed for the optimistic bilevel programs with convex lower-level problems and the MPEC solution maps described in Exercise 6.49, derive necessary optimality conditions for *nonconvex* bilevel programs with replacing the partial calmness as in Theorem 6.21 by the *calmness property* of the perturbation mapping (6.74) at $(0, \bar{x}, \bar{y})$ in the sense of Exercise 3.51.

(iv) Compare the results of [341] with those presented in Section 6.3 in the same smooth and convex setting, and then investigate the possibility of extending the approach of [341] to the more general frameworks studied above.

Exercise 6.51 (Two-Level Value Function in Bilevel Programming). Consider the cost function $f_{opt}(x)$ in (6.31) with $S(x)$ taken from (6.30); $f_{opt}(x)$ is labeled as the *two-level optimal value function* in bilevel programming [198].

(i) Evaluate the basic and singular subdifferentials of f_{opt}, and then establish verifiable conditions for the local Lipschitz continuity of this function around a local solution to the optimistic bilevel program (6.31) by using Corollary 4.3 and the Lipschitz-like property of $S(x)$ via the coderivative criterion from Theorem 3.3.

(ii) Apply (i) to deriving necessary optimality conditions in the original optimistic model (6.31), which may not be locally equivalent to model (6.34) studied above; see Proposition 6.13. Compare it with the results presented in Section 6.3.

(iii) Implement this approach to justifying various types of *stationarity* in optimistic bilevel programming as formulated, e.g., in [198].

Hint: Consult [198] for the results, proofs, and additional material.

Exercise 6.52 (Necessary Optimality Conditions in Pessimistic Bilevel Programming). Consider the class of pessimistic bilevel programs (6.32) with the cost function $f_{pes}(x)$ under the same constraints as in (6.31).

(i) Employing the results of Exercise 6.51(i) on the local Lipschitz continuity Lipschitz continuity of f_{opt} with taking into account that $f_{pes} = -f_{opt}$, derive necessary optimality and stationarity conditions for (6.32) from those in Exercise 6.51(ii,iii).

(ii) Derive upper subdifferential conditions for pessimistic bilevel programs from the corresponding results of Section 6.1.

Hint: Consult [199] for more details on both (i) and (ii).

Exercise 6.53 (Multiobjective Approach to Bilevel Programming). Given an upper-level objective function $f\colon X \times Y \to \mathbb{R}$ and the solution map $S\colon X \rightrightarrows Y$ to the lower-level problem as described in (6.30) in the cases of finite-dimensional or infinite-dimensional spaces X and Y, consider the set-valued mapping $F\colon X \rightrightarrows \mathbb{R}$ given in the composition form $F(x) := f(x, S(x))$ for $x \in X$, and rewrite the upper-level problem of bilevel programming as follows:

$$\text{minimize} \quad F(x) \quad \text{subject to } x \in \Omega \tag{6.75}$$

with respect to the standard order on $\mathbb{R}$, where the upper-level constraint set Ω in (6.75) can be represented as or added by some other types of constraints (functional, operator, complementarity, equilibrium, etc.).

(i) Applying to (6.75) the coderivative and subdifferential types of necessary optimality conditions obtained in Section 9.4 for multiobjective optimization together with the coderivative/subdifferential chain rules for the composition $f(x, S(x))$ and then evaluating the coderivative of S, derive necessary optimality conditions for bilevel programs in terms of their initial data.

(ii) Specify the results obtained in this way for optimistic and pessimistic models of bilevel programming, and compare it with those derived and discussed above.

6.5 Commentaries to Chapter 6

Section 6.1. Deriving *necessary optimality conditions* for optimization problems with *nonsmooth data* has been among early motivations to develop constructions and machinery of modern variational analysis and generalized differentiation. Nonsmoothness naturally appeared in the original framework of *optimal control* problems starting from the mid-1950s; see [645]. A simple albeit typical problem of this type was formulated as minimizing a cost function $\varphi(x(1))$ depending on the right endpoints of trajectories for the ODE control system

$$\frac{dx}{dt} = f(x, u), \; x(0) = x_0 \in \mathbb{R}^n, \; u(t) \in U, \; t \in T := [0, 1] \tag{6.76}$$

over measurable (or piecewise continuous) control functions $u(t)$ on T with values belonging to the prescribed closed set $U \subset \mathbb{R}^m$. Since the feasible control region U may be arbitrary (a typical case is when U consists of finitely many points as in systems of automatic control), the formulated optimal control problem can be treated as an optimization problem with irregular geometric constraints regardless of smoothness assumptions on the given functions φ and f. Furthermore, this problem can be equivalently rewritten in form (6.1) studied above, where $\Omega \subset \mathbb{R}^n$ is the reachable set of trajectory endpoints generated by feasible controls in (6.76). Optimal control theory from the very beginning, while revolving around different proofs and extensions of the Pontryagin maximum principle, has been seeking appropriate techniques to deal with this *intrinsic nonsmoothness*. It was a major driving force to develop modern forms of variational analysis that invoke generalized differentiation.

Another remarkable class of intrinsically nonsmooth optimization problems was also discovered in the mid-1950s and named *dynamic programming* by Bellman [77]. His "Principle of Optimality" led him to the so-called Bellman equation for the corresponding optimal value function while *assuming* the smoothness of the latter. Since this assumption fails even in simple examples, the Bellman equation plays just a heuristic role in some practical problems but generally may result in wrong conclusions; see, e.g., [645]. Comprehensive theories of the Hamilton-Jacobi-Bellman and related PDE equations with numerous applications have been developed in the frameworks of *viscosity* and *minimax* solutions by using tools of generalized differentiation; see the books [66, 136, 167, 268, 698] and the references therein.

In fact, intrinsic (often hidden) nonsmoothness already appears at the very fundamental level of modern optimization for problems with *inequality constraints*

$$\varphi_i(x) \le 0, \quad i \in I, \tag{6.77}$$

where the index set I may be finite (while fairly large as, i.e., in linear programming) or infinite as in semilinear programming studied below in Chapters 7 and 8. It is well recognized that the development of efficient machinery for studying and solving optimization problems with inequality constraints is probably the most monumental contribution of mathematical optimizers to society. Saying this, we observe that the inequality constraints (6.77) closely relate to nonsmoothness even in the case of finitely many linear functions φ_i. Geometrically it is manifested by the vertices of convex polyhedra that are described by (6.77) and play a crucial role in the groundbreaking simplex algorithm to solve linear programs. Analytically nonsmoothness is revealed via the equivalent replacement of (possibly great many) inequality constraints in (6.77) by the *single one*

$$\phi(x) := \max\big\{\varphi_i(x)\big|\ i \in I\big\} \le 0$$

given by the *maximum function* $\phi(x)$, which is nondifferentiable even in the case of two linear functions on the real line: $\phi(x) = \max\{x, -x\} = |x|$. As the reader can see in this book, among other numerous publications, maximum/supremum functions and their generalized differentiation are highly important for the study and applications of various types of optimization and equilibrium problems.

To complete these discussions on the role of nonsmoothness in optimization, observe that nondifferentiable functions unavoidably arise while applying *perturbation* and *approximation techniques*, which are central in modern variational analysis, to problems with smooth initial data. Also powerful *variational principles* (notably the Ekeland one) lead us to considering nonsmooth optimization problems.

Now we comment on some specific results presented in Section 6.1 and the corresponding exercises from Section 6.4. *Lower subdifferential optimality conditions* in terms of basic normals and subgradients were derived by using the method of metric approximations in the original publications by the author [502, 503, 504, 507] and those joint with Kruger [439, 440, 528]. Their infinite-dimensional extensions were given in [441, 426, 430] for problems in Fréchet smooth spaces under certain Lipschitzian assumptions and in [516, 523] for the case of Asplund spaces under SNC-type requirements imposed on the sets, mappings, and functions in question. Clarke's version (6.21) of the generalized Lagrange multiplier rule was obtained in [164, 165], and Warga's rule (6.22) was derived in [736, 737]. Other results in this direction in both Fritz John and KKT forms under various qualification conditions can be found in, e.g., [16, 84, 273, 326, 328, 366, 523, 678, 685] and the references therein.

Applications of necessary optimality conditions presume that optimal solutions exist. This is not always the case, especially in infinite dimensions. One of the primary motivations for developing of Ekeland's variational principle was to obtain the "almost stationarity" condition for "almost optimal" (suboptimal) solutions formulated in (2.24). More general (lower) *necessary suboptimality conditions* for problems of nonlinear and nondifferentiable programming presented in Exercise 6.35 are based on the *lower subdifferential variational principle* formulated in Exercise 2.39 and are taken from [523, 587], where the reader can find more discussions and references.

Upper subdifferential optimality conditions for *minimization* problems were initiated by the author [519] who obtained the results presented in Section 6.1 and their counterparts for other optimization problems in general Banach spaces; see also [523, Chapter 5]. As discussed in Remark 6.2, upper subdifferential conditions may have serious advantages over lower subdifferential ones provided that $\widehat{\partial}^+\varphi(\bar{x}) \neq \emptyset$. Various classes of such functions were discussed in [523, Subsection 5.5.4].

It is interesting to observe as in Proposition 6.3 that for problems of minimizing the *DC* (*difference of convex*) functions $\varphi_1(x) - \varphi_2(x)$, the upper subdifferential condition (6.3) reduces to the well-known one $\partial\varphi_2(\bar{x}) \subset \partial\varphi_1(\bar{x})$ as in [350]. Note that the class of DC functions as well as its specifications and modifications play an important role in various qualitative and quantitative issues of optimization including its global aspects and numerical algorithms; see, e.g., [203, 302, 311, 327, 329, 350, 487, 355] among many other publications. Problems of this type will be also studied in Chapter 7 below in the framework of semi-infinite programming.

Sections 6.2 and 6.3. *Bilevel programs* constitute a broad class of problems in hierarchical optimization that is very interesting mathematically and important in applications. We refer the reader to the book by Dempe [193] and more recent publications [177, 194, 195, 196, 197, 198, 199, 200, 201, 202, 341, 469, 540, 750, 763, 769, 764] for various versions in bilevel programming, different approaches to their study, and numerous applications. A characteristic feature of bilevel programs, which can be seen in all of their versions, reformulations, and transformations, is *intrinsic nonsmoothness* that creates serious theoretical and algorithmic challenges. Furthermore, it has been well recognized that standard constraint qualifications in nonlinear and nondifferentiable programming fail to fulfill in bilevel optimization.

The *optimistic* version is by far the most investigated one in bilevel programming, while there are many unsolved theoretical questions therein, not even mentioning numerical algorithms. Among several approaches to deriving necessary optimality conditions in optimistic bilevel programs, we present in Sections 6.2–6.3 the *value function approach*, which was initiated by Outrata [619] for a particular bilevel optimization model. This approach explicitly manifests nonsmoothness in bilevel programming via the nondifferentiable lower-level value function (6.33).

The value function approach to optimistic bilevel programs was greatly developed by Ye and Zhu [748], who introduced the *partial calmness* condition that allowed them to reduce bilevel programs to nonsmooth single-level ones via penalization. Combining it with Clarke's generalized differentiation, they derived in [748] necessary optimality conditions for bilevel programs in terms of their initial data.

In this book we mainly follow the papers [195, 540] and further develop the value function approach by employing our basic tools of generalized differentiation to express optimality conditions for nondifferentiable programs from Section 6.1 and then to evaluate basic subgradients of marginal/optimal value functions via the results of Section 4.1. Such a device allows us to essentially improve necessary optimality conditions for optimistic bilevel programs obtained in [748] and other publications. Note the importance of the rather surprising *difference rule* for regular subgradients from Lemma 6.22 established in [546] by using the smooth variational description of regular subgradients in Theorem 1.27.

The partial calmness assumption from Definition 6.14 plays an essential role in the value function approach to bilevel programming. Although it is satisfied in many important settings, it may fail in rather simple nonlinear examples; see the discussions above as well as the results in [133, 201, 198, 748, 749]. A sufficient (while far from being necessary) condition for the validity of partial calmness was introduced by Ye and Zhu [748] under the name of "uniformly weak sharp minima," which could be seen as a version of sharp minima by Polyak [643, 644] and weak sharp minima by Ferris [264]. In contrast to the latter two notions, which have been well investigated and applied in the literature (see, e.g., [132, 133, 237, 335, 462, 495, 546, 608, 697, 744, 782, 785]), uniform weak sharp minima have drawn much less attention. We refer the reader to [133, 327, 744, 748, 749] for some efficient conditions ensuring the validity of the uniform weak sharp minimum estimate (6.43) and also to the discussions right before Proposition 6.18, which seems to be new.

There are several approaches to deriving necessary optimality and stationary conditions that don't employ partial calmness; see [51, 198, 199, 200, 201, 341, 750, 763]. We particularly emphasize remarkable developments by Henrion and Surowiec [341] for the class of optimistic bilevel programs with $\mathcal{C}^2$-smooth data and *convex* lower-level problems, where the solution map to the lower-level problem can be equivalently rewritten in the *MPEC form* (6.73) with $q(x, y) = \nabla_y \varphi(x, y)$ and $Q(x, y) = N(y; G)$; see Exercise 6.49(i). They replace the partial calmness assumption by the weaker *calmness* property of the perturbation mapping (6.74) in the sense defined in Exercise 3.51. Imposing in addition the *constant rank constraint qualification* in the lower-level problem (see [477, 499] for more details about the latter notion), Henrion and Surowiec derive necessary optimality conditions (more precisely, M(ordukhovich)-stationarity conditions) for optimistic bilevel programs, which have serious advantages in comparison with the corresponding results of [195] in such settings. The reader may find more information about MPECs and their applications in the fundamental monographs [482, 624] and the subsequent publications [3, 78, 267, 314, 338, 341, 346, 290, 523, 620, 623, 684, 745, 746, 780] among other works with numerous references therein. See, in particular, the papers by Outrata [620] and Scheel and Scholtes [684] for introducing various notions of stationarity for MPECs, which have been similarly developed later in bilevel programming. Note to this end that, although MPECs [482, 624] and bilevel programs have many things in common, these two classes of optimization problems are essentially different in general; see the papers by Dempe and Dutta [194] and by Dempe and Zemkoho [202] for various results and comprehensive discussions.

Section 6.4. This section contains exercises of different levels of difficulties on necessary optimality conditions in nonsmooth optimization and bilevel programming with hints and references when needed. At the same time, we present here some *challenging* and largely *open questions* concerning various issues of bilevel optimization. They include Exercise 6.38(i), Exercise 6.39(ii,iii), and Exercise 6.40(iv) on uniform weak sharp minima, Exercise 6.43(v) on the usage of the symmetric subdifferential of marginal functions for deriving necessary optimality conditions for bilevel programs, Exercise 6.50 on deriving necessary optimality conditions for optimistic bilevel programs without the partial calmness assumption by using the approaches described therein, Exercise 6.52 and beyond on deriving necessary optimality and stationarity conditions for pessimistic bilevel programs that are considerably underinvestigated in the literature, and Exercise 6.53 on developing a new multiobjective optimization approach to bilevel programs by using the procedure described therein.